Extinderi ale regresiei cu puncte multiple

George Daniel Mateescu

Abstract: In a previous article, we introduced the technique of multiple regression points. This technique is an extension of regression methods in the case of multi valued functions. Specifically, we study the time series for which a value is an interval. This is the case, for example, exchange rate values in one day, which varies between a minimum and a maximum. Also, we can study the associated data series stock indicators, which varies continuously throughout the day. In this paper we present the extension of this technique to non-linear regression.

Classification-JEL: C02, C32

Keywords: multiple points regression

Introducere

In regresia clasica se presupune ca un set de puncte din plan, de forma $(x_i, y_i), i = 1..n$ apartin graficului unei functii, de exemplu, o variabila *y* care depinde de alta variabila, *x*, printr-o lege de corespondenta (functionala) necunoscuta.

Spre deosebire de regresia clasica, in (Mateescu, 2013) am definit o regresie cu puncte multiple, aceasta notiune fiind definita pornind de la un set de valori observate de forma $(x_i, [u_i, v_i]), i = 1..m$. Adica, pentru o valoare a variabilei independente, pot corespunde mai multe valori observate, chiar un interval de valori. Acesta este cazul seriilor de timp cu un anumit pas, zile, luni, in care valorile observate variaza intr-un interval, asa cum este cazul cursului de schimb pe parcursul unei zile, de asemenea cum este cazul indicatorilor bursieri care de asemenea variaza continuu pe parcursul unei zile, intre o valoare minima si o valoare maxima, zilnica.

Am definit regresia liniara pornind de la functia de distanta, de forma ([1]):

$$\varphi(a,b)=\sum_{i=1}^{m}\int_{u_i}^{v_i}\left(ax_i+b-t\right)^2 dt$$

Iar valorile coeficientilor *a* si *b* se obtin prin:

$$\min\varphi(a,b)$$

Adica:

$$a=\frac{\left(\frac{1}{2}\sum_{i=1}^{m}x_i\left(v_i^2-u_i^2\right)\right)\cdot\left(\sum_{i=1}^{m}\left(v_i-u_i\right)\right)-\left(\frac{1}{2}\sum_{i=1}^{m}\left(v_i^2-u_i^2\right)\right)\cdot\left(\sum_{i=1}^{m}\left(v_i-u_i\right)x_i\right)}{\Delta_2}$$

$$b=\frac{\left(\sum_{i=1}^{m}\left(v_i-u_i\right)x_i^2\right)\cdot\left(\frac{1}{2}\sum_{i=1}^{m}\left(v_i^2-u_i^2\right)\right)-\left(\sum_{j=1}^{n_i}\left(v_i-u_i\right)x_i\right)\cdot\left(\frac{1}{2}\sum_{i=1}^{m}x_i\left(v_i^2-u_i^2\right)\right)}{\Delta_2}$$

In continuare, vom extinde definitia noastra, introducand o expresie polinomiala de forma

$$ax_i^2+bx_i+c$$

precum si functia de distanta

$$\varphi(a,b,c)=\sum_{i=1}^{m}\int_{u_i}^{v_i}\left(ax_i^2+bx_i+c-t\right)^2 dt$$

pe care o minimzam, adica determinam valorile coeficientilor a, b, c care satisfac sistemul de ecuatii:

$$\frac{\partial\varphi}{\partial a}(a,b,c)=\sum_{i=1}^{m}\int_{u_i}^{v_i}2x_i^2\left(ax_i^2+bx_i+c-t\right)dt=0$$

$$\frac{\partial\varphi}{\partial b}(a,b,c)=\sum_{i=1}^{m}\int_{u_i}^{v_i}2x_i\left(ax_i^2+bx_i+c-t\right)dt=0$$

$$\frac{\partial\varphi}{\partial c}(a,b,c)=\sum_{i=1}^{m}\int_{u_i}^{v_i}2\left(ax_i^2+bx_i+c-t\right)dt=0$$

de unde rezulta:

$$\sum_{i=1}^{m}\left(ax_i^4\left(v_i-u_i\right)+bx_i^3\left(v_i-u_i\right)+cx_i^2\left(v_i-u_i\right)-x_i^2\frac{v_i^2-u_i^2}{2}\right)=0$$

$$\sum_{i=1}^{m}\left(ax_i^3\left(v_i-u_i\right)+bx_i^2\left(v_i-u_i\right)+cx_i\left(v_i-u_i\right)-x_i\frac{v_i^2-u_i^2}{2}\right)=0$$

$$\sum_{i=1}^{m}\left(ax_i^2\left(v_i-u_i\right)+bx_i\left(v_i-u_i\right)+c\left(v_i-u_i\right)-\frac{v_i^2-u_i^2}{2}\right)=0$$

si in continuare

$$a\sum_{i=1}^{m}x_i^4\left(v_i-u_i\right)+b\sum_{i=1}^{m}x_i^3\left(v_i-u_i\right)+c\sum_{i=1}^{m}x_i^2\left(v_i-u_i\right)=\frac{1}{2}\sum_{i=1}^{m}x_i^2\left(v_i^2-u_i^2\right)$$

$$a\sum_{i=1}^{m}x_i^3\left(v_i-u_i\right)+b\sum_{i=1}^{m}x_i^2\left(v_i-u_i\right)+c\sum_{i=1}^{m}x_i\left(v_i-u_i\right)=\frac{1}{2}\sum_{i=1}^{m}x_i\left(v_i^2-u_i^2\right)$$

$$a\sum_{i=1}^{m}x_i^2\left(v_i-u_i\right)+b\sum_{i=1}^{m}x_i\left(v_i-u_i\right)+c\sum_{i=1}^{m}\left(v_i-u_i\right)=\frac{1}{2}\sum_{i=1}^{m}\left(v_i^2-u_i^2\right)$$

