

Business Mathematics (BK/IBA) – Quantitative Research Methods I (EBE)

Computer tutorial 4

Introduction

In the last tutorial session, we will continue to work on using Microsoft Excel for quantitative modelling. But first, we will train ourselves in translating Excel syntax into readable mathematical formulas, as well as in designing formulas from scratch.

Making formulas

So far, we practiced retyping existing formulas. But in the end, you should be able to come up with a formula yourself. If you have modelled a business case in Excel, you must be able to write your model equations in the report, not as Excel syntax, but as real formulas. Thus, a syntax like

“=1/((1/(A1-1))+1/(B1-1))”

should then be typed as

$$C = \frac{1}{\frac{1}{A-1} + \frac{1}{B-1}}$$

where of course the choice of symbols (A, B, C) could have been made differently (e.g., x, y, z).

Create a readable formula to report the following Excel syntax:

Exercise 1. “=(EXP(-T1)-EXP(T1))/(EXP(-T1)+EXP(T1))”

Exercise 2. “=SQRT((SUMSQ(A:A))/(COUNT(A:A)-1))”

Exercise 3. “={=MMULT(B1:F5;H1:H5)-J1:J5}”

Exercise 4. “=IF(Z1<=0;-Z1^2;Z1^2)”

Some suggested answers are:

1. $x \cdot 1 = \frac{z^{\partial+2-\partial}}{z^{\partial-1-\partial}}$

2. $x \cdot 2 = \sqrt{\frac{1-u}{z^{\frac{1}{2}}}}$

3. $x \cdot 3 = \frac{c - qv}{z}$

4. $x \cdot 4 = \begin{cases} z^z & \text{if } z \leq 0 \\ z^z & \text{otherwise} \end{cases}$

1. $x \cdot 1$

2. $x \cdot 2$

3. $x \cdot 3$

4. $x \cdot 4$

Of course the choice of symbols may be different, and the expressions may be typed in a slightly different form as well. For instance, number 2 might have been typed as $z = \sqrt{\frac{1}{n-1} \sum_{j=1}^n ((a_j)^2)}$.

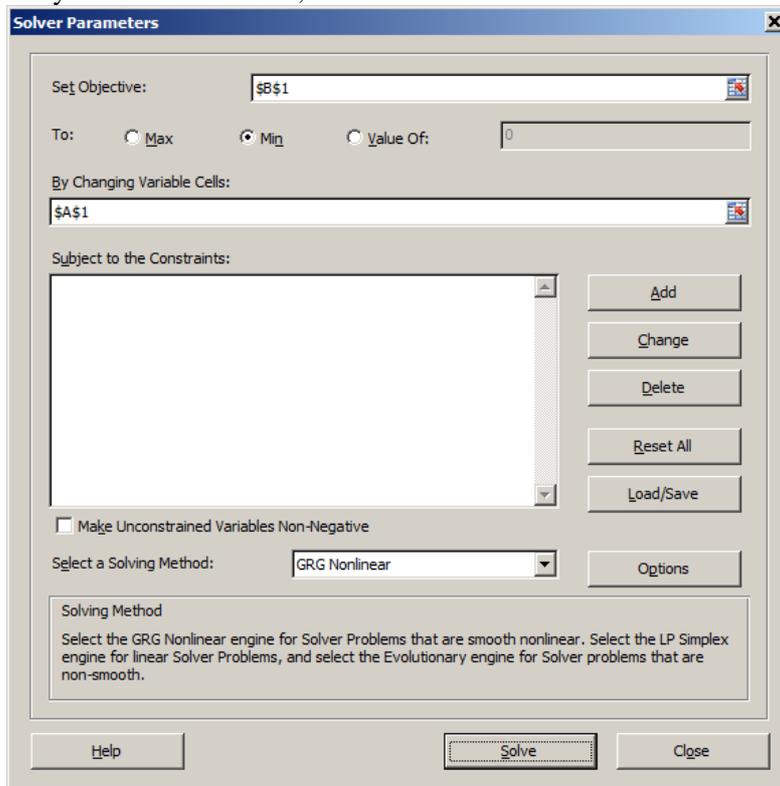
Mathematical modelling in Excel – doing more with the Solver

Last week, we discussed how to use the Solver for solving “difficult” equations. But the Solver can do much more:

- it can also find extreme values;
- it can even find constrained extreme values.

