

Forecast Production with Variable Shifts, Seasonal Cycles and Growth

Sometimes forecasting production seems like it is just a step ahead of using crystal balls, Ouija boards and reading goat entrails. Complicate the problem with variable shift schedules, seasonal fluctuations and underlying growth, and it seems almost insolvable. But recognizing these facts and compensating for them make it puts the problem back into the realm of solvability.

The accompanying spreadsheet shows an analysis of such a situation. The raw data consists of dates, and the number of units produced on the dates. The shift structure is two 10-hour shifts (20 hours) on Monday, Tuesday, Wednesday and Thursday, and a 12-hour shift on Friday, Saturday and Sunday.

It's About Time

Forecasting is about looking forward in time and making a projection based on looking back in time. Time can be expressed as Excel expresses it namely, the number of days since January 1, 1900. Using this scale would produce extreme coefficients for the function that describes the events. So a scale starting with August 15, 2007 (the first date for which data is available) is used. This value is referred to as the Elapsed or E-Date.

The formula in cell C2 duplicated to the end of data is:

=A2-\$A\$2+1

Converting Apples to Oranges

Production output varies greatly from day to day primarily because of the difference in the number of hours worked in each day (20 or 12).

The first adjustment to be made is to convert the daily production values into a common scale. The logical thing to do is to convert the number of units produced per shift to units per hour.

A couple of intermediate steps have to happen first:

- Determine the day of the week
- Determine the shift
- Determine the number of hours on the shift.

To determine the day of the week, use the weekday formula in cell D2:

=WEEKDAY(A2,2)

The second argument, 2, means that Monday is the first day of the week (1) and Sunday is the seventh day of the week (7). Using this option makes the following if statement that determines the number of hours worked during the day easier to formulate.

The number of hours that the shifts on duty work during a day depends on the day of the week. This information is displayed in cell E2

=IF(D2<5,20,12)

