

19. december 2011

**„Vygooglime“ si budúcnosť**Pripravili: *Monika Pécsyová*

Nedávna recesia ukázala, že žiaden z tradičných komplexných ekonometrických modelov nedokázal predvídať dostatočne rýchlo tak výrazný prepád produkcie na začiatku krízy. V dnešných časoch vysokej neistoty môže sledovanie tradičných dát nepostačujúce, keďže tie sú publikované s výrazným oneskorením. Preto sa v súčasnosti pozornosť ekonómov upriamuje na údaje, ktoré dokážu priniesť informáciu o momentálnej ekonomickej situácii v aktuálnom čase. Okrem predstihových „mäkkých“ indikátorov s čoraz častejším využitím internetu rastie vypovedacia schopnosť internetových vyhľadávaní. V zahraničí sa potvrdila ich významnosť pri predikovaní maloobchodných tržieb, predaja automobilov, nehnuteľností, návštevnosti turistických destinácií a trhu práce. Komentár IFP naznačuje, že internetové vyhľadávania dokážu pomôcť vysvetliť mieru nezamestnanosti na Slovensku a zlepšujú jej krátkodobú prognózu o viac ako 30%. S rastúcim používaním internetu tak vzniká zaujímavá analytická databáza, ktorá rozhodne má potenciál slúžiť ako alternatíva k tradičným doteraz používaným údajom.

*Prečo sú štatistiky internetových vyhľadávaní zaujímavé?*

Internet je fenomén posledných rokov, pričom záujem o neho rastie každým rokom. Vďaka internetu môžeme zaplatiť mesačnú faktúru, sledovať televíziu, telefonovať s rodinou v zahraničí ale najmä vyhľadávať informácie rôzneho druhu. O každom vyhľadávaní pritom existuje záznam, je preto možné si pre ľubovoľnú tému pozrieť, či v nejakom momente v čase záujem surfujúcich ľudí o ňu zrazu rapídne nevzrástol. To (okrem iných) poskytuje nástroj v rámci vyhľadávača Google<sup>1</sup> – tzv. *Google insights for search data set* (viac v Boxe). Takto napríklad vieme zistiť, že za celé obdobia sledovania internetových štatistík (od roku 2004) bolo najčastejšie hľadaným heslom v Google vyhľadávači slovo „facebook“. Okrem Slovákov si však „facebook“ spolu s heslom „youtube“ googlia najviac aj internetoví surferi na celom svete. Mnoho užívateľov totiž používa Google vyhľadávanie, aj keď pozná stránku, kam sa chce dostať. Namiesto zadania presnej URL adresy tak zadajú kľúčové slovo stránky do vyhľadávača. Nástroj Google insights for search preto okrem vyhľadávaných hesiel zaznamenáva aj stránky o ktorých užívatelia dopredu vedia, že ich chcú navštíviť.

*Poskytujú informáciu v reálnom čase*

Zo štatistík internetových vyhľadávaní vieme tiež zistiť, že ľudí na celom svete spája zvýšený záujem o tragédie. Za posledných 90 dní zaznamenal na Slovensku najrýchlejší rast počet vyhľadávaní informácií o tragickej smrti hokejistu Pavla Demitru, na celom svete to bolo meno zakladateľa Applu Steve Jobs. Štatistiky internetových vyhľadávaní vieme aplikovať aj v prípade ekonomie. Vieme takto napríklad sledovať, ako sa v čase mení záujem ľudí o kúpu auta, alebo chladničky alebo či narástol počet ľudí, ktorý si cez internet hľadajú prácu. **Heslo „profesia“ sa zaradilo medzi desať najrýchlejšie rastúcich internetových vyhľadávaní v roku 2011, teda v časoch, kedy miera nezamestnanosti takmer neustále rastie. Internetové vyhľadávania tak poskytujú potenciálne užitočný zdroj dát na monitorovanie a predpovedanie aktuálnej ekonomickej aktivity – tzv. *nowcasting*.**

