

Využitie programu Microsoft Excel pri ekonometrickom modelovaní

Martin Lukáčik, Adriana Lukáčiková, Karol Szomolányi

Aplikovanú ekonometriu, najmä odhad parametrov modelu a testovanie predpokladov si už v súčasnosti bez počítačového vybavenia a špecializovaných softvérových produktov nevieme predstaviť. Špecializované komerčné ekonometrické programy nie sú lacnou záležitosťou a takisto zvládnutie každého nového softvéru vyžaduje istý čas a námahu, preto pri jednoduchších úlohách stojí za zváženie, či na ich realizáciu nepostačuje „obyčajný“ tabuľkový procesor so svojimi možnosťami. Kvôli zváženiu tejto alternatívy popíšeme možnosti využitia programu Microsoft Excel pri ekonometrickej analýze. Ako zdrojové údaje pre analýzu využijeme ročné údaje o importe tovarov a služieb, celkovej konečnej spotrebe a tvorbe hrubého kapitálu za roky 1995 až 2009 v mld. Eur bežných cien publikované Štatistickým úradom SR v septembri 2010.

Úvod

Ekonometrické metódy sa dajú v Exceli (abstrahujúc od možností programovania) aplikovať tromi základnými spôsobmi:

- pomocou zabudovaných funkcií pre odhad parametrov modelu;
- pomocou zabudovaných funkcií pre prácu s maticami;
- pomocou nástroja Regrese v štandardne dodávanom doplnku Analytické nástroje.

Základnou funkciou v Exceli určenou pre odhad parametrov lineárneho modelu je funkcia *LINEST* a jej analógiou pre exponenciálny model je funkcia *LOGEST*. Tieto funkcie sa zadávajú ako tzv. vzorec poľa, lebo ich výsledok sa zobrazuje do viacerých buniek. Ich doplnkom sú funkcie *TREND* a funkcia *GROWTH*, ktoré slúžia pre získanie ďalších hodnôt takýmto spôsobom odhadnutej priamky (respektíve krivky), tzv. prognóz.

Program Excel umožňuje základné výpočtové operácie s maticami a vektormi pomocou niekoľkých funkcií, a to: pomocou funkcie *MMULT*, ktorou sa dá získať súčin matic; pomocou funkcie *MINVERSE*, ktorá slúži na výpočet inverznej matice; pomocou funkcie *TRANSPOSE*, pomocou ktorej sa transponuje vektor alebo matica a pomocou funkcie *MDETERM*, ktorá vypočíta determinant určenej štvorcovej matice.

S výnimkou poslednej funkcie pre prácu s maticami *MDETERM* a prípadov, keď výsledkom je skalár sa tieto funkcie zadávajú rovnako ako *LINEST* a *LOGEST* ako tzv. vzorec poľa. Pripomeňme, že pri takomto type funkcie, po zvolení „správnej veľkosti“ oblasti výstupu, sa po zadaní funkcie s argumentmi musí stlačiť kombinácia kláves *Ctrl+Shift+Enter*, aby sa výsledok zobrazil úplne celý.

Doplnok Analytické nástroje je doplnkový program, ktorý pridáva novú funkčnosť do programu Microsoft Excel. Tento doplnok je vhodné pridať v rámci voliteľnej inštalácie programu Excel. Aj po pridaní pri inštalácii prípadne pri zvolenej úplnej inštalácii, sa pred prvým použitím musí doplnok najprv načítať. Načítanie doplnku sa od verzie Excelu 2007 spúšťa kliknutím na tlačidlo *Office* a v jeho ponuke na položku *Možnosti programu Excel*. V tomto okne sa potvrdí položka *Doplnky* a v dolnom zozname s názvom *Správa* sa vyberú *Doplnky programu Excel*. Zo zoznamu sa zvolí tlačidlo *Spustiť*. Otvorené okno *Doplnky* sa v starších verziách Excelu spúšťalo v ponuke *Nástroje* cez možnosť *Doplnky*. V tomto okne v poli *Dostupné doplnky* sa začiarkne políčko *Analytické nástroje* a voľba sa potvrdí tlačidlom OK. Po načítaní doplnku sa na karte *Údaje* v skupine *Analýza* zobrazuje príkaz *Analýza údajov*. V starších verziách bol tento príkaz dostupný v poslednej sekcii ponuky *Nástroje*.

