

## Chapter 5 Growth Models

<b>5.1</b>	<b>Forecasting with no data (GROWCOMP)</b>	<b>42</b>
<b>5.2</b>	<b>Linear growth (LIN)</b>	<b>47</b>
<b>5.3</b>	<b>Exponential growth (EXP)</b>	<b>51</b>
<b>5.4</b>	<b>Modified exponential growth (MODEXP)</b>	<b>54</b>
<b>5.5</b>	<b>Logistic growth (LOG)</b>	<b>58</b>

*This chapter presents growth models for cases in which when little historical data exist. The growth models are also used in scenario analysis for long-term trend extrapolation based on alternative assumptions about the future.*

*If you have no data at all, use the worksheet in section 5.1, which projects a variety of assumed growth patterns into the future. When data develop, use the models in sections 5.2-5.5 to fit growth curves or functions of time using ordinary-least-squares regression. The linear growth model (Section 5.2) predicts a constant amount of growth each time period, while the exponential (Section 5.3) predicts constant percentage growth. In both linear and exponential growth, the forecasts are unbounded. In the modified exponential (Section 5.4) and logistic (Section 5.5) models, the amount of growth declines each period, so the forecasts are bounded by a saturation level.*

## 5.1 Forecasting with no data (GROWCOMP)

It is heroic to forecast with no data but there may be no alternative. Data may not exist for new products, markets, or business locations or for budgeting new categories of costs or expenses. GROWCOMP in Figure 5-1 is a handy tool for analyzing assumptions about future data. The basic idea is to assume data values for the first period in the future and for some later period, perhaps when we believe the data or market will stabilize. Models are then used to fill in unknown data between these two periods, using a variety of alternative growth patterns. One of these patterns should be more credible than the others and should help develop a final forecast. This procedure saves a lot of work in creating assumed growth patterns and makes the forecasting process more objective.

In Figure 5-1, suppose the year is 1950 and we are interested in predicting the development of the television market, specifically the annual percentage of households in the U.S. that will purchase one or more television sets. Alternative growth patterns or trends are computed based on the set of assumptions shown at the top of the worksheet. First, you specify the data for the first year (1950) in cell C5. Assume perfect hindsight and use the actual penetration value listed in column Q (not shown in Figure 5-1). Thus the first year's data in C5 is set at 9.00, meaning that we expect 9 per cent of households will own at least one television set in 1950. Next, in cell C6, estimate the saturation level, the maximum value of the data in the future or the point at which growth stops. Here we assume saturation will occur at 99% penetration. In cells G5 and G6, two additional assumptions are required: a target period and value in the future. We assume that in period 10 (1959), penetration will reach 85%.

Using these assumptions, four alternative growth patterns are computed in columns C, F, I, and L: linear, simple exponential, modified exponential, and logistic growth. Equations used to compute each type of growth are shown in row 9. (Note: these equations are equivalent to those presented below for fitting regression models.) Model coefficients are listed in rows 10-12. Estimated data values are plotted in Figure 5-2, with percentage growth by period plotted in Figure 5-3.

In Figure 5-2, all growth curves start at the same point, the first-year data assumption of 9.00, and run through the target value of 85 in 1959. The first option, linear growth, is a middle-ground assumption. Growth is simply a constant amount each time period. In percentage terms, linear growth declines more rapidly than the other models.