Finally, the number of units per hour can be calculated in F2

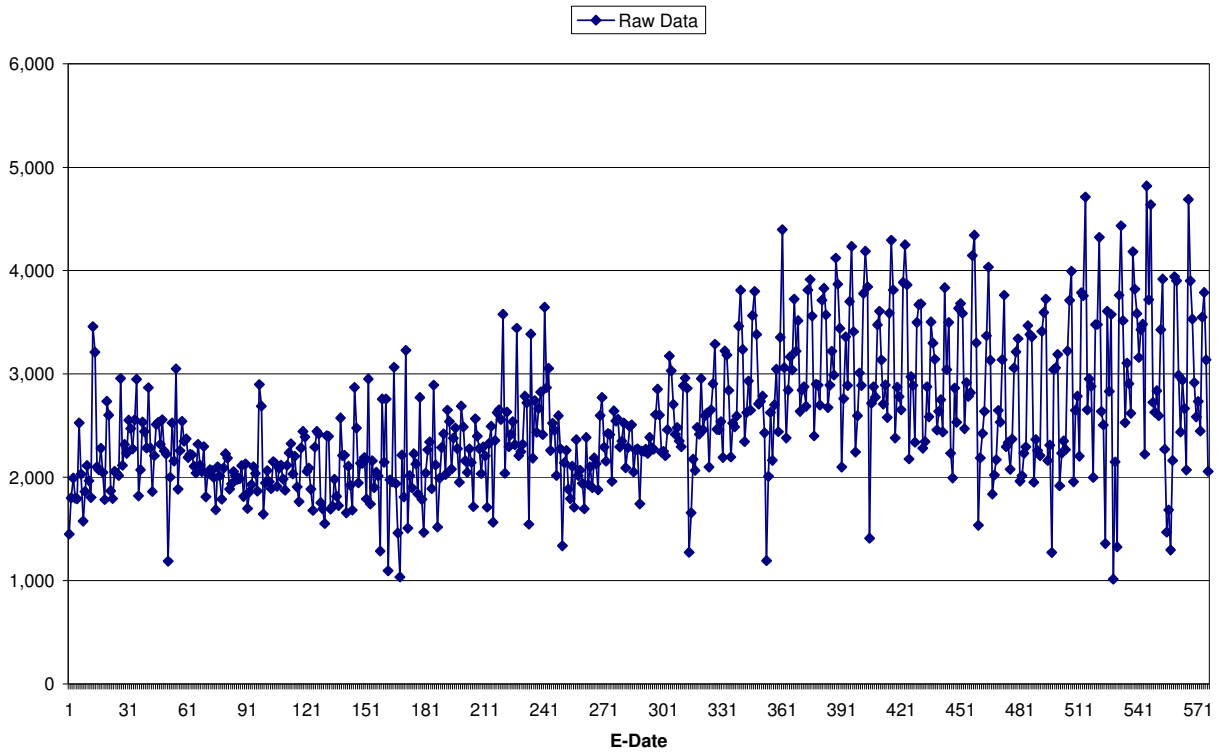
=B2/E2

The Charts

The function is in the eye of the beholder. Visual inspection will reveal what components are in it.

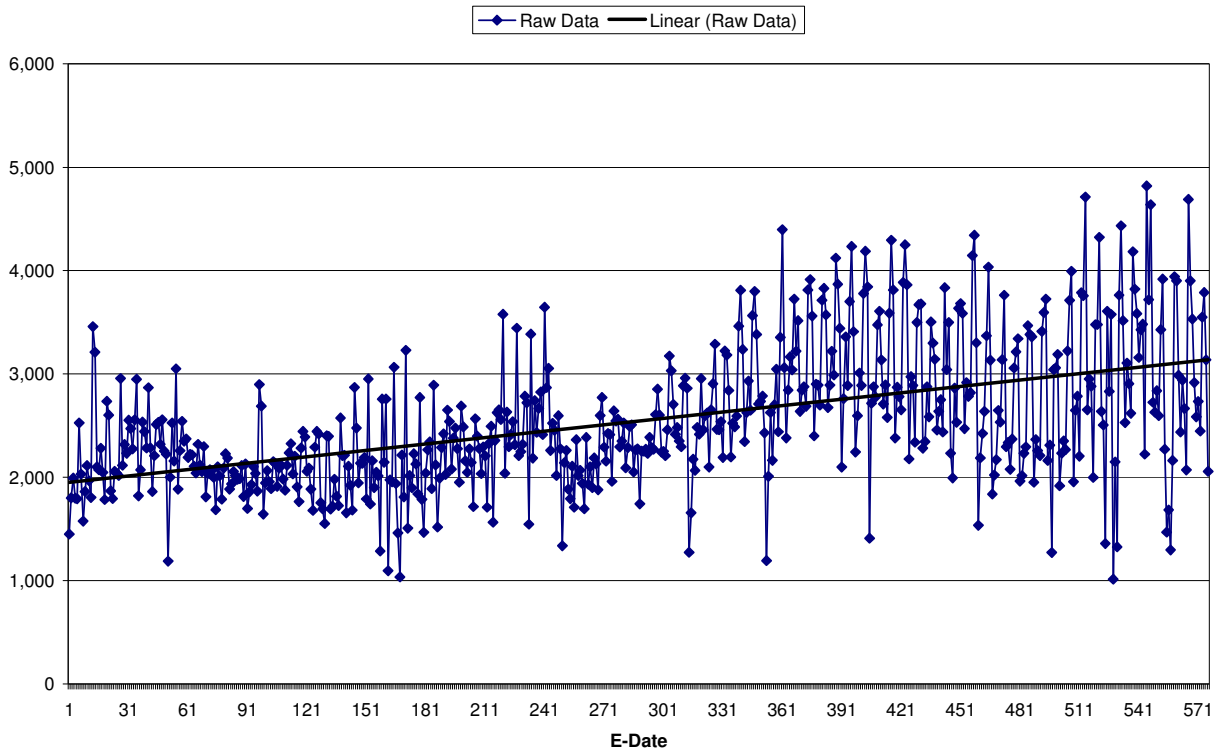
First look is at the raw data.

Productivity (Units per Hour) 2007-2009



Looking at this data, it appears that the data shifts up as time goes on. There is a growth trend. To see it better, add a trend line to the graph to see if the data “wraps around” it.