Usporiadanie údajov pred odhadom parametrov

Pred odhadom parametrov odporúčame vždy usporiadať údaje tak, aby stĺpce s vysvetľujúcimi premennými tvorili súvislú oblasť – rady x1, x2 (prípadne ďalšie, ak ich je viac) musia byť usporiadané vedľa seba. (Pri odhadoch pomocou matíc je pred nimi stĺpec 1!)

Obrázok 1: Usporiadanie dát – vyznačené stĺpce s vysvetľujúcimi premennými susedia

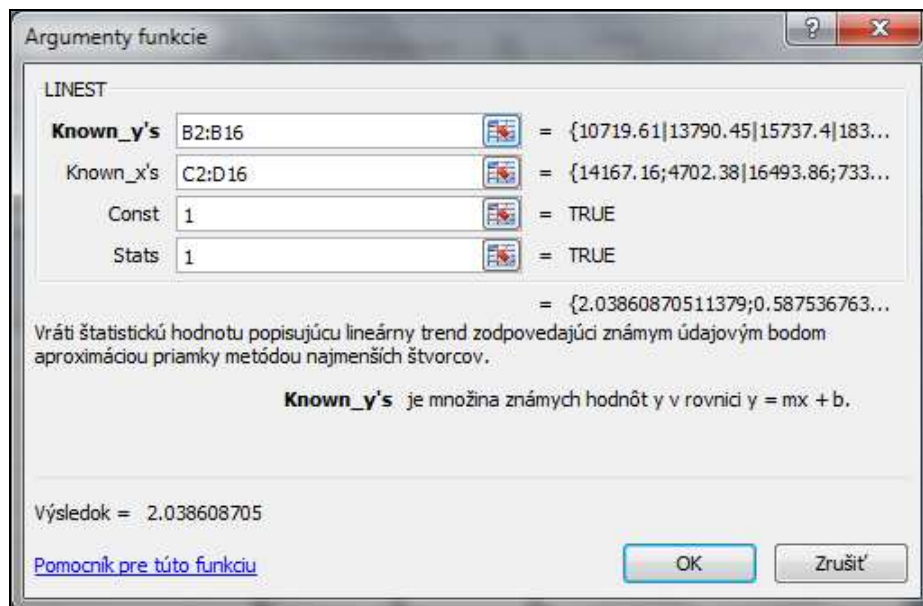
	A	B	C	D	E
1		y	x1	x2	
2	1995	10719.6	14167.2	4702.38	
3	1996	13790.5	16493.9	7333.67	
4	1997	15737.4	17962.3	8182.3	
5	1998	18304.9	20192.7	8781.75	
6	1999	18430.4	21512.1	7813.38	
7	2000	22738.4	23826	8100.44	
8	2001	27345.7	26545.7	10019.3	
9	2002	28825.8	28725.3	10704.4	
10	2003	31545.3	31332.9	9992.7	
11	2004	34883.2	34449.4	11896.8	
12	2005	39864.8	37302.7	14239	
13	2006	48649.5	41815.2	15409.2	
14	2007	54017.6	45096.5	17095.4	
15	2008	57327.4	49876.2	18857.5	
16	2009	44564.9	50723.9	13044.8	

Odhad parametrov lineárneho modelu pomocou funkcie *LINEST*

Pred spustením funkcie *LINEST* zo zoznamu všetkých funkcií, najskôr vyznačíme oblasť výstupu v hárku, ktorú tvorí vždy päť riadkov a taký počet stĺpcov, koľko odhadujeme parametrov. Napríklad pre model s dvomi vysvetľujúcimi premennými a absolútnym členom vyznačíme tri stĺpce.

