

## Correlaties in Crystal Ball modellen 2

Memo van André Koch (Stachanov) voor de Dienst Landelijke Gebieden in opdracht van André Bijl – Weisz.  
Oktober 2010

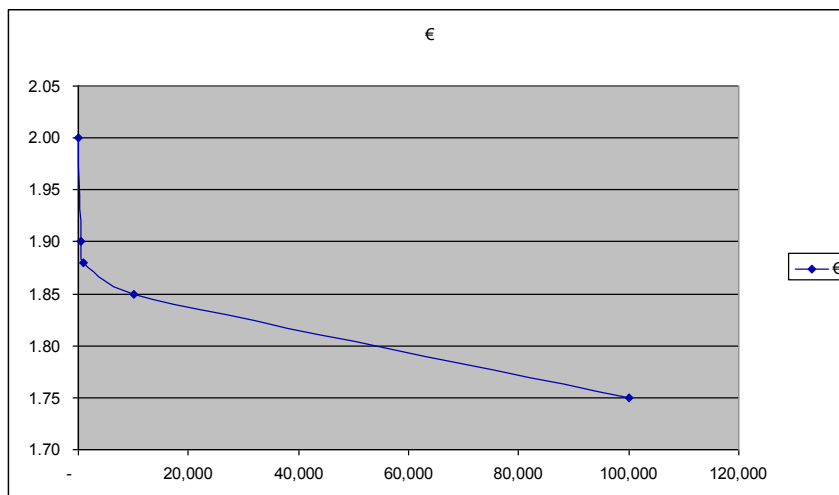
Een fout die vaak gemaakt wordt is dat men een causaal verband zoals *prijs X hoeveelheid = omzet*, modelleert als een zeer hoge correlatie tussen *prijs* en *hoeveelheid*. Modelmatig is dit verkeerd omdat men deze relatie in Excel moet definiëren en niet moet proberen te pakken met correlaties.

André Bijl heeft me het volgende voorbeeld aangereikt:

M3	€
100	2.00
500	1.90
1,000	1.88
10,000	1.85
100,000	1.75

Het voorbeeld laat duidelijk een relatie zien tussen hoeveelheid en prijs. Hoe groter de hoeveelheid, hoe lager de prijs.

Nu is de verleiding groot om dit te modelleren aan de hand van de correlatie. Met de Excel formule =CORREL(Range1,Range2) vind je deze correlatie: 82%. Het gaat in dit geval om de *Pearson* correlatie. Zou men de *Spearman* correlatie nemen, en dat is de correlatie zoals Crystal Ball deze uitrekent, dan krijgt men een correlatie van  $-1^1$ . Zoals bekend heeft het geen zin een correlatie van min of plus één te modelleren omdat de twee variabelen zich in feite als één dezelfde variabele gedragen.



<sup>1</sup> In de modellen van DLG komen vaker correlaties van 1 of -1 voor. Kan het zijn dat dit is aangereikt door de Spearman correlaties die in dit soort gevallen immers vaak op 1 uitkomt?

Principieel is het ook fout: je weet immers dat de relatie tussen prijs en hoeveelheid er is en men zou dit in Excel kunnen modelleren. De reden waarom men naar de correlatie grijpt is omdat dit eenvoudig lijkt. De correcte methode is op zoek te gaan naar de relatie tussen Hoeveelheid (de X-variabele) en Prijs (de Y-variabele). Het zoeken naar de relaties heeft regressie-analyse.

De eenvoudigste regressie-analyse is de lineaire regressie waarbij de grafiek een rechte lijn is. Je kan deze lijn vinden door twee punten te nemen en hierbij de delta(x) en delta(y) te bereiken en de richtingscoëfficiënt te bereiken. De intercept, of het snijpunt met de Y-as kan men vinden door voor X nul aan te houden. Het snelst werkt echter de Regression functie in Excel onder het Data Analysis package.

Excel geeft -0.0000017 als richtingscoëfficiënt en als intercept met de Y-as 1.91.