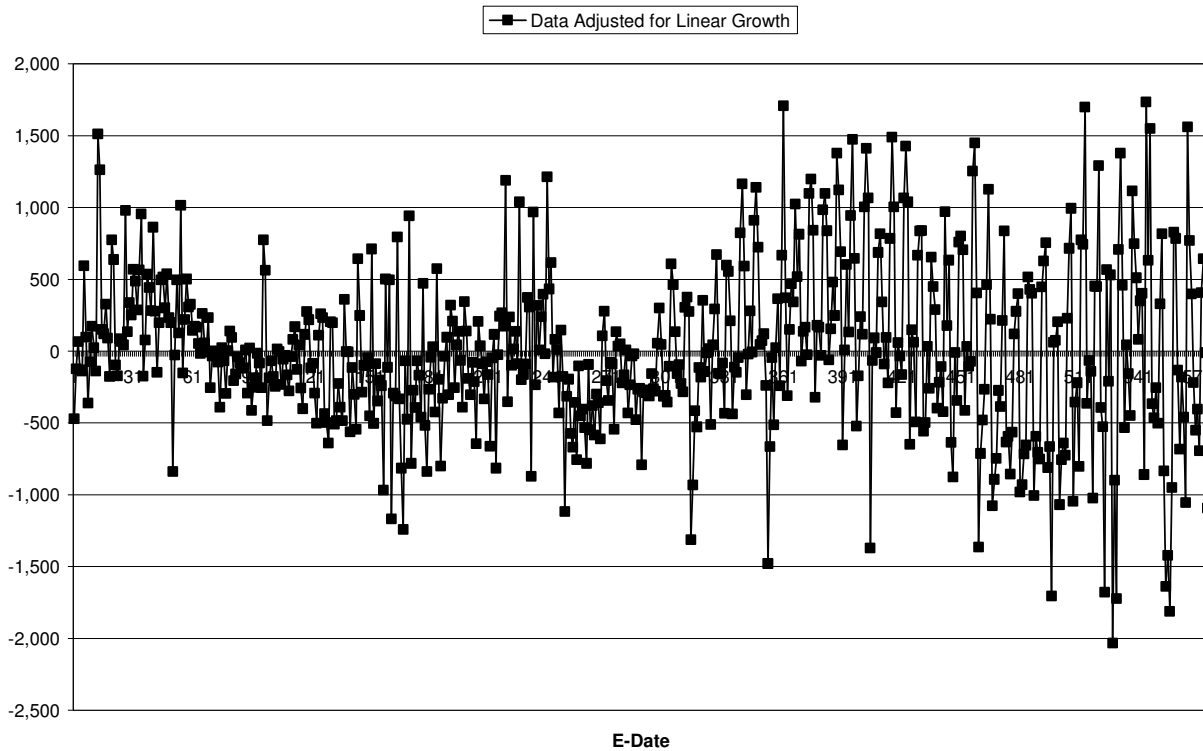
Productivity (Units per Hour) 2007-2009



This exercise confirms that part of the formula is $mx + b$. Where m is the growth rate and b is a steady-state baseline. Sometimes growth is not linear. It may be exponential or part of a larger cycle. Both of these can be approximated by a second degree or higher order equation. In this particular case, the linear fit looks good.

It is also apparent that there is a cyclical component to the function. This can be better seen if the trend line (growth + baseline) is subtracted from the raw data.

Productivity (Units per Hour) 2007-2009



One of the obvious things about this chart is that the signal has a cyclical component. It is interesting to note, that although lines usually obscure patterns in scattered data, in this case keeping them in makes seeing the cyclical component easier.

This chart reveals another factor, namely that the amplitude of the cycle is also increasing with respect to time. So in addition to the growth and steady-state component, the function can be described with: $A * \sin(Bx + C)$.

So the entire function looks like $mx + b + A * \sin(Bx + C)$

Solver

Once the model for the function exists, the parameters can be computed using solver. The solver parameters are set up in cells N1:S4

	N	O	P	Q	R	S
1		A	B	C	m	b
2		1	1	1	1	1
3						
4	Minimize	3.1241E+09				

The initial parameter settings for the Solver solution are set up in cells O2:S2 and in cell O4.