Funkcia má štyri argumenty, preto pri zadávaní funkcie *LINEST* potrebujeme postupne vyplniť štyri informačné riadky. Prvým argumentom je **Known_y's**, kde zadávame súvislú oblasť buniek obsahujúcu hodnoty modelovanej (závislej) premennej – podľa usporiadania dát na obrázku 1, to bude rozsah buniek B2:B16. Druhým argumentom je **Known_x's**, kde zadávame súvislú oblasť buniek obsahujúcu hodnoty vysvetľujúcich premenných – podľa usporiadania dát na obrázku 1, to bude rozsah buniek C2:D16. Tretím argumentom je **Const**, kde zadávame 1, ak odhadujeme parametre modelu s absolútnym členom a 0, ak odhadujeme parametre modelu bez absolútneho člena. Posledným štvrtým argumentom je **Stats**, kde zadávame 1, ak chceme vypočítať okrem odhadnutých parametrov aj ďalšie štatistiky pomáhajúce zhodnotenie zvoleného modelu a kvalitu realizovaného odhadu.

Obrázok 2: Funkcia *LINEST* v programe Microsoft Excel



Po vyplnení argumentov podľa obrázku 2 a stlačení kombinácie *Ctrl+Shift+Enter* získame výsledok, ktorý vidíme na obrázku 3.

Obrázok 3: Odhad parametrov modelu pomocou funkcie *LINEST*

	G	H	I	J
1	2.0386087	0.5875368	-9486.416103	
2	0.3304413	0.1081108	1488.094932	
3	0.9865355	1882.4278	#NEDOSTUPNÝ	
4	439.61637	12	#NEDOSTUPNÝ	
5	3.116E+09	42522414	#NEDOSTUPNÝ	

Usporiadané výsledné hodnoty funkcie zapísané zaužívanými symbolmi sú zobrazené na obrázku 4. Význam jednotlivých hodnôt získaných funkciou *LINEST* je nasledovný:

- $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ – odhadnuté hodnoty parametrov modelu;
- $\hat{\sigma}_{\beta_0}, \hat{\sigma}_{\beta_1}, \hat{\sigma}_{\beta_2}, \dots, \hat{\sigma}_{\beta_k}$ – odhadnuté hodnoty štandardných odchýlok parametrov modelu;
- R^2 – koeficient determinácie odhadnutého lineárneho modelu;
- $\hat{\sigma}$ – odhad štandardnej odchýlky náhodnej zložky;
- F – vypočítaná hodnota F štatistiky modelu;
- df – počet stupňov voľnosti;
- ESS – suma štvorcov vysvetlená modelom;
- RSS – suma štvorcov reziduálov.

Obrázok 4: Usporiadanie výsledných hodnôt výstupu funkcie *LINEST*

$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_0$
$\hat{\sigma}_{\beta_2}$	$\hat{\sigma}_{\beta_1}$	$\hat{\sigma}_{\beta_0}$
R^2	$\hat{\sigma}$	
F	df	
ESS	RSS	

resp.

$\hat{\beta}_k$	$\hat{\beta}_{k-1}$...	$\hat{\beta}_1$	$\hat{\beta}_0$
$\hat{\sigma}_{\beta_k}$	$\hat{\sigma}_{\beta_{k-1}}$...	$\hat{\sigma}_{\beta_1}$	$\hat{\sigma}_{\beta_0}$
R^2	$\hat{\sigma}$			
F	df			
ESS	RSS			

Prv ako zapíšeme výsledný model obvyklým spôsobom môžeme dopočítať hodnoty t štatistík jednotlivých parametrov, ktoré získame ako podiel odhadnutých hodnôt parametrov (údajov v prvom riadku) a odhadnutých hodnôt štandardných odchýlok parametrov (údajov z druhého riadku). Napríklad $t_{\beta_2} = \hat{\beta}_2 / \hat{\sigma}_{\beta_2} = 6,16935 = 2,0386 / 0,3304$ a odhadnutý model má tvar:

$$\hat{y}_i = -9486,4 + 0,5875x_{1i} + 2,0386x_{2i} \quad n = 15$$

(1488,1) (0,108) (0,330) $\hat{\sigma}^2 = 1882,4^2$

[-6,375] [5,435] [6,169] $R^2 = 0,9865$

Odhad parametrov lineárneho modelu pomocou funkcií pre prácu s maticami

Postup získania odhadu parametrov pomocou matic, sleduje maticový zápis estimátora metódy najmenších štvorcov, teda vzťahu $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Preto prvým krokom po usporiadaní údajov a vytvorení matice vysvetľujúcich premenných \mathbf{X} (tvorí ju stĺpec jednotiek nasledovaný stĺpcami x_1 a x_2) je vytvorenie transponovanej matice \mathbf{X}^T za pomoci funkcie *TRANSPOSE*, pričom musíme správne vyznačiť výstupnú oblasť tejto funkcie (rozmer matice \mathbf{X} je 15×3 , preto rozmer matice \mathbf{X}^T musí byť 3×15). Jej jediným argumentom je rozsah buniek, ktorý obsahuje pôvodnú maticu \mathbf{X} .

Ďalším krokom je postupné získanie dvoch súčinov matíc $\mathbf{X}^T\mathbf{X}$ a $\mathbf{X}^T\mathbf{y}$ pomocou funkcie *MMULT*, pričom musíme správne vyznačiť výstupnú oblasť tejto funkcie (jej rozmer bude v prvom prípade 3×3 a v druhom prípade 3×1). Prvým argumentom funkcie je matica (prípadne vektor), ktorý je v súčine zľava a druhým argumentom je matica (prípadne vektor), ktorý je v súčine sprava.

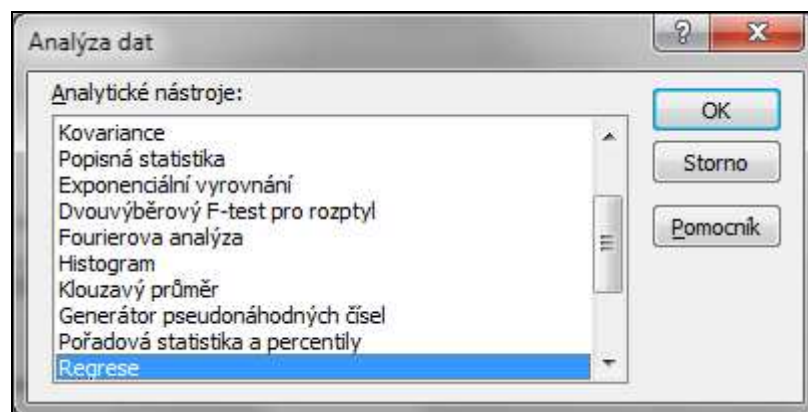
Nasleduje získanie inverznej matice $(\mathbf{X}^T\mathbf{X})^{-1}$ za pomoci funkcie *MINVERSE*, pričom musíme správne vyznačiť výstupnú oblasť tejto funkcie (jej rozmer bude 3×3). Jediným argumentom funkcie je rozsah buniek, ktorý obsahuje maticu $\mathbf{X}^T\mathbf{X}$, ktorú chceme invertovať.

Posledným krokom je opätovné použitie funkcie *MMULT* pre získanie súčinu matíc $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ s rozmerom 3×1 . Celý postup sa dá spojiť do jednej takejto oblasti pomocou vnorených funkcií. Takisto sa dajú následne dopočítať všetky sprievodné štatistiky ako v prípade použitia funkcie *LINEST*.

Odhad parametrov lineárneho modelu pomocou doplnku Analytické nástroje

Po načítaní doplnku *Analytické nástroje* a jeho spustení sa nám otvorí okno *Analýza dat* zobrazené na obrázku 5, v ktorom máme možnosť vybrať si niektorý z ponúkaných dostupných nástrojov slúžiacich pre rôzne typy analýz. Na odhad parametrov je vhodné použiť nástroj s názvom *Regrese*.