Let us first consider an easy case: $f(x) = 3x^2 - 6x + 2$. Clearly, the extreme value occurs when $f'(x) = 6x - 6 = 0$, so when $x = 1$. It is a minimum, because $f''(x) = 6 > 0$. How would Excel’s Solver approach this, assuming that you didn’t study the theory of optimization? As before, we type a trial value, say $x = 2$ in a cell, e.g., A1. Cell B1 contains the expression for $f(x)$, so “=3*A1^2-6*A1+2”.

Now, we use the Solver, but specifying that we want to minimize the function, not to equate it to a certain value. The Solver easily finds the value of 1, with a function value of -1 .



The same procedure may be used to maximize a function. More interestingly, it may also be used to minimize or maximize functions of more than one variable.

- Exercise 1. Use the Excel Solver to find the minimum of $f(x, y) = x^2 - xy + y^2 - 4y + 5$.
 Exercise 2. Use the Excel Solver to find the maximum of $f(x, y, z) = 2x - x^2 + 10y - y^2 - 2xy - 2xz$.

Answers:

1. Let x be in A1 with a trial value of 1, y in B1 with a trial value of 2, and let C1 contain the formula $=A1\sqrt{2}-A1*B1+B1\sqrt{2}-4*B1+5$. Use the Solver with objective minimize C1, changing variables $\{A1:B1\}$. Turn off the non-negativity constraints. Clicking Solve yields $(x, y) = (1.333, 2.667)$.
 2. Let x be in A2 with a trial value of 1, y in B2 with a trial value of 2, z in C2 with a trial value of 3, and let D2 contain the formula $=2*A2-A2\sqrt{2}+10*B2-B2\sqrt{2}+2*A2*B2-2*A2*C2$. Use the Solver with objective maximize D2, changing variables $\{A2:C2\}$. Turn off the non-negativity constraints. Clicking Solve yields no maximum.

Finding an extreme value with the Solver sounds easy, but it is sometimes a tricky thing. Last week we saw that finding solutions to an equation may in some cases yield a wrong answer. But for extreme values, the situation is even more delicate. Take the quadratic function $y = x^2$.

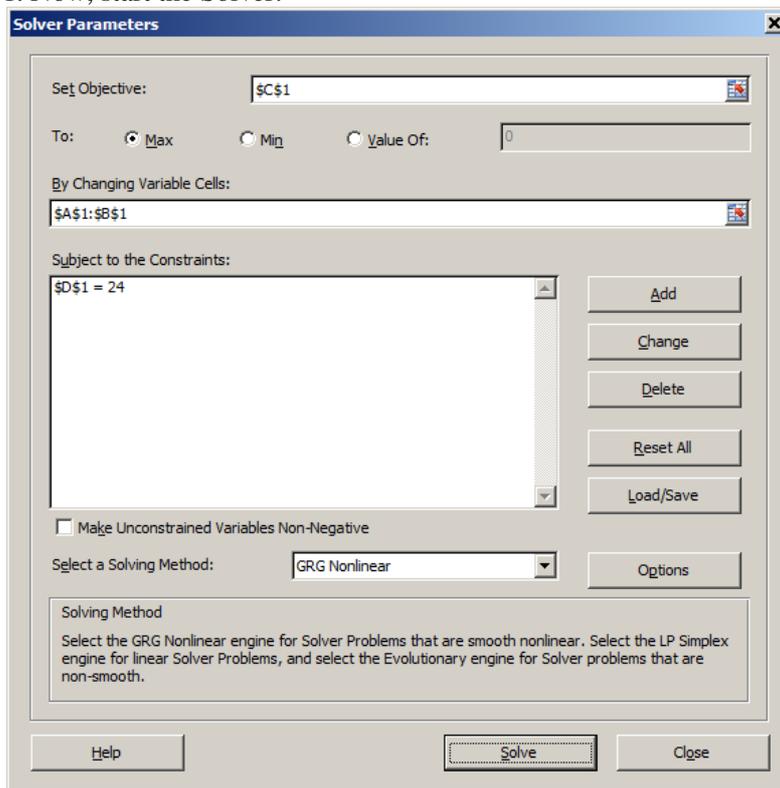
- Exercise 3. Use the Excel Solver to find the minimum of $f(x) = x^2$. This should give $x = 0$.
 Exercise 4. Now, keep $x = 0$ as initial value, and use the Excel Solver to find the maximum of $f(x) = x^2$.
 Exercise 5. Change the initial value into $x = 0.001$, and again use the Excel Solver to find the maximum of $f(x) = x^2$.