These Parameters (A, B, C, m and b) need to be applied to E-Date:

Cell G2 contains $=R\$2*C2+SS\$2= mx + b$

Cell H2 contains $=O\$2 *SIN(2*PI()*C2/PS\$2+Q\$2) = A * \sin (Bx + C)$

Cell I2 contains $=G2+H2 = mx + b + A * \sin(Bx + C)$

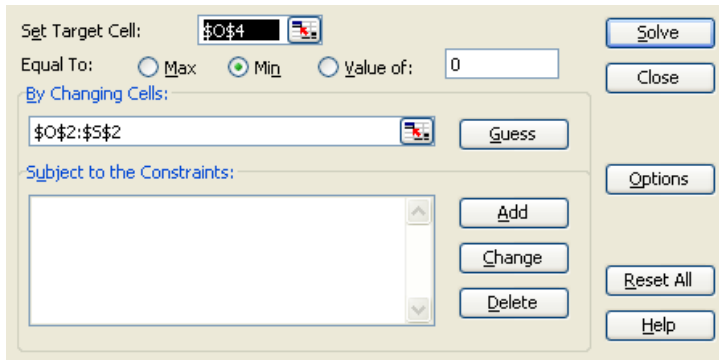
Column I contains the computed values according to the formula.

Column J contains the formula $= (F2 - I2)^2$ which is the square of the difference between the computed value and the actual value.

Which brings us back to cell N4. Since the objective is to minimize the sum of the squares of the difference to get the least squares fit, cell O4 contains the formula $= \text{SUM}(J2:J577)$

It is now time to engage solver. Note that Solver may have to be added using Tools → Addins.

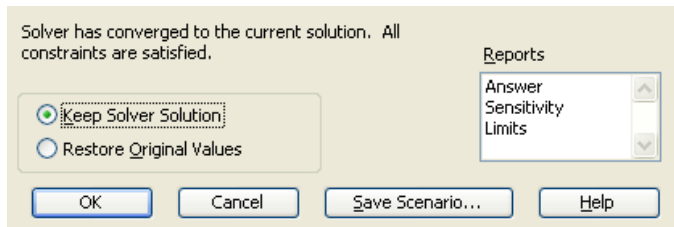
Select Tools → Solver



In the *Set Target Cell* Block, select (or type in) \$O\$4 – this is the cell to minimize. Assure that the Min radio button is selected.

With the cursor in the *By Changing Cells* block, select cells O2:S2 – these are the parameters to vary to fit the function to the data.

Click on solve.



Keep the solver solution and click OK.

Re-engage Solver several times until the parameters are no longer changing. Usually this takes only two or three iterations.