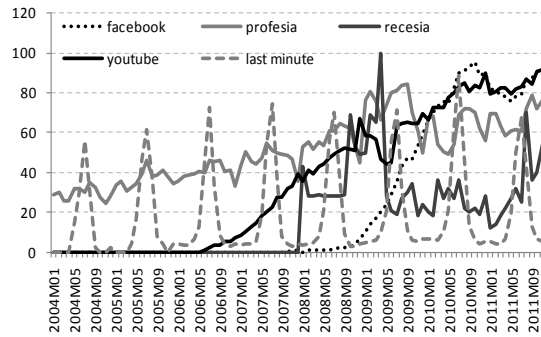
*Až dve tretiny slovenských domácností má prístup k internetu*

Eurostat každý rok zverejňuje výsledky o prístupe vybraných domácností a jednotlivcov na internet. **V roku 2010 bolo na Slovensku pripojených 67%** domácností, čo je len mierne pod priemerom EÚ (70%). Počet domácností s prístupom na internet sa na Slovensku od roku 2004 viac ako strojnásobil, čo zo Slovenska (po Rumunsku, Bulharsku a Česku) robí jednu z krajín s najrýchlejšim šírením internetu v EÚ. Najviac domácností je pripojených v Holandsku (91%) a Luxembursku (90%), najmenej

<sup>1</sup> Internetový prehliadač Google má najväčší, viac ako 65%-ný podiel na trhu ([http://blog.nielsen.com/nielsenwire/online\\_mobile/nielsen-reports-february-2010-u-s-search-rankings/](http://blog.nielsen.com/nielsenwire/online_mobile/nielsen-reports-february-2010-u-s-search-rankings/)). V nasledujúcom texte pod pojmom internetové vyhľadávania budeme rozumieť vyhľadávania v prehliadači Google.

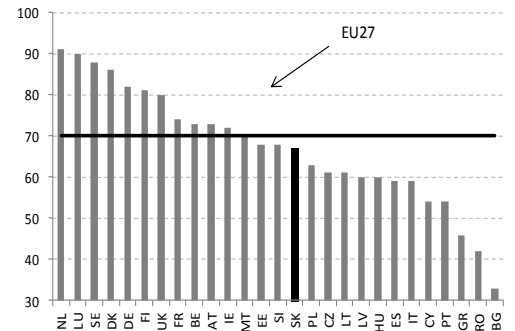
pripojených domácností je v Bulharsku (33 %), Rumunsku (42 %) a v Grécku (46 %). Dáta za jednotlivcov vo veku 16-74 ukazujú, že až **70% dospelých je pripojených denne**, pričom výrazne poklesol počet takých, ktorý nikdy neboli pripojený na internet. V najbližších rokoch sa prirodzene očakáva, že rady Slovákov pripojených na internet sa budú aj naďalej rozširovať.

**Graf 1: Príklady vývoja internetových vyhľadávani na Slovensku**



Zdroj: Google

**Graf 2: Podiel domácností (v %) s prístupom na internet v krajinách EÚ v roku 2010**



Eurostat

Zdroj:

*Plusy vs. mínusy štatistik internetových vyhľadávani*

Dáta o internetovom vyhľadávaní majú v porovnaní s bežne používanými dátami niekoľko výhod. Kvartálne údaje za národné účty a mesačné údaje za vybrané odvetvia (tržby, zamestnanosť, mzdy, priemyselná produkcia, nové objednávky) sú totiž publikované s určitým oneskorením (2-3 mesiace), kým **internetové dáta sú dostupné okamžite**, pričom dávajú informáciu o vyhľadávaní za predchádzajúci týždeň. Porovnateľnú časovú dostupnosť poskytujú aj tzv. „mäkké“ dáta, ktoré sú výsledkom prieskumu vybraných domácností a firiem v priemysle, službách, stavebníctve a maloobchode. Tie však zodpovedajú len vopred preddefinované otázky, v prípade firiem napr. o očakávanom dopyte alebo zamestnanosti a v prípade domácností o ich plánoch realizovať v najbližšom čase veľké nákupy. Navyše, prieskumy nevedia vyriešiť problém s nepresne zodpovedanými resp. nezodpovedanými otázkami. Rozdiel je aj vo veľkosti vzorky respondentov, kde prieskumy pokrývajú len vybranú časť domácností, pričom **prístup na internet má dnes viac ako 1,3 mil. domácností na Slovensku**.