Answers:

5. Same as above. Now, the Solver does not converge.
 4. Same as above, but now let the Solver minimize the objective function. You will find $x=0$.
 3. Let x be in A3 with a trial value of 1, and let B3 contain the formula “=A3^2”. Use the Solver with objective minimize B3, changing variables “A2?”. Turn off the non-negativity constraints. Clicking Solve yields $x = 0$.

In Exercise 5, we correctly find that there is no maximum, but in Exercise 4 we find a completely wrong answer. The Solver started in the minimum, where $f'(x) = 0$, and apparently, it never was able to move out of the stationary point, and incorrectly concluded there is no way to climb higher. This again illustrates that a numerical approach to solving an equation or find an extreme value has not only benefits but also grave dangers. Of course, this has important repercussions in the real world. Suppose that you fine-tuned your business operation in the point of minimum profit instead of maximum profit, because Excel’s solver suggested so!

The Solver can also be used for finding constrained extreme values. For this, we need to use the settings “Subject to the Constraints”. Suppose we want to maximize $f(x,y) = xy$ under the constraint that $x + 3y = 24$. Type trial values in A1 for x and in B1 for y , and type the equations “=A1*B1” in C1 and “=A1+3*B1” in D1. Now, start the Solver:



Click Solve, and you will find a solution at $x = 12$ and $y = 4$.

Exercise 6. Use the Excel Solver to find the maximum of $f(K,L) = 120K^{0.3}L^{0.7}$, subject to the constraint $12K + 14L = 100$.
 Exercise 7. Use the Excel Solver to find the maximum of $f(x,y,z) = x^2 + y^2 + z^2$, subject to the constraints $xyz = 1000$ and $x + y + z \leq 400$.

Answers:

6. Let K be in A4 with a trial value of 1, L in B4 with a trial value of 2, and let C4 contain the formula $=120*A4\sqrt{0.3*B4\sqrt{0.7}}$. Further, let D2 contain the formula $=12*A4+14*B4$. Use the Solver with objective maximize C4, changing variables $A4:B4$, and add a constraint $D4<=100$. Turn off the non-negativity constraints. Clicking Solve yields $(K, L) = (2.5, 5)$.

7. Let x be in A5 with a trial value of 1, y in B5 with a trial value of 2, z in C5 with a trial value of 3, and let D5 contain the formula $=A5\sqrt{2+B5\sqrt{2+C5\sqrt{2}}}$. Further, let E5 contain the formula $=A5*B5*C5$ and F5 the formula $=A5+B5+C5$. Use the Solver with objective maximize D5, changing variables $A5:C5$, and add two constraints: $E5=1000$ and $F5<=100$. Turn off the non-negativity constraints. Clicking Solve yields $(x, y, z) = (1.59, 1.59, 396.82)$. Observe that the problem is completely symmetrical: the answer $(396.82, 1.59, 1.59)$ is also OK, and it may be found with a different choice of trial values.

One final note on the Solver: there is a checkbox for “Make Unconstrained Variables Non-Negative”. This option adds for all decision variables (say, a , b , etc.) the constraint $a \geq 0$, $b \geq 0$, etc, unless there is an explicit constraint for these variables. This option is turned on by default! We recommend you to turn it off. Of course, in business applications, many decision variables are non-negative (the number of employees, the number of products sold), but some decision variables can be negative (the change of your inventory, for instance).