Analiza datelor

Am ales pentru analiza noastra datele indicatorului Dow Jones, iar pentru exemplificare am selectat datele zilnice din ultimele trei luni (august-octombrie 2014).

Datele sunt preluate prin Internet de la http://finance.yahoo.com/q/hp?s=^DJI intr-o foaie de lucru Excel, in forma:

0.001	17.72044	17.60389
0.002	17.71226	17.6245
0.003	17.73571	17.64203
0.004	17.67507	17.60681
0.005	17.66415	17.6132
0.006	17.70548	17.58388
0.007	17.62671	17.53617
0.008	17.63821	17.58494

0.009	17.62187	17.54751
0.01	17.57533	17.49337
0.011	17.56031	17.44035
0.012	17.48659	17.38576

Toate valorile au fost impartite prin 1000, coloana a doua reprezentand valoarea maxima a zilei, iar coloana a treia valoarea minima.

Ulterior, datele sunt aduse intr-o foaie de lucru in Mathcad, in forma:

date :=

...\dj1000.xls

$v := date^{\langle 1 \rangle}$

$u := date^{\langle 2 \rangle}$

$x := date^{\langle 0 \rangle}$

n := length(x

Urmeaza calculul coeficientilor regresiei polinomiale:

$$s4 := \sum_{i=0}^{n-1} \left[(x_i)^4 \cdot (v_i - u_i) \right] \qquad s3 := \sum_{i=0}^{n-1} \left[(x_i)^3 \cdot (v_i - u_i) \right]$$

$$s2 := \sum_{i=0}^{n-1} \left[(x_i)^2 \cdot (v_i - u_i) \right] \qquad s1 := \sum_{i=0}^{n-1} \left[x_i \cdot (v_i - u_i) \right]$$

$$s0 := \sum_{i=0}^{n-1} (v_i - u_i) \qquad \sigma 2 := 0.5 \cdot \sum_{i=0}^{n-1} \left[(x_i)^2 \cdot \left[(v_i)^2 - (u_i)^2 \right] \right]$$

$$\sigma 1 := 0.5 \cdot \sum_{i=0}^{n-1} \left[x_i \cdot \left[(v_i)^2 - (u_i)^2 \right] \right] \qquad \sigma 0 := 0.5 \cdot \sum_{i=0}^{n-1} \left[(v_i)^2 - (u_i)^2 \right]$$

si apoi determinarea solutiei, respectiv a coeficientilor regresiei polinomiale de gradul 2:

$a := 0$

$b := 0$

$c := 0$

Given

$a \cdot s4 + b \cdot s3 + c \cdot s2 = \sigma 2$

$a \cdot s3 + b \cdot s2 + c \cdot s1 = \sigma 1$

$a \cdot s2 + b \cdot s1 + c \cdot s0 = \sigma 0$

$$\text{Find}(a, b, c) = \begin{pmatrix} 837.211 \\ -57.872 \\ 17.699 \end{pmatrix}$$

Desigur, trebuie sa verificam daca valorile obtinute reprezinta un punct de minim, adica daca matricea hessiana:

$$\begin{pmatrix} s4 & s3 & s2 \\ s3 & s2 & s1 \\ s2 & s1 & s0 \end{pmatrix}$$

este pozitiv definite, iar pentru aceasta verificam criteriul lui Sylvester, respective verificam pozitivitatea determinantilor de pe diagonala principala:

$$\left| \begin{pmatrix} s4 & s3 & s2 \\ s3 & s2 & s1 \\ s2 & s1 & s0 \end{pmatrix} \right| = 2.154 \times 10^{-8}$$

$$\left| \begin{pmatrix} s4 & s3 \\ s3 & s2 \end{pmatrix} \right| = 3.099 \times 10^{-8}$$

$$s4 = 2.86 \times 10^{-5}$$

References

Dolado, J.J. and Lütkepohl, H., 1996. Making Wald Test Work for Cointegrated VAR Systems. *Econometric Reviews*, 15, pp.369-386.

INNSE, 2009. Monthly Statical Bulletin. [online] Available at: http://www.insse.ro/cms/files/arhiva_ buletine2009/bsl_4.pdf. [Accessed in April 2009].

Lütkepohl, H. and Kratzig, M., 2004. *Applied Time Series Econometrics (Themes in Modern Econometrics)*. Cambridge University Press.

Mateescu, G.D., 2013. Multiple Points Regression, Romanian Journal of Ecomomic Forecasting, 2/2013, p.141-144