Figure 5-1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	<b>GROWTH</b>													
2	<b>Growth model projections</b>													
3	<b>TV penetration</b>													
4														
5	<b>First period data</b>	<b>9.00</b>				<b>Target period</b>	<b>10</b>							
6	<b>Saturation level</b>	<b>99.00</b>				<b>Target data</b>	<b>85.00</b>							
7														
8			<b>LINEAR GROWTH</b>			<b>SIMPLE EXPONENTIAL</b>			<b>MODIFIED EXPONENTIAL</b>			<b>LOGISTIC</b>		
9			$Y = a + bx$			$Y = a(1+r)^n$			$Y = a - br^n$			$Y = 1 / (a-br^n)$		
10			a =	0.556		a =	7.0128		a =	99.0000		a =	0.0101	
11			b =	8.444		b =	#N/A		b =	110.6708		b =	-0.1594	
12			r =	#N/A		r =	0.2834		r =	0.8132		r =	0.6337	
13														
14				<b>Growth</b>	<b>Growth</b>		<b>Growth</b>	<b>Growth</b>		<b>Growth</b>	<b>Growth</b>		<b>Growth</b>	<b>Growth</b>
15	<b>Year</b>	<b>Period</b>	<b>Forecast</b>	<b>amount</b>	<b>%</b>	<b>Forecast</b>	<b>amount</b>	<b>%</b>	<b>Forecast</b>	<b>amount</b>	<b>%</b>	<b>Forecast</b>	<b>amount</b>	<b>%</b>
16	1950	1	9.00			9.00			9.00			9.00		
17	1951	2	17.44	8.44	93.83%	11.55	2.55	28.34%	25.81	16.81	186.78%	13.49	4.49	49.93%
18	1952	3	25.89	8.44	48.41%	14.82	3.27	28.34%	39.48	13.67	52.96%	19.74	6.25	46.29%
19	1953	4	34.33	8.44	32.62%	19.02	4.20	28.34%	50.60	11.12	28.16%	27.93	8.19	41.50%
20	1954	5	42.78	8.44	24.60%	24.41	5.39	28.34%	59.64	9.04	17.87%	37.90	9.97	35.68%
21	1955	6	51.22	8.44	19.74%	31.33	6.92	28.34%	66.99	7.35	12.33%	48.97	11.07	29.22%
22	1956	7	59.67	8.44	16.49%	40.21	8.88	28.34%	72.97	5.98	8.92%	60.10	11.13	22.72%
23	1957	8	68.11	8.44	14.15%	51.61	11.40	28.34%	77.83	4.86	6.66%	70.20	10.11	16.82%
24	1958	9	76.56	8.44	12.40%	66.23	14.62	28.34%	81.78	3.95	5.08%	78.58	8.37	11.93%
25	1959	10	85.00	8.44	11.03%	85.00	18.77	28.34%	85.00	3.22	3.93%	85.00	6.42	8.18%
26	1960	11	93.44	8.44	9.93%	109.09	24.09	28.34%	87.61	2.61	3.08%	89.64	4.64	5.46%
27	1961	12	101.89	8.44	9.04%	140.00	30.91	28.34%	89.74	2.13	2.43%	92.86	3.21	3.59%
28	1962	13	110.33	8.44	8.29%	179.67	39.67	28.34%	91.47	1.73	1.93%	95.02	2.16	2.33%
29	1963	14	118.78	8.44	7.65%	230.58	50.91	28.34%	92.88	1.41	1.54%	96.44	1.42	1.50%
30	1964	15	127.22	8.44	7.11%	295.93	65.34	28.34%	94.02	1.14	1.23%	97.36	0.92	0.96%
31	1965	16	135.67	8.44	6.64%	379.78	83.86	28.34%	94.95	0.93	0.99%	97.96	0.59	0.61%
32	1966	17	144.11	8.44	6.22%	487.40	107.62	28.34%	95.71	0.76	0.80%	98.34	0.38	0.39%
33	1967	18	152.56	8.44	5.86%	625.52	138.12	28.34%	96.32	0.62	0.64%	98.58	0.24	0.25%
34	1968	19	161.00	8.44	5.54%	802.78	177.26	28.34%	96.82	0.50	0.52%	98.73	0.15	0.16%
35	1969	20	169.44	8.44	5.24%	1030.26	227.49	28.34%	97.23	0.41	0.42%	98.83	0.10	0.10%
36	1970	21	177.89	8.44	4.98%	1322.21	291.95	28.34%	97.56	0.33	0.34%	98.89	0.06	0.06%
37	1971	22	186.33	8.44	4.75%	1696.89	374.68	28.34%	97.83	0.27	0.28%	98.93	0.04	0.04%
38	1972	23	194.78	8.44	4.53%	2177.75	480.85	28.34%	98.05	0.22	0.22%	98.96	0.03	0.03%
39	1973	24	203.22	8.44	4.34%	2794.86	617.11	28.34%	98.23	0.18	0.18%	98.97	0.02	0.02%
40	1974	25	211.67	8.44	4.16%	3586.85	791.99	28.34%	98.37	0.14	0.15%	98.98	0.01	0.01%
41	1975	26	220.11	8.44	3.99%	4603.26	1016.42	28.34%	98.49	0.12	0.12%	98.99	0.01	0.01%
42	1976	27	228.56	8.44	3.84%	5907.71	1304.44	28.34%	98.58	0.10	0.10%	98.99	0.00	0.00%
43	1977	28	237.00	8.44	3.69%	7581.79	1674.08	28.34%	98.66	0.08	0.08%	99.00	0.00	0.00%
44	1978	29	245.44	8.44	3.56%	9730.27	2148.48	28.34%	98.72	0.06	0.06%	99.00	0.00	0.00%
45	1979	30	253.89	8.44	3.44%	12487.56	2757.30	28.34%	98.78	0.05	0.05%	99.00	0.00	0.00%
46	1980	31	262.33	8.44	3.33%	16026.20	3538.64	28.34%	98.82	0.04	0.04%	99.00	0.00	0.00%
47	1981	32	270.78	8.44	3.22%	20567.59	4541.39	28.34%	98.85	0.03	0.03%	99.00	0.00	0.00%
48	1982	33	279.22	8.44	3.12%	26395.90	5828.30	28.34%	98.88	0.03	0.03%	99.00	0.00	0.00%
49	1983	34	287.67	8.44	3.02%	33875.79	7479.89	28.34%	98.90	0.02	0.02%	99.00	0.00	0.00%
50	1984	35	296.11	8.44	2.94%	43475.27	9599.49	28.34%	98.92	0.02	0.02%	99.00	0.00	0.00%
51	1985	36	304.56	8.44	2.85%	55794.99	12319.72	28.34%	98.94	0.01	0.02%	99.00	0.00	0.00%