Mathematical modelling in Excel – more advanced modelling

Finally, we will explore a number of issues that are useful for more advanced modelling. The first is the use of auxiliary columns and/or variables. Suppose we have a data matrix with floor areas in column A and house prices in column B. Now we wish to find the regression line which best fits the data. Denoting the independent variable (floor area) with x and the dependent variable (price) with y , we seek the line

$$y = ax + b$$

with coefficient a and b such that sum of squares of deviations is minimal. As discussed at the lectures, the least-squares approach yields the following expressions:

$$a = \frac{\frac{1}{n} \sum x \sum y - \sum xy}{\frac{1}{n} \sum x \sum x - \sum x^2}$$

and

$$b = \frac{\sum y - a \sum x}{n}$$

We will use Excel to implement the least-squares method.

First we type in the data. To speed this up, the data has already been typed by us; you can find them on BlackBoard (together with the present tutorial document).

	A	B
1	Floor area (X)	Price (Y)
2	142	649000
3	80	125000
4	117	599000
5	46	145000
6	110	249000
7	95	259000
8	164	1195000
9	80	187500
10	188	750000
11	89	215000
12	103	265000
13	244	1750000
14	148	315000
15	80	210000
16	145	1495000

Now, without using auxiliary variables, we would need to type something like “ $= (1/\text{COUNT}(A:A) * \text{SUM}(A:A) * \text{SUM}(B:B) - \text{SUMPRODUCT}(A:A;B:B)) / (1/\text{COUNT}(A:A) * \text{SUM}(A:A) * \text{SUM}(A:A) - \text{SUMSQ}(A:A))$ ” to get the coefficient a .¹

Actually, we will implement these formulas in the spreadsheet, in four alternative ways. Try all approaches. This seems a bit redundant, but in real problems you sometimes have just one approach available, and in the present exercise the results provide cross checks.

Exercise 8. In the first implementation, we indeed make a tedious formula for a in C2 and for b in D2, directly, without using any auxiliary variables.

Exercise 9. The second implementation is done by defining five auxiliary variables: n in cell C2, $\sum x$ in cell D2, $\sum y$ in cell E2, $\sum xy$ in cell F2, and $\sum x^2$ in cell G2. These cells are then combined in the much simpler formulas for a and b .

Exercise 10. A third approach is similar to approach 2, but without using the functions SUMPRODUCT and SUMSQ. To do this, you need to make extra columns, say C containing $x_i y_i$ and D containing x_i^2 .

Exercise 11. The final method computes a sum of squares and uses the Solver to find values for a and b such that the sum of squares of observed prices and predicted price is a minimum. For instance, you make a new column containing $\hat{y}_i = ax_i + b$, a column containing $(\hat{y}_i - y_i)^2$, and you use the Solver to find the values of a and b that minimize $\sum_i (\hat{y}_i - y_i)^2$.

Answers:

¹ Note the use of the functions SUMSQ and SUMPRODUCT. See Excel’s Help if you don’t understand them.

8. Let a be in $C1$ as $=(1/COUNT(A:A)*SUM(B:B)*SUM(A:A)-SUMPRODUCT(B:B;A:A)/(1/COUNT(B:B)*SUM(B:B)-SUMSQ(B:B)))$ and b in $D1$ in as $=(1/COUNT(A:A)*SUM(B:B)-SUMSQ(B:B))$. This yields $a = 6151$ and $b = -264749$.

9. Now let n be in $C1$ as $=(COUNT(A:A); \sum x$ in $D1$ as $=(SUM(B:B); \sum y$ in $E1$ as $=(SUM(A:A); \sum xy$ in $F1$ as $=(SUMPRODUCT(B:B;A:A); \sum x^2$ in $G1$ as $=(SUMSQ(B:B))$. Next define a in $H1$ as $=(1/C1*(D1*E1-F1)/(1/C1*D1^2-G1))$ and b in $I1$ as $=(E1-H1*D1)/C1$. This yields the same a and b as in the previous question.