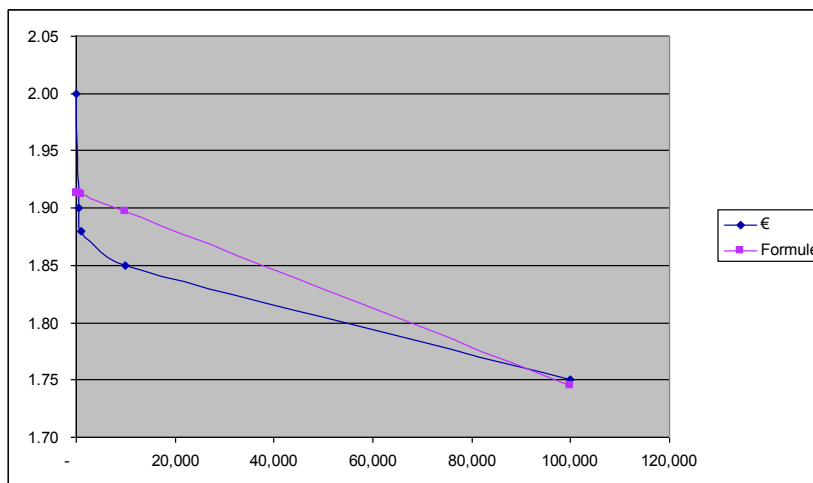
De formule van het formaat  $y = ax + b$  kan dan worden ingevuld als

$$y = -0.0000017x + 1.91$$

Deze formule “voorspelt” als het ware de prijs als ik de hoeveelheid weet.

M3	€	Lineaire regressie
100	2.00	1.91
500	1.90	1.91
1,000	1.88	1.91
10,000	1.85	1.90
100,000	1.75	1.75

Zet men dit grafisch uit dan is te zien dat de formule redelijk klopt.



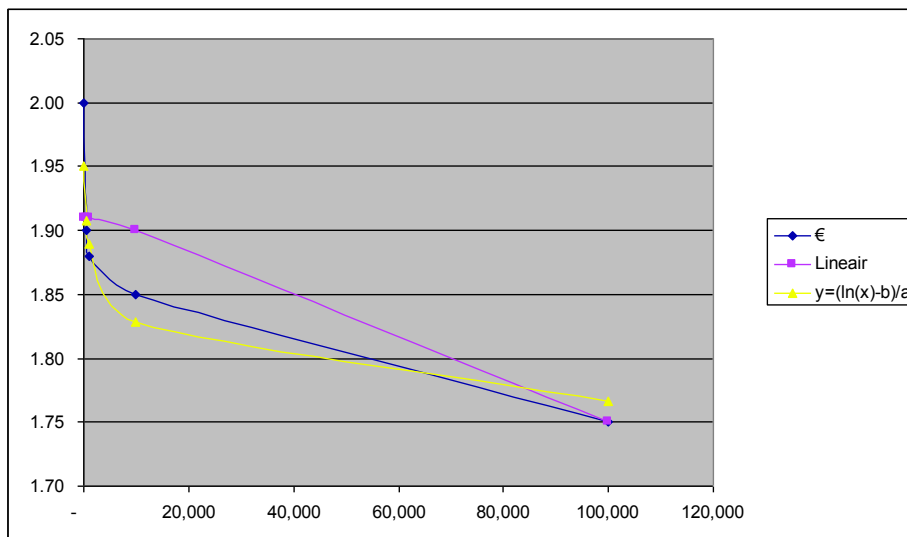
Het is desalniettemin duidelijk dat het verband complexer is dan een lineaire functie. Wil men dit preciezer definiëren dat moet dan grijpen naar niet-lineaire regressie methodes die helaas niet in Excel zijn opgenomen.

Voor de liefhebber: De volgende niet-lineaire functie komt in de buurt van hetgeen we hier zien

$$y = (\ln(x) - b) / a$$

met  $a = -37,63$  en  $b = 78$

M3	€	Lineaire regressie	Niet-Lineaire regressie
100	2.00	1.91	1.95
500	1.90	1.91	1.91
1,000	1.88	1.91	1.89
10,000	1.85	1.90	1.83
100,000	1.75	1.75	1.77



#### Conclusie:

Het is principieel beter om oorzakelijke verbanden ook als een formule in Excel te modelleren en hiervoor niet de correlatie te gebruiken. Deze formule kan worden gevonden door middel van lineaire regressie. Als men dieper in de materie duikt geven niet-lineaire functie vaak nog betere oplossingen.

André Koch  
 Stachanov Solutions & Services BV  
 020-5091010  
[andre@stachanov.com](mailto:andre@stachanov.com)