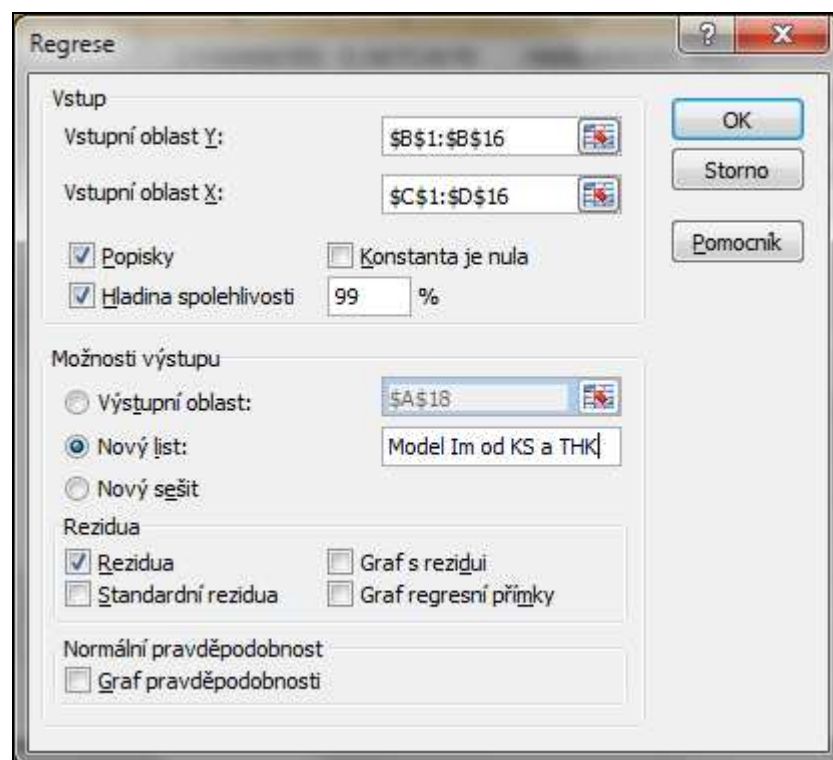
Obrázok 5: Ponuka nástrojov doplnku Analytické nástroje



Po označení nástroja *Regrese* a potvrdení *OK* sa zobrazí dialógové okno *Regrese* zobrazené na obrázku 6, v ktorom musíme špecifikovať, aký model budeme odhadovať a aké informácie majú doplniť preddefinovaný výpis tejto aplikácie. Charakterizovať musíme minimálne, čo je vstupom a kde sa má realizovať výstup nástroja. Do **Vstupní oblast Y**: zadáme súvislú oblasť buniek obsahujúcu hodnoty modelovanej (závislej) premennej, pričom teraz do rozsahu môžeme zahrnúť aj bunku s názvom premennej (oblasť buniek B1:B16). Rozsah sa v prípade vyznačovania automaticky zadá ako absolútna adresa (\$B\$1:\$B\$16).

Do **Vstupní oblast X**: obdobne zadáme súvislú oblasť buniek obsahujúcu hodnoty vysvetľujúcich premenných rovnako aj s bunkami s názvami premenných (C1:D16). Rozsah sa v prípade vyznačovania takisto automaticky zadá ako absolútna adresa (\$C\$1:\$D\$16). Vo vstupnej oblasti zaškrtneme možnosť **Popisky**, ak sme zadali rozsahy údajov aj s bunkami s názvami premenných. Pre prípad, keď by sme chceli získať intervalový odhad aj pre inú hladinu spoľahlivosti ako je 95 %, môžeme túto informáciu zadať v oblasti *Vstupu* (my sme zvolili hodnotu 99 %). Ak by sme chceli odhadovať model bez absolútneho člena, čo nie náš prípad, musíme zaškrtnúť možnosť **Konstanta je nula**.

Obrázok 6: Nástroj Regrese



V oblasti *Možnosti výstupu* je nevyhnutnou zadanie informácie, kde sa zobrazí výstup aplikácie. Na výber je **Výstupná oblasť**: v tom istom hárku ako sú zadané vstupné údaje, pričom stačí zadať bunku od ktorej sa vloží výpis. Ďalej je tu možnosť zvoliť si **Nový list**: (teda nový hárok) a zároveň si ho aj pomenovať v tom istom súbore ako sú zadané vstupné údaje. A poslednou možnosťou je voľba výstupu do nového súboru cez možnosť **Nový sešit**.