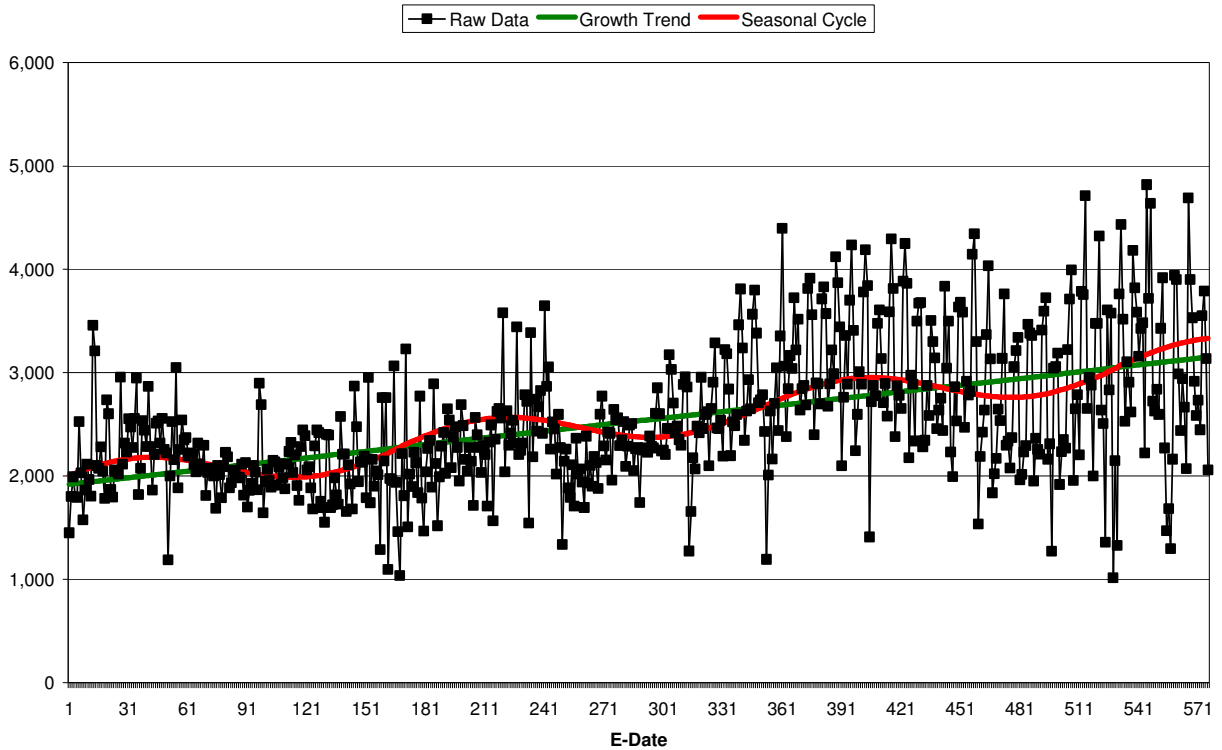
Solver should present a display like the following:

	N	O	P	Q	R	S
1		A	B	C	m	b
2		182.476112	183.0525	911.5016	2.111363	1918.061
3						
4	Minimize	1.8331E+08				

Which means the function is $2.03x + 1950 - 51.87 * \text{Sin}(2 * \text{PI}() * x / -19.26 + 16.39)$ give or take a couple of decimal points. Where x is E-Date.

A plot of raw data, trend line and seasonality all on one graph looks like:

Productivity (Units per Hour) 2007-2009



Note that the sine wave does a pretty good job cutting through the center of mass for the raw data.

Back Testing

One way to test the accuracy of a model is to see how well it predicts “history.” The Back Test tab of the spreadsheet does exactly this. To get the data needed, the units per hour is re-converted into production units per day. In other words, the model (which shows data in units per hour) is multiplied by 12 or 20 depending on the shifts on duty for the day.

It is obvious from looking at the graph that there is a significant amount of noise in this signal so that the ability to predict any specific day’s production volume is an exercise in frustration. However, according to the least squares fit methodology, the combined magnitude of the noise on either side of the function is the same: the area under both curves should be equal. In other words, the longer the period of measurement, the more likely it is that the model will fit reality – the noise will cancel itself out.

A look at the actual and predicted values for monthly periods shows that the model is fairly accurate.

	Actual	Predicted	Variance
0707	585,690	584,840	-0.15%
0708	1,181,080	1,115,467	-5.56%
0709	1,036,626	1,044,209	0.73%
0710	1,066,372	1,063,632	-0.26%
0711	954,755	954,976	0.02%
0712	970,609	1,051,837	8.37%
0801	1,088,060	1,212,371	11.43%
0802	1,115,242	1,213,757	8.83%
0803	1,197,767	1,283,847	7.19%
0804	1,121,256	1,218,390	8.66%
0805	1,184,025	1,215,195	2.63%
0806	1,357,151	1,259,567	-7.19%
0807	1,493,863	1,456,562	-2.50%
0808	1,538,734	1,467,480	-4.63%
0809	1,422,061	1,409,136	-0.91%
0810	1,418,605	1,446,121	1.94%
0811	1,202,892	1,299,745	8.05%
0812	1,330,287	1,475,422	10.91%
0901	1,464,018	1,550,668	5.92%
0902	1,431,553	1,540,984	7.64%

There are some anomalous peaks in the variance around year end 2007 and year end 2008. Beyond the plain numbers, there is a knowledge that there is a drop in production around the holiday season. An adjustment to the model could be made to incorporate this knowledge, or in a more practical sense, a manual adjustment can be made.