Figure 5-2

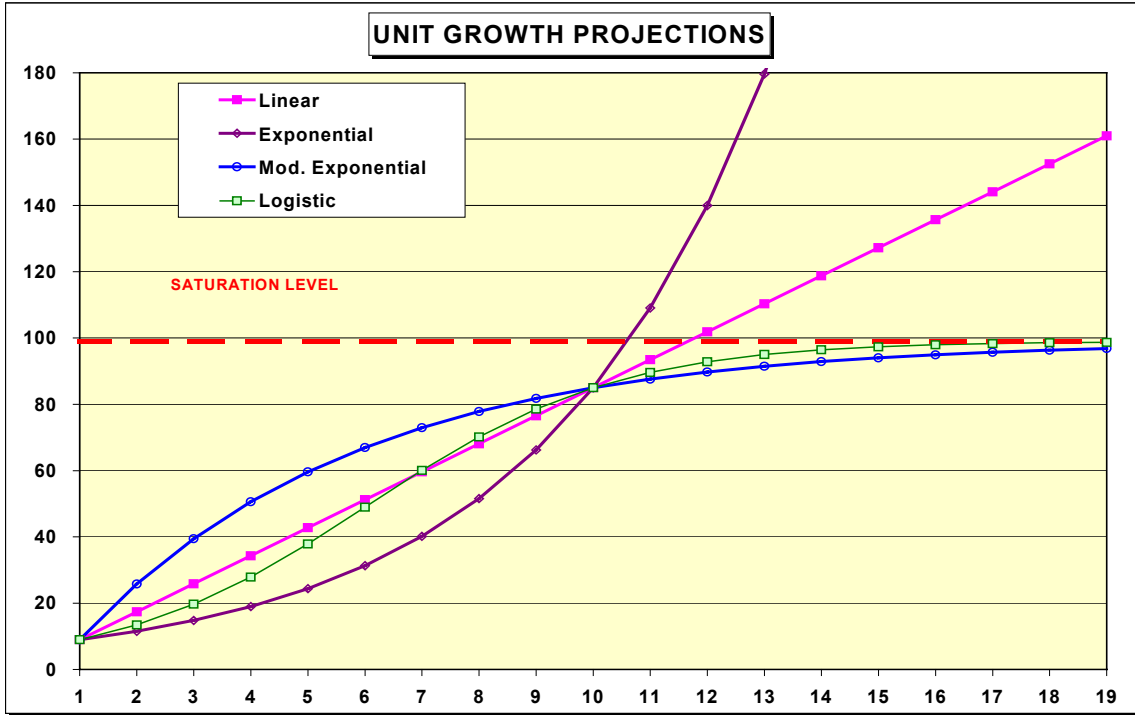
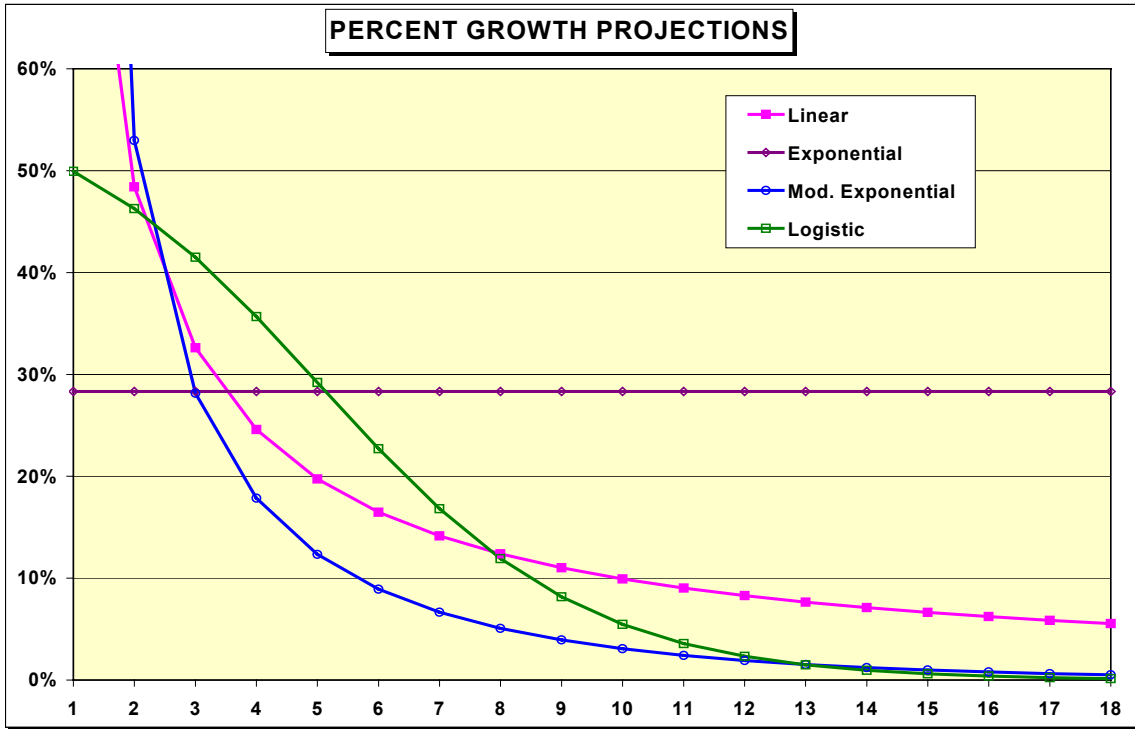


Figure 5-3



The second option, the simple exponential model, produces constant percentage growth. After the target value is reached, exponential growth quickly gets out of hand and reaches extreme penetration values. But it takes some time for the simple exponential to build up steam compared to the other models. The result is that simple exponential growth between any starting point and any target value is always more pessimistic than the other models.

The modified exponential gives an excellent fit to the actual data in column Q (actual data are not plotted to avoid cluttering the graph). Early growth from the modified exponential is sensational, more than 186% from 1950 to 1951. The modified-exponential curve runs far above the other possibilities for every year until the target penetration is reached. After that, the modified exponential gradually approaches the saturation level. Growth slows dramatically as the saturation level is approached. The modified exponential always produces the most optimistic growth pattern between any two data values. If you expect strong early growth in your data, use the modified exponential model as a planning tool.

Finally, logistic growth should be used when you expect growth to follow an S-shaped pattern. This curve is one of a large family of S-shaped curves that display slow growth at first, followed by a period of steep growth, and finally by a gradual decline in growth to the saturation level. This type of forecasting model originated in the study of the spread of epidemics during the early part of this century. Today the logistic is used for the same purpose as well as in forecasting the development of markets for both industrial and consumer goods. The logistic is another middle-ground assumption, falling between the modified exponential and the simple exponential. In terms of percentage growth in Figure 5-3, the logistic always starts out somewhere below the modified exponential. Gradually, the logistic catches up as the saturation level is approached.

Forecasting with no data is hazardous but the process is certainly more objective when you make some attempt to develop a model of the future. Although an overwhelming variety of growth models are available in the literature of forecasting, the ones in this section are commonly used in practice and should at least give you a starting point for making predictions. You can think of the area between the modified exponential and the simple exponential curves as the range of likely values for future data, often called the envelope of growth potential. The size of this range depends on your assumptions about first-period data, target data, and saturation level. Whatever your assumptions, we can summarize as follows: the modified-exponential model generates the best-case scenario, while the simple exponential generates the worst case. The linear and logistic models are middle-ground possibilities.

The most critical assumption in GROWCOMP worksheet is the first-period data value. You should spend most of your time estimating this number and you should do sensitivity analysis on the effects of alternative values. Once you get an actual value for the first period, update the worksheet and look again at the assumptions for the saturation level and the target data in the future.

After three or four data values have been observed, you should attempt to fit a forecasting model to your data. Each of the growth models in GROWCOMP can be fitted to data as discussed in sections 5.2-5.5.

Some will protest that you need a lot more than three or four data values to fit a forecasting model but in the real world you cannot afford to wait. Something must be done to adjust the forecasts based on how the early data behave. The best way to make such adjustments is to switch to a forecasting model based on the data.

## 5.2 Linear growth (LIN)

Perhaps the simplest forecasting model is a straight line fitted to the data using ordinary-least-squares regression. The explanatory or X variable in the regression is time, a series of consecutive integers. The regression line is computed to minimize the sum of squared differences between actual data and the line. Linear-growth regression is appropriate when you expect the amount of growth each time period to be relatively constant. An equivalent assumption is that percentage growth is expected to decline continuously. It is important to understand that all past data are equally weighted in linear-growth regression. If recent data should receive more weight, the trend-smoothing model in Chapter 4 is a better choice.