Aj internetové dáta však majú svoje mínusy. Keďže internet je relatívne nový fenomén, časové rady za internetové vyhľadávania sú krátke – len od roku 2004. Navyše existuje vysoká pravdepodobnosť, že sú korelované s faktormi ako príjem alebo vek používateľov internetu, takže vzniká problém s reprezentatívnosťou vzorky. V niektorých prípadoch môže nastať situácia, že dvaja rôzni užívatelia internetu si chcú vyhľadať tú istú vec, do vyhľadávača však zadajú niečo iné. V prípade Slovenska je tento problém o to väčší, keďže užívatelia môžu alebo nemusia zadávať slová s diakritikou. Potom dostávame dva rozdielne rady za vyhľadávané heslá „pracovná ponuka“ a „pracovna ponuka“.

**Box: Použitie nástroja Google Insights for Search Data**

Nástroj Google Insights for Search umožňuje porovnať si záujem a popularitu ľubovoľnej témy, ktorú si užívatelia vyhľadávajú cez internet, v čase a v rámci jednotlivých krajín. Záujem o vyhľadávanú tému je počítaný ako týždenný index, ktorý vyjadruje, koľko vyhľadávani sa uskutočnilo pre konkrétny výraz v porovnaní s celkovým počtom vyhľadávani v službe Google za dané časové obdobie. Index teda predstavuje pravdepodobnosť toho, že náhodný používateľ vyhľadáva konkrétny výraz z určitého miesta v určitý čas. Týmto spôsobom výpočtu sa zároveň kontroluje rastúci trend internetových vyhľadávani. Podiely sú potom normalizované tak, aby sa najvyššia hodnota podielu rovnala hodnote 100. Keďže index nevyjadruje absolútny objem vyhľadávani, treba dávať pozor na interpretáciu jeho vývoja. Jeho nárast (pokles) potom totiž znamená, že rastie (klesá) rýchlejšie ako celkový počet vyhľadávani. Pozor treba dávať aj v prípade nulových hodnôt, ktoré neznamenajú nulový záujem užívatelov o danú tému, ale návštevnosť pod hranicu zadanú nástrojom Google Insights. Systém zároveň vylučuje opakovaný dopyt od jedného používateľa v krátkom časovom období, aby tento typ dopytov umelo neovplyvňoval úroveň záujmu.

Nástroj Google Insights for Search je voľne dostupný na stránke <http://www.google.com/insights/search>.

Internetové vyhľadávania sú štatisticky signifikantné a môžu pomôcť zlepšiť prognózu miery nezamestnanosti

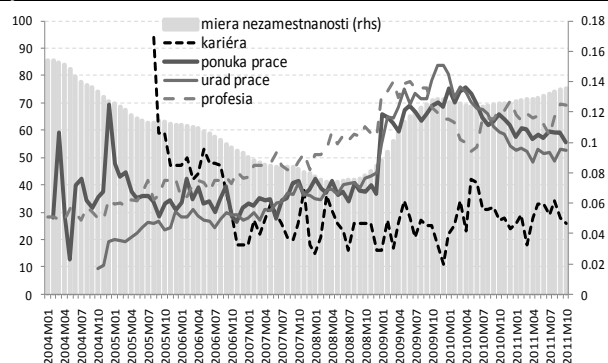
Aj napriek tomu, že ide o relatívne novú tému, v súčasnosti existuje pomerne veľké množstvo aplikovaných článkov, ktoré skúmali použitie internetových vyhľadávanií pri modelovaní a predikovaní rôznych ekonomických premenných. **Ukázala sa ich významnosť pri predikovaní maloobchodných tržieb<sup>2</sup>, predaja automobilov, nehnuteľností<sup>3</sup>, návštevností turistických destinácií a miery nezamestnanosti<sup>4</sup>.**