10. Define $C2$ as $=(B2*A2)$ and copy this until $D72$. Likewise, $D2$: $=(B2^2)$ until $D72$. Now define $\sum x$ in $E1$ as $=(SUM(B:B); \sum y$ in $F1$ as $=(SUM(A:A); \sum xy$ in $G1$ as $=(SUM(C:C); \sum x^2$ in $H1$ as $=(SUM(D:D))$. Further, n in $I1$ is $=(1/I1*(F1-G1)/(1/I1*E1^2-H1))$ and b in $K1$ is $=(F1-J1*E1)/I1$. The values are again as before.

11. Start with trial value for a (1) in $C1$ and b (2) in $D1$. Next define y in $C2$ as $=(C1*B2+D1)$ and copy to $C72$. Also define $y - y$ in $D2$ as $=(C2-A2)$ and copy downward, and $(y - y)^2$ in $E2$ as $=(D2^2)$ and copy downward. $\sum (y - y)^2$ is in $F1$. Start the Solver with objective function $F1$, to be minimized and $C1:C2$ as decision variables. Click Solve, and you will not find the right solution! However, run the Solver with the same settings again, and you will now obtain an improved solution, which is (almost) the same as those obtained before.

The great advantage of the last option is that it is very flexible. Suppose that you were wondering why we use least squares, and not least absolute values. The formulas for a and b are known for minimizing the least squares sum, but not for minimizing the absolute sum. The Solver can, however, easily find the coefficients a and b

Exercise 12. Make in one spreadsheet the OLS estimator of a and b and the estimator of a and b on the basis of minimizing $\sum |\hat{y}_i - y_i|$.

Exercise 13. Make one graph with the data, the OLS regression line, and the regression line on the basis of the absolute sum of squares.

Answers:

13. You need to define a scatterplot with three series. Series 1 ("data") has the empirical x and y values. Series 2 ("OLS") has the same x values, but \hat{y} according to OLS. Series 3 has the alternative \hat{y} according to the absolute deviation.

12. Much is similar compared to 12. But we also create $|y - y|$ in $G2$, we use $=(ABS(D2))$. Notice that it is wise to reserves cells for two versions of a and b : one with OLS and one with the absolute value.

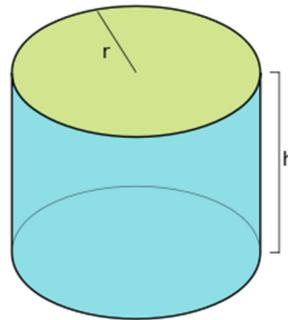
More sophisticated models can also be easily addressed with the Solver.

Exercise 14. Add the OLS estimate for the quadratic model $y = ax^2 + bx + c$.

Answers:

14. Now, we have an extra variable c (in E1), and define y in column C as $=C\$1*B2\sqrt{2}+D\$1*B2+E\$1$, and onward to C72. The Solver yields $a = 43699$, $b = 1.289$, and $c = 1.002$.

Now, we will take up another problem. Suppose you are a producer of canned meat, and you wish to use the cheapest tin cans, so tin cans with an “optimal” shape. In this problem, we will assume that the can’s form is a cylinder with radius r and height h :



In this assignment, you need some facts that you may have forgotten, such as the area of circle (πr^2) and its circumference ($2\pi r$).

Exercise 15. Develop a spreadsheet in which you model this problem, and use the Solver to find the values of r and h such that the amount of tin is a minimum, while allowing the tin can to have a volume of 1 liter.