The linear growth model is written:

$$Y_t = a + bt \quad (5-1)$$

**Y** is the dependent variable and the subscript **t** refers to time, a series of consecutive integers. The regression constant or intercept (starting point) for the line is **a**. The parameter **b** is the slope or amount of growth each period; **b** is also called the X variable coefficient.

Linear growth models are commonly used in modeling population data, as shown in Figures 5-4 and 5-5. The population of DeWalt, Texas, has marched along a straight line since the city was incorporated in 1986 and the forecasts reflect a constant amount of annual growth.

LIN requires installation of the Excel Data Analysis Toolbox. Enter data in columns A and B, then enter the number of warm-up data and the last period to forecast in cells C10 - C11. Next, select Tools, Data Analysis, Regression. In the dialog box for the Y range, paint over cells in column B, starting at row 25. Click on the box for the X range and paint over cells in Column C. Ensure that the number of data used to fit the regression model agrees with the number of warm-up data in cell C10. In the box for output range, enter O1. Finally, click OK.

The estimated Y values in column D are computed using equation 5-1. The **a** and **b** coefficients are found in the regression output (see Figure 5-6) in cells P17 - P18. The intercept is 7105.45 and the slope is 232.60 per time period.

Figure 5-4

	A	B	C	D	E	F	G	H	I
1	LIN.XLS			DeWalt, Texas population			INTERMEDIATE CALCULATIONS:		
2	Linear regression on time						RMSE (Square root of MSE)		45.58
3	$Y_t = a + bt$						3 X RMSE		136.73
4	Select Tools, Data Analysis, Regression. Enter ranges as follows:								
5	Y range includes column B, X range includes column C, Output range = O1.							Warm-up SSE	33234.94
6	Ensure number of data selected for Y and X ranges agree with the							Warm-up MSE	2077.18
7	nbr. of warm-up data entered in cell C10.							Forecasting SSE	33234.94
8							Fcst SSE - Warm-up SSE		0.00
9	INPUT:						Nbr. of forecast periods		0
10	Nbr. of warm-up data		16				Forecasting MSE		0.00
11	Last period to forecast		25						
12							Warm-up Sum abs. err.		574.22
13							Warm-up MAD		35.89
14	OUTPUT:						Forecasting Sum abs. err.		574.22
15	Number of data		16				Fcst Sum abs. - Warm-up Sum at		0.00
16	Nbr. of outliers		0				Nbr. of forecast periods		0
17	Warm-up MSE		2077.18				Forecasting MAD		0.00
18	Forecasting MSE		0.00				Last forecast at period		25
19	Warm-up MAD		35.89						
20	Forecasting MAD		0.00						
21									
22		Regression ranges							
23		Y range	X-range						
24	Mon/yr.	Data	t	Est Y	$e = Y - Est Y$	$e^2$			
25	1986	7,350	1	7338.05	11.95	142.77			
26	1987	7,521	2	7570.65	-49.65	2465.41			
27	1988	7,810	3	7803.25	6.75	45.50			
28	1989	8,070	4	8035.86	34.14	1165.82			
29	1990	8,254	5	8268.46	-14.46	209.02			
30	1991	8,488	6	8501.06	-13.06	170.53			
31	1992	8,812	7	8733.66	78.34	6137.11			
32	1993	9,010	8	8966.26	43.74	1913.03			
33	1994	9,165	9	9198.86	-33.86	1146.72			
34	1995	9,324	10	9431.46	-107.46	11548.66			
35	1996	9,666	11	9664.07	1.93	3.74			
36	1997	9,874	12	9896.67	-22.67	513.82			
37	1998	10,202	13	10129.27	72.73	5289.78			
38	1999	10,342	14	10361.87	-19.87	394.84			
39	2000	10,632	15	10594.47	37.53	1408.35			
40	2001	10,801	16	10827.07	-26.07	679.83			
41	2002		17	11059.68	#N/A	#N/A			
42	2003		18	11292.28	#N/A	#N/A			
43	2004		19	11524.88	#N/A	#N/A			
44	2005		20	11757.48	#N/A	#N/A			
45	2006		21	11990.08	#N/A	#N/A			
46	2007		22	12222.68	#N/A	#N/A			
47	2008		23	12455.28	#N/A	#N/A			
48	2009		24	12687.89	#N/A	#N/A			
49	2010		25	12920.49	#N/A	#N/A			