**Cieľom komentára IFP bolo preskúmať užitočnosť použitia internetových dát pri krátkodobom predikovaní miery nezamestnanosti na Slovensku.** Predpokladáme, že ak je človek nezamestnaný alebo mu to potenciálne hrozí, môže použiť internet ako prostriedok pri hľadaní informácií o tom, ako postupovať na úrade práce resp. pri hľadaní novej práce. Preto by nárast takýchto internetových vyhľadávanií mal indikovať nárast miery nezamestnanosti. Samozrejme, v tomto prípade je kľúčové správne určiť, ktoré heslá (potenciálne) nezamestnaný človek na internete vyhľadá. Skúsili sme viacero slovných kombinácií, pre niektoré však neexistoval dostatočne veľký počet vyhľadávanií. Nakoniec sme použili nasledovné heslá: **ponuka práce, práca, úrad práce, profesia a kariéra** (aj bez diakritiky). Mieru nezamestnanosti<sup>5</sup> ( $ur_t$ ) sme kvôli krátkosti časových radov za internetové vyhľadávania modelovali jednoduchým autoregresným modelom nasledovne:

$$\Delta ur_t = \beta_0 + \beta_1 \Delta ur_{t-1} + \beta_2 \Delta \log(y_t) + \beta_3 \Delta \log(i_t),$$

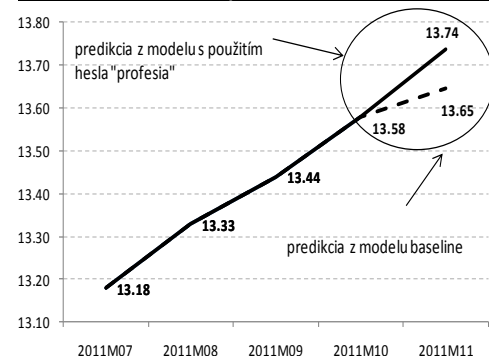
pričom  $y_t$  reprezentuje ekonomickú aktivitu danú tržbami za vybrané odvetvia resp. priemyselnou produkciou a  $i_t$  predstavuje časový rad internetových vyhľadávanií. Pracovali sme so sezónne očistenými mesačnými časovými radmi za obdobie 2004M1-2011M10. Skúmali sme, nakoľko pridanie internetových vyhľadávanií lepšie vysvetlí mieru nezamestnanosti a zlepši prognostickú vlastnosť odhadnutých modelov.

**Graf 3: Miera nezamestnanosti vs. použité príbuzné internetové vyhľadávania**



Zdroj: Google, UPSVAR

**Graf 4: Porovnanie predikcie miery nezamestnanosti z „baseline“ modelu a z „profesia“ modelu**



Zdroj: UPSVAR, MF SR

V prvej fáze sme odhadli rovnicu bez použitia premennej za internetové vyhľadávania (tzv. *baseline*). Išlo v podstate o **veľmi zjednodušenú formu Okunovho zákona**, ktorý kvantifikuje vzťah medzi mierou nezamestnanosti a produkcie. V našom prípade bol koeficient  $\beta_2$ , ktorý meral citlivosť zmeny miery nezamestnanosti na zmenu priemyselnej produkcie odhadnutý na úrovni -0,5 až -0,8. Hodnota Okunovho koeficientu je v súlade s odhadmi pre slovenskú ekonomiku<sup>6</sup>.