Okrem informácie o oblasti výstupu sa tu dá zvoliť veľmi užitočná možnosť **Rezidua**, ktorá doplní preddefinovaný výpis tejto aplikácie o tabuľku obsahujúcu vyrovnané hodnoty modelovanej premennej a reziduály. Tieto umožňujú realizáciu ďalších ekonometrických metód užitočných pri analýze. Okrem tejto možnosti sa môže k bežnému výpisu pridať zobrazenie viacerých rôznych grafov, napríklad reziduálov, či regresnej priamky.

Obrázok 7: Výsledok odhadu parametrov pomocou nástroja Regrese

	A	B	C	D	E	F	G	H	I
1	VÝSLEDEK								
2									
3	Regresní statistika								
4	Násobné R	0.993244936							
5	Hodnota spoľehlivosti R	0.986535503							
6	Nastavená hodnota spoľehlivosti	0.984291421							
7	Chyba stf. hodnoty	1882.427826							
8	Pozorování	15							
9									
10	ANOVA								
11		Rozdíl	SS	MS	F	Významnost F			
12	Regrese	2	3115591535	1557795767	439.616366	5.95855E-12			
13	Rezidua	12	42522414.26	3543534.522					
14	Celkem	14	3158113949						
15									
16		Koeficienty	Chyba stf. hodnoty	t Stat	Hodnota P	Dolní 95%	Horní 95%	Dolní 99.0%	Horní 99.0%
17	Hranice	-9486.4161	1488.094932	-6.37487293	3.5315E-05	-12728.69643	-6244.13578	-14031.861	-4940.97123
18	x1	0.587536764	0.108110832	5.43457811	0.00015138	0.351983495	0.823090032	0.257307946	0.917765581
19	x2	2.038608705	0.330441291	6.169352191	4.8036E-05	1.318638981	2.758578429	1.0292627	3.047954711
20									
21	REZIDUA								
22									
23		Pozorování	Očekávané y	Rezidua					
24		1	8423.624035	2295.985965					
25		2	15154.81652	-1364.366522					
26		3	17747.58001	-2010.180011					
27		4	20280.066	-1975.175997					
28		5	19081.12862	-650.6986155					
29		6	21025.86232	1712.567676					
30		7	26535.57625	810.0737505					
31		8	29212.88888	-387.1188814					
32		9	29294.01976	2251.260237					
33		10	35006.70239	-123.5423871					
34		11	41458.10204	-1593.332037					
35		12	46494.86316	2154.67684					
36		13	51860.24331	2157.31669					
37		14	58260.69927	-933.3392725					
38		15	46909.06743	-2344.127434					

Výsledok odhadu parametrov zobrazený na obrázku 7 je rozdelený na viacero častí. Prvou tabuľkou výsledkov úplne hore je **Regresní statistika**. Tu sa nachádza informácia o hodnotách koeficientu determinácie, odhade štandardnej odchýlky náhodnej zložky a počte pozorovaní. Druhou tabuľkou v poradí je **ANOVA**, kde nájdeme informácie o sumách štvorcov a *F* štatistike. Nasleduje tabuľka s odhadmi parametrov, odhadmi ich štandardných odchýlok a *t* štatistikami parametrov obsahujúca aj intervalové odhady. Ak sme zvolili možnosť **Rezidua**, tak k štandardnému výpisu pribudla tabuľka obsahujúca vyrovnané hodnoty modelovanej premennej **Očekávané y** a reziduály v stĺpci **Rezidua**. Tento postup je najrýchlejší a najkomplexnejší.

Literatúra:

- [1] LUKÁČIKOVÁ, A. – LUKÁČIK, M.: Ekonometrické modelovanie s aplikáciami. Bratislava: EKONÓM, 2008.
- [2] SURMANOVÁ, K.: Úvod do ekonometrie – Praktikum. Bratislava: EKONÓM, 2009.
- [3] Pomocník programu Microsoft Excel a internetové stránky programu Microsoft Excel
- [4] Databáza ŠÚ SR – SLOVSTAT
- [5] <http://www.fhi.sk/files/katedry/kove/veda-vyskum/prace/2010/Ekonometria-a-excel.xls>