Figure 5-5

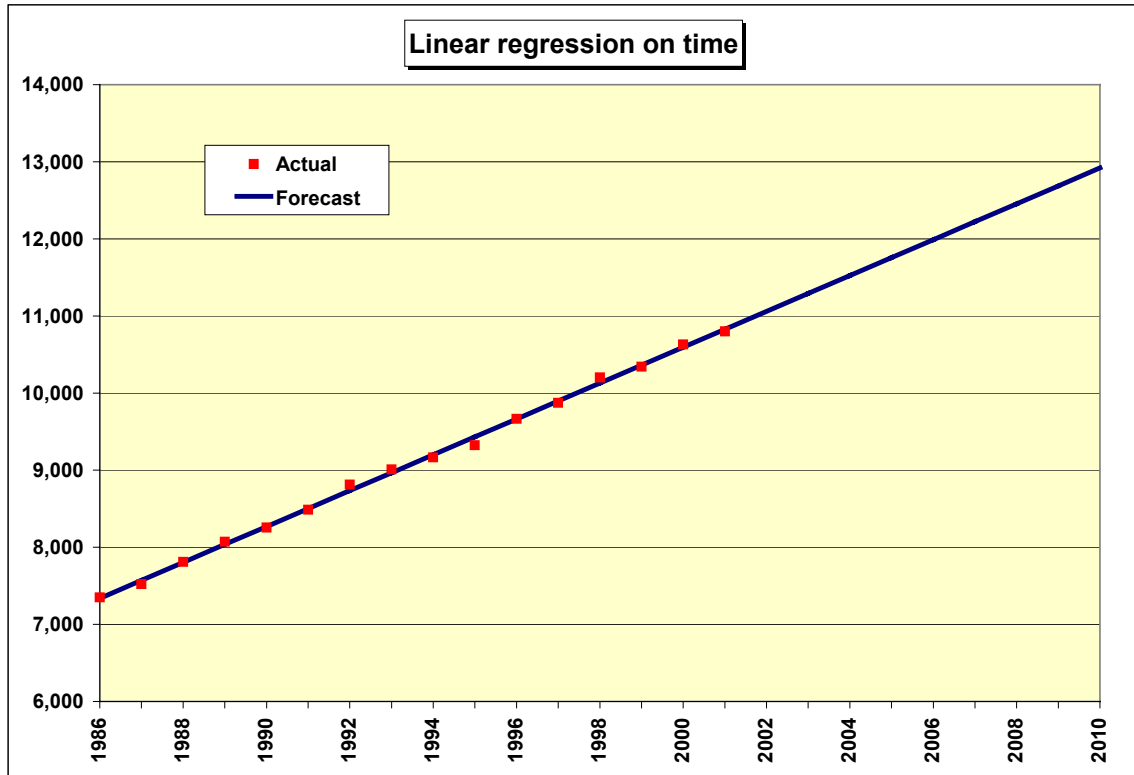


Figure 5-6

	O	P	Q	R	S	T	U
1	SUMMARY OUTPUT						
2							
3	<i>Regression Statistics</i>						
4	Multiple R	0.999097862					
5	R Square	0.998196538					
6	Adjusted R Square	0.998067719					
7	Standard Error	48.72293151					
8	Observations	16					
9							
10	<i>ANOVA</i>						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
12	Regression	1	18395171	18395171	7748.845615	1.30083E-20	
13	Residual	14	33234.93676	2373.924055			
14	Total	15	18428405.94				
15							
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
17	Intercept	7105.45	25.55052084	278.0941354	1.33298E-27	7050.649534	7160.250466
18	X Variable 1	232.6014706	2.64237196	88.0275276	1.30083E-20	226.9341413	238.2687998

A variety of other statistics are found in the regression output. R square (often denoted  $r^2$ ) in cell P5 is a measure of the goodness of fit of the regression line, the proportion of total variation in the Y variable (population) that is explained by the relation with the X variable (time). If all of the data fell on the regression line, R square would be 1.0. The standard error in cell P7 is also called the standard error of the estimate and is the square root of the sum of the squared differences between actual and estimated Y values divided by  $n-2$ , where  $n$  is the number of data. The reason for dividing by  $n-2$  is that two degrees of freedom are lost in computing the regression line.

The regression standard error is similar to the RMSE or square root of the mean-squared-error value in cell I5. Recall that three times the RMSE is used to set control limits in the control-chart graph (not shown here). We will continue to use the RMSE to set control limits for consistency with the control charts for exponential smoothing in Chapter 4.

Several statistical tests can be made on the significance of the linear relationship between X and Y. First, an F statistic is computed in cell S12. Although a full explanation is beyond our scope here, the F statistic is the ratio of explained to unexplained variation in Y. When F is large, the regression line explains a large proportion of the variation in Y. The significance of F is computed in cell T12, interpreted as the probability that the true regression slope or X coefficient is actually equal to zero. This probability is very small, as it should be given the excellent fit of the regression line. An equivalent test can be made with the t-statistic for the X variable in cell R18. The t-statistic is the X coefficient divided by the standard error of the coefficient. The P-value is the probability that the true X coefficient is actually zero and again is very small.

### 5.3 Exponential growth (EXP)

A classic case of exponential growth is the behavior of IBM world-wide sales during the early days of the computer (Figures 5-7 and 5-8). Percentage growth is roughly constant, with increasing amounts of growth each period.

The exponential growth model is written:

$$\ln(Y_t) = a + bt \quad (5-2)$$

Data in column B are converted to logarithms in column D. Logs convert exponential growth to linear growth so the standard linear regression functions can be used. Column D is used as the Y input to the regression, with the X input as the series of consecutive integers in column C. Estimated logs of the Y data are in column E, with antilogs (final forecasts) in column F.

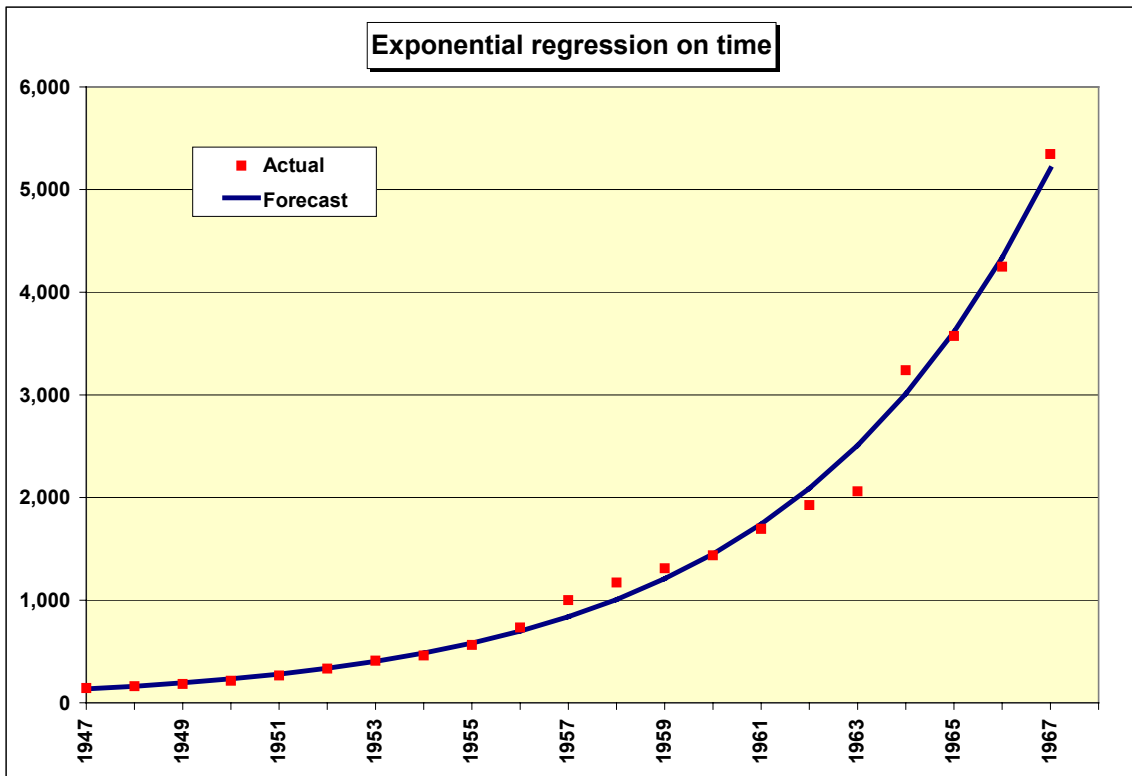
This example presents a forecast simulation because the warm-up data are set to 18 in cell C10. Thus the Y-range for the regression is D25:D42 and the X-range is C25:C42 (the last 3 data are not used to compute the regression equation). The model performs well in the forecast periods (1965-1967) with a forecasting MAD of 89.89 units per year. Also note that the fit of the model is excellent with an R square value of 0.9916 in cell P5 (not shown here).