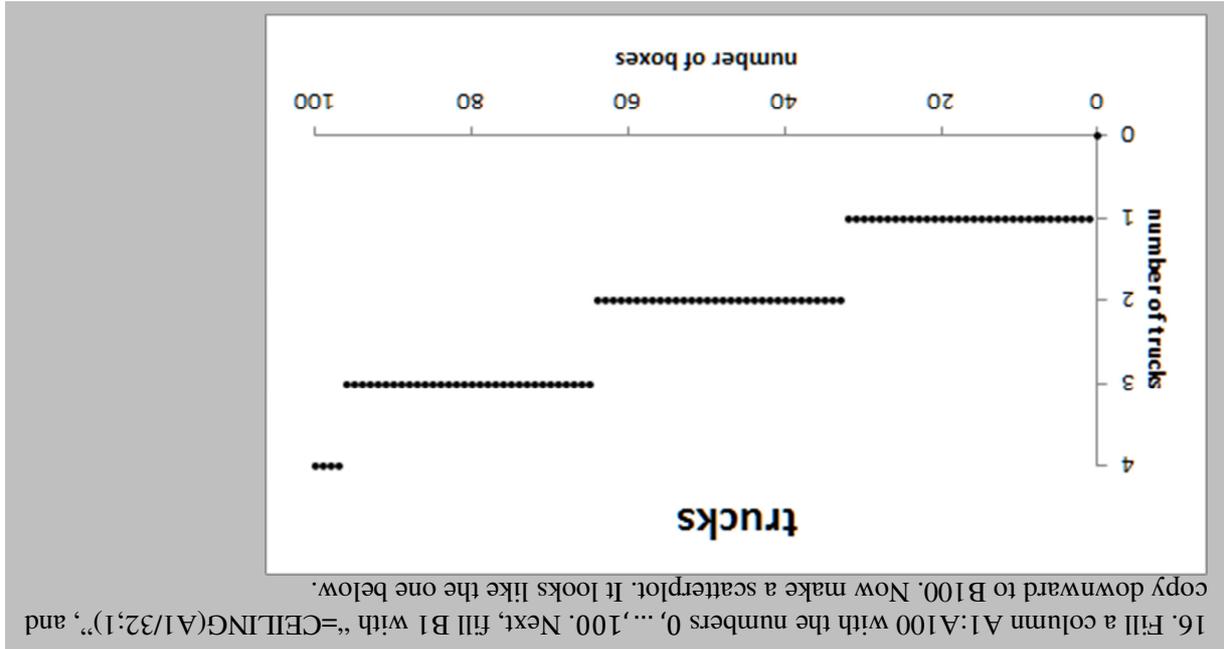
Answers:

15. Put r in A1 and h in B1. To facilitate coding, it may help to assign the names “rad” and “hgh” respectively. Further define area (“area”) in C1 as “=2*PI()*rad²+2*PI()*rad*hgh” and volume (“vol”) in D1 as “=PI()*rad²*hgh”. Use the Solver to set the objective “vol” equal to “1”, by changing “rad;hgh”. You may keep the non-negativity condition turned on (rad and hgh can never be negative). Click Solve, and you find $r = 0.598$ and $h = 0.889$.
 [Why these formulas in C1 and D1? The area of tin is the area of circle (πr^2), multiplied by 2 because there are two circles, plus the area of a rectangle of length $2\pi r$ and height h . The volume is the area of the circle (πr^2) multiplied by the height (h).]

In real-world problems related to logistics, we often encounter non-smooth variables. Suppose that you want to transport carton boxes with cargo, and that a truck has a capacity of 32 of such boxes. Excel has a variety of functions to deal with integer functions; see ROUND, TRUNC, FLOOR and CEILING.

Exercise 16. Develop a spreadsheet in which you calculate the amount of trucks needed to transport a supply of n boxes, with $n = 0,1, \dots, 100$. Plot the results in a graph.

Answers:



Modelling business processes and creating economic models is an important element in daily practice of all analysts. Sometimes, you use specialized software (e.g., for problems in logistics), or you use more advanced tools than Excel (e.g., Matlab). But Excel is also widely used, by professionals and scientists alike. The famous paper *Growth in a Time of Debt*, by Carmen Reinhart and Kenneth Rogoff, in the top level *American Economic Review* in 2010, was based on analysis in Excel. Unfortunately, the authors had not taken today's course, and the work contained Excel coding errors. Let's take this lesson. Use \$-signs carefully, use intermediate variables, build checks into your spreadsheet, and compare all results with common sense. Good luck!