The exponential growth model is most useful early in the life cycle of a product. Obviously, exponential growth cannot be sustained indefinitely so use this model with caution. The model works well with relatively stable, well-behaved data. If the growth rate in the data changes frequently or if the data contain a great deal of randomness or noise, a better choice is the trend exponential smoothing model from Chapter 4. With the trend modifier set above 1.0, this smoothing model produces an exponential trend.

Figure 5- 7

	A	B	C	D	E	F	G	H	I
1	EXP			IBM World-wide sales			INTERMEDIATE CALCULATIONS:		
2	Exponential regression on time						RMSE (Square root of MSE)		139.34
3	ln(Y <sub>t</sub> ) = a + bt						3 X RMSE		418.02
4	Select Tools, Data Analysis, Regression. Enter ranges as follows:								
5	Y range includes column D, X range includes column C, Output range = O1.						Warm-up SSE		349478.33
6	Ensure number of data selected for Y and X ranges agree with the						Warm-up MSE		19415.46
7	nbr. of warm-up data entered in cell C10.						Forecasting SSE		378736.56
8							Fcst SSE - Warm-up SSE		29258.23
9	INPUT:						Nbr. of forecast periods		3
10	Nbr. of warm-up data		18				Forecasting MSE		9752.74
11	Last period to forecast		21						
12							Warm-up Sum abs. err.		1471.82
13							Warm-up MAD		81.77
14	OUTPUT:						Forecasting Sum abs. err.		1741.49
15	Number of data		21				Fcst Sum abs. - Warm-up Sum abs.		269.67
16	Nbr. of outliers		1				Nbr. of forecast periods		3
17	Warm-up MSE		19415.46				Forecasting MAD		89.89
18	Forecasting MSE		9752.74				Last forecast at period		21
19	Warm-up MAD		81.77						
20	Forecasting MAD		89.89						
21									
22			Regression ranges						
23			X range	Y range					
24	Mon/yr.	Data	t	ln(Y)	Est ln(Y)	Est Y	e = Y - Est Y	e^2	
25	1947	144.54	1	4.97	4.91	135.30	9.24	85.35	
26	1948	161.98	2	5.09	5.09	162.39	-0.41	0.17	
27	1949	183.46	3	5.21	5.27	194.90	-11.44	130.88	
28	1950	214.91	4	5.37	5.45	233.92	-19.01	361.37	
29	1951	266.79	5	5.59	5.64	280.75	-13.96	194.91	
30	1952	333.72	6	5.81	5.82	336.96	-3.24	10.49	
31	1953	409.98	7	6.02	6.00	404.42	5.56	30.94	
32	1954	461.35	8	6.13	6.18	485.38	-24.03	577.61	
33	1955	563.54	9	6.33	6.37	582.56	-19.02	361.71	
34	1956	734.34	10	6.60	6.55	699.19	35.15	1235.63	
35	1957	1,000.40	11	6.91	6.73	839.17	161.23	25995.80	
36	1958	1,171.70	12	7.07	6.91	1007.17	164.53	27069.63	
37	1959	1,309.70	13	7.18	7.10	1208.81	100.89	10178.81	
38	1960	1,436.00	14	7.27	7.28	1450.82	-14.82	219.54	
39	1961	1,694.20	15	7.43	7.46	1741.27	-47.07	2215.99	
40	1962	1,925.20	16	7.56	7.64	2089.88	-164.68	27120.17	
41	1963	2,059.60	17	7.63	7.83	2508.28	-448.68	201315.52	
42	1964	3,239.30	18	8.08	8.01	3010.45	228.85	52373.83	
43	1965	3,572.80	19	8.18	8.19	3613.15	-40.35	1627.81	
44	1966	4,247.70	20	8.35	8.37	4336.51	-88.81	7886.80	
45	1967	5,345.20	21	8.58	8.56	5204.69	140.51	19743.61	

Figure 5-8



## 5.4 Modified exponential growth (MODEXP)

The modified-exponential growth model is widely used to model market penetration of consumer goods and services. A classic example of this type of growth is the penetration of TVs in U.S. households, shown in Figures 5-9 and 5-10. Penetration is defined as the percentage of households that own one or more TVs. Early percentage growth with this model is quite large, so the forecasts grow explosively at first. However, percentage growth declines continuously until the data reach a saturation level, where growth stops.

From 1950 to 1960, TV penetration grew almost ten-fold, from 9% to about 87%. Later growth slowed dramatically as the market approached saturation. In 1981, penetration reached 98.1% and has been virtually constant since then.

Data entry for this model is similar to the earlier models except that you must enter the expected saturation level in cell C12. The equation for the modified-exponential curve is:

$$Y_t = K - ab^t \quad (5-3)$$

**K** is the saturation level. As in the earlier models, **t** refers to time, **a** is the intercept, and **b** is the slope. After taking logs and rearranging, the modified exponential can be written as:

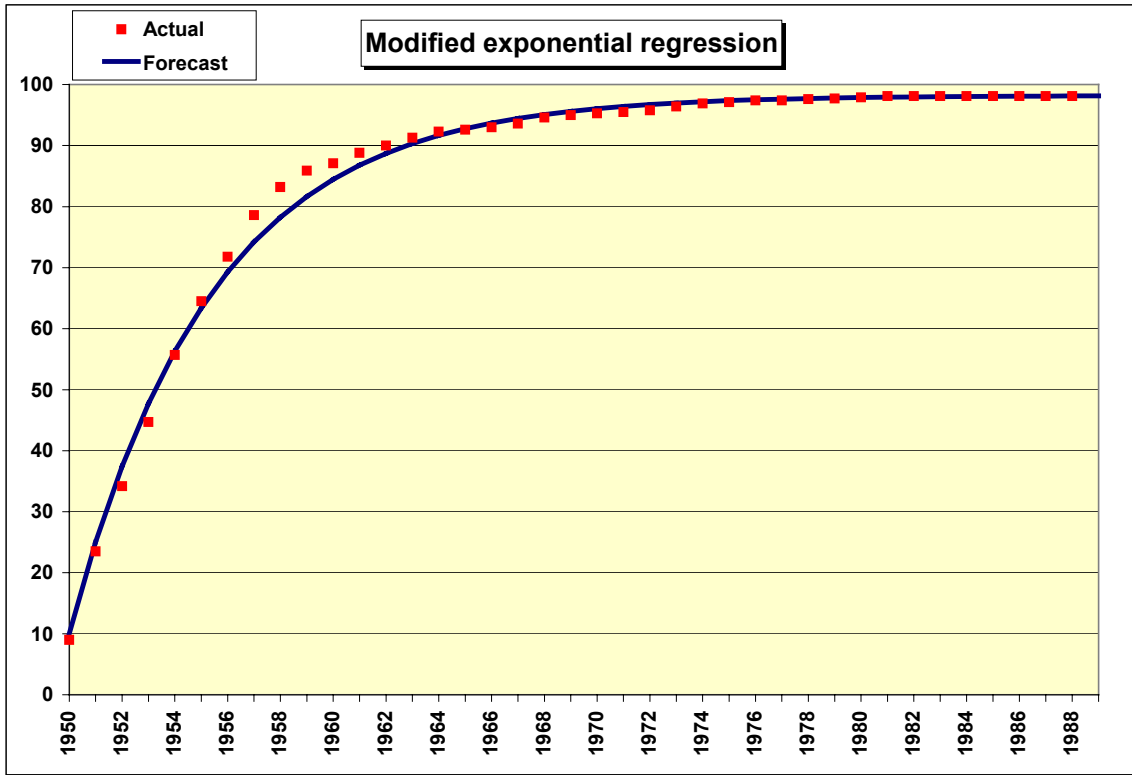
$$\ln(K - Y_t) = \ln(a) + t(\ln b) \quad (5-4)$$

Fitting this equation by regression and forecasting the Y data works as follows: The Y variable in the regression is the log of (K - Y) in column D. The X variable is the series of integers in column C. Estimated values of the logs of (K-Y) are computed using the regression coefficients in column F. Column G takes antilogs of column F. Finally, in column H, forecasts are computed by subtracting column G from the saturation level.

Figure 5-9

	A	B	C	D	E	F	G	H	I
1	MODEXP			TV penetration			INTERMEDIATE CALCULATIONS:		
2	Modified exponential regression on time							RMSE (Square root of MSE)	1.69
3	$Y_t = K - ab^t$	or	$\ln(K - Y_t) = \ln(a) + t(\ln(b))$				3 X RMSE	5.07	
4	Select Tools, Data Analysis, Regression. Enter ranges as follows:								
5	Y range includes column D, X range includes column C, Output range = O1.							Warm-up SSE	111.20
6	Ensure number of data selected for Y and X ranges agree with the							Warm-up MSE	2.85
7	nbr. of warm-up data entered in cell C10.							Forecasting SSE	111.20
8							Fcst SSE - Warm-up SSE	-	
9	INPUT:						Nbr. of forecast periods	-	
10	Nbr. of warm-up data	39					Forecasting MSE	-	
11	Last period to forecast	40							
12	Saturation level (K)	98.21					Warm-up Sum abs. err.	41.82	
13							Warm-up MAD	1.07	
14	OUTPUT:						Forecasting Sum abs. err.	41.82	
15	Number of data	39					Fcst Sum abs. - Warm-up Sum abs.	-	
16	Nbr. of outliers	0					Nbr. of forecast periods	-	
17	Warm-up MSE	2.85					Forecasting MAD	-	
18	Forecasting MSE	-					Last forecast at period	40	
19	Warm-up MAD	1.07							
20	Forecasting MAD	-							
21									
22			Regression ranges						
23			X-range	Y range					
24	Mon/yr.	Data	t	LN (K-Y)	K-Y	Est LN K-Y	Est K-Y	Est Y	e = Y - Est Y
25	1950	9.0	1	4.49	89.21	4.48	88.14	10.07	-1.07
26	1951	23.5	2	4.31	74.71	4.29	73.20	25.01	-1.51
27	1952	34.2	3	4.16	64.01	4.11	60.78	37.43	-3.23
28	1953	44.7	4	3.98	53.51	3.92	50.48	47.73	-3.03
29	1954	55.7	5	3.75	42.51	3.74	41.92	56.29	-0.59
30	1955	64.5	6	3.52	33.71	3.55	34.81	63.40	1.10
31	1956	71.8	7	3.27	26.41	3.36	28.91	69.30	2.50
32	1957	78.6	8	2.98	19.61	3.18	24.00	74.21	4.39
33	1958	83.2	9	2.71	15.01	2.99	19.93	78.28	4.92
34	1959	85.9	10	2.51	12.31	2.81	16.55	81.66	4.24
35	1960	87.1	11	2.41	11.11	2.62	13.75	84.46	2.64
36	1961	88.8	12	2.24	9.41	2.44	11.42	86.79	2.01
37	1962	90.0	13	2.11	8.21	2.25	9.48	88.73	1.27
38	1963	91.3	14	1.93	6.91	2.06	7.87	90.34	0.96
39	1964	92.3	15	1.78	5.91	1.88	6.54	91.67	0.63
40	1965	92.6	16	1.72	5.61	1.69	5.43	92.78	-0.18
41	1966	93.0	17	1.65	5.21	1.51	4.51	93.70	-0.70
42	1967	93.6	18	1.53	4.61	1.32	3.74	94.47	-0.87
43	1968	94.6	19	1.28	3.61	1.13	3.11	95.10	-0.50
44	1969	95.0	20	1.17	3.21	0.95	2.58	95.63	-0.63
45	1970	95.3	21	1.07	2.91	0.76	2.14	96.07	-0.77
46	1971	95.5	22	1.00	2.71	0.58	1.78	96.43	-0.93
47	1972	95.8	23	0.88	2.41	0.39	1.48	96.73	-0.93
48	1973	96.4	24	0.59	1.81	0.21	1.23	96.98	-0.58
49	1974	96.9	25	0.27	1.31	0.02	1.02	97.19	-0.29
50	1975	97.1	26	0.10	1.11	-0.17	0.85	97.36	-0.26

Figure 5-10





It is important to understand that the modified-exponential curve is the best-fitting curve, given a prespecified saturation level. Thus an infinite number of curves exists, one for each saturation level. If you do not have marketing information that suggests a reasonable saturation level, experiment with entries in cell C12 until you find the saturation level that minimizes the warm-up MSE. This was done with the TV data and the best saturation level found to be 98.21.

Because the saturation level is critical, it is always wise to do simulated forecasting to test the validity of the model. To demonstrate, set the warm-up data to 7 periods (covering the years 1950 - 1956) and run the regression again. The forecasts are not significantly different from those produced using all the data. You may protest the simulation on the grounds that in 1956 there was no way to know that 98.21 was the minimum-MSE saturation level. However, the point of this example is that forecasts for the next few time periods are often insensitive to the saturation level as long as it is somewhere in the ballpark. It seems obvious from the early growth in TV data that saturation will occur somewhere in the high 90s. Any saturation level in the range 95 to 100 produces reasonable forecasts for 1957 - 1960.

## 5.5 Logistic growth (GROWTH)

Perhaps the most widely-used long-range forecasting model is the logistic curve, one of a family of S-shaped curves. In the example in Figures 5-11 and 5-12, the Y data are DVD penetration, the percentage of households that own one or more DVDs, in Fort Bend County, Texas. The curve is typical, with slow growth at first, followed by a period of steep growth, and finally by a gradual decline in growth to the saturation level. This type of forecasting model originated in the study of the spread of epidemics during the early part of this century. Today the logistic is used for the same purpose as well as in forecasting the development of markets for both industrial and consumer goods.

Like the modified exponential, the expected saturation level **K** must be entered before running the model. The equation for the logistic curve is:

$$Y_t = K / (1 + a (\exp (-bt))) \quad (5-5)$$

After taking logs and rearranging, the logistic can be written as:

$$\ln ((K/Y_t)-1) = \ln (a) - bt \quad (5-6)$$

The Y variable in the regression is the log of (K/Y)-1, column D in Figure 5-11. The X is column C. The remainder of the calculations are similar to the exponential.

The same observations about choosing the saturation level for the modified exponential also apply to the logistic. There are an infinite number of logistic curves for each saturation level. Always do some sensitivity analysis before choosing a final model.

**Figure 5-11**

	A	B	C	D	E	F	G	H	I
1	<b>LOG</b>			<b>DVD penetration in Fort Bend</b>			<b>INTERMEDIATE CALCULATIONS:</b>		
2	Logistic regression on time						<b>RMSE (Square root of MSE)</b>		0.83
3	$Y = K / [1 + a \exp(-bt)]$	or	$\ln((K/Y)-1) = \ln(a) - bt$				<b>3 X RMSE</b>		2.49
4	Select Tools, Data Analysis, Regression. Enter ranges as follows:								
5	Y range includes column D, X range includes column C, Output range = O1.						<b>Warm-up SSE</b>		4.14
6	Ensure number of data selected for Y and X ranges agree with the						<b>Warm-up MSE</b>		0.69
7	nbr. of warm-up data entered in cell C10.						<b>Forecasting SSE</b>		4.14
8							<b>Fcst SSE - Warm-up SSE</b>		-
9	<b>INPUT:</b>						<b>Nbr. of forecast periods</b>		-
10	<b>Nbr. of warm-up data</b>	<b>6</b>					<b>Forecasting MSE</b>		-
11	<b>Last period to forecast</b>	<b>16</b>							
12	<b>Saturation level (K)</b>	<b>70.00</b>					<b>Warm-up Sum abs. err.</b>		3.64
13							<b>Warm-up MAD</b>		0.61
14	<b>OUTPUT:</b>						<b>Forecasting Sum abs. err.</b>		3.64
15	<b>Number of data</b>	6					<b>Fcst Sum abs. - Warm-up Sum ab</b>		-
16	<b>Nbr. of outliers</b>	0					<b>Nbr. of forecast periods</b>		-
17	<b>Warm-up MSE</b>	0.69					<b>Forecasting MAD</b>		-
18	<b>Forecasting MSE</b>	0.00					<b>Last forecast at period</b>		16
19	<b>Warm-up MAD</b>	0.61							
20	<b>Forecasting MAD</b>	0.00							
21									
22			<b>Regression ranges</b>						
23			<b>X-range</b>	<b>Y range</b>					
24	<b>Mon/yr.</b>	<b>Data</b>	<b>t</b>	<b>LN (K/Y)-1</b>	<b>(K/Y)-1</b>	<b>LN (K/Y)-1</b>	<b>Est (K/Y)-1</b>	<b>Est Y</b>	<b>e = Y - Est Y</b>
25	1996	2.1	1	3.48	32.33	3.39	29.74	2.28	-0.18
26	1997	5.5	2	2.46	11.73	2.55	12.82	5.06	0.44
27	1998	10.6	3	1.72	5.60	1.71	5.53	10.72	-0.12
28	1999	20.8	4	0.86	2.37	0.87	2.38	20.68	0.12
29	2000	36.0	5	-0.06	0.94	0.03	1.03	34.51	1.49
30	2001	47.2	6	-0.73	0.48	-0.81	0.44	48.49	-1.29
31	2002		7	#N/A	#N/A	-1.65	0.19	58.76	#N/A
32	2003		8	#N/A	#N/A	-2.50	0.08	64.67	#N/A
33	2004		9	#N/A	#N/A	-3.34	0.04	67.60	#N/A
34	2005		10	#N/A	#N/A	-4.18	0.02	68.94	#N/A
35	2006		11	#N/A	#N/A	-5.02	0.01	69.54	#N/A
36	2007		12	#N/A	#N/A	-5.86	0.00	69.80	#N/A
37	2008		13	#N/A	#N/A	-6.70	0.00	69.91	#N/A
38	2009		14	#N/A	#N/A	-7.54	0.00	69.96	#N/A
39	2010		15	#N/A	#N/A	-8.38	0.00	69.98	#N/A
40	2011		16	#N/A	#N/A	-9.22	0.00	69.99	#N/A

Figure 5-12