<sup>2</sup> Choi, H and Varian, H (2009a), 'Predicting the present with GoogleTrends', Chamberlain, G (2010), 'Googling the present'

<sup>3</sup> Nick McLaren (2011), 'Using internet search data as economic indicators'

<sup>4</sup> Choi, H and Varian, H (2009b), 'Predicting initial claims for unemployment benefits, Askitas, N and Zimmerman, K F (2009), 'Google econometrics and unemployment forecasting, Nick McLaren (2011), 'Using internet search data as economic indicators'

<sup>5</sup> Pracovali sme s mierou nezamestnanosti podľa metodiky UPSVAR, keďže tá je vykazovaná mesačne. Rok 2004 je poznačený legislatívnou zmenou v registrácii nezamestnaných. V tom čase došlo k zvýšenému vyradovaniu uchádzačov o prácu na úrade práce.

<sup>6</sup> Použili sme priemyselnú produkciu v dôsledku potreby mesačných dát. Pre Okunov koeficient pozri Ekonomická analýza IFP: Šrámková L. (2010): 'Output gap and NAIRU'

V druhej fáze sme skúmali štatistickú významnosť internetových vyhľadávaní, ktoré sa pridali do odhadnutej *baseline* rovnice. Zistili sme, že významné sú nasledovné vyhľadávané výrazy: „ponuka práce“ (ozn.  $i_1$ ), „urad práce“ (ozn.  $i_2$ ), „profesia“ (ozn.  $i_3$ ) a „kariéra“ (ozn.  $i_4$ ). Ďalej sme skúmali, či pridanie internetových vyhľadávaní pomôže lepšie vysvetliť mieru nezamestnanosti porovnaním štatistiky *upravené R<sup>2</sup>* (ozn.  $R_{adj}^2$ ). Toto kritérium (narozdiel od najčastejšie používanej štatistiky  $R^2$ ) hovorí, nakoľko presne náš model vysvetľuje závislú premennú nezávisle od pridania dodatočnej vysvetľujúcej premennej. Zistili sme, že  $R_{adj}^2$  je vyššia vo všetkých prípadoch odhadnutých rovníc, kde vystupujú internetové vyhľadávania, takže **internetové vyhľadávania sú nielen štatisticky významné, ale aj zlepšujú odhad miery nezamestnanosti**.

Na záver sme porovnali prognózovacu schopnosť *baseline* modelu a modelov s použitím internetových vyhľadávaní na základe strednej absolútnej percentuálnej chyby (MAPE). Pozreli sme sa na statickú „*in-sample*“ prognózu, kde sme vypočítali štatistiku MAPE pre celý časový horizont. Porovnanie MAPE štatistik ukázalo, že **pridanie internetových vyhľadávaní zlepšuje prognózy miery nezamestnanosti v porovnaní s *baseline* modelom o viac ako 30%**. Výsledky sú zosumarizované v nasledujúcej tabuľke:

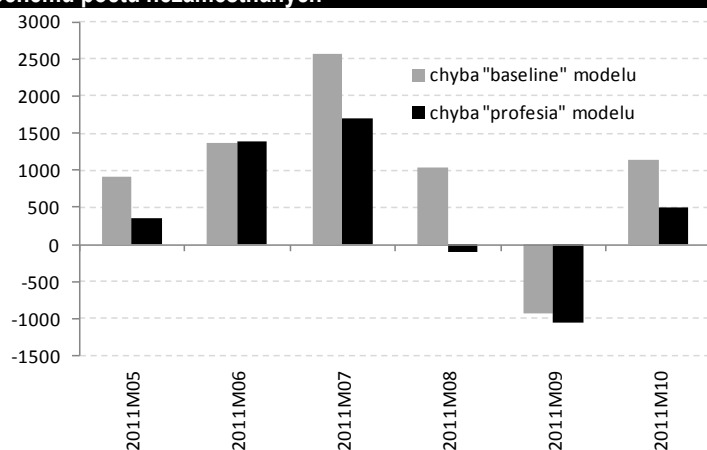
Tabuľka: Výstupy z regresii					
	baseline	„ponuka práce“	„urad práce“	„profesia“	„kariéra“
$\beta_0$	0,05	0	0	0	0
$\Delta u_{t-1}$	0,91	0,91	0,9	0,91	0,9
$\Delta y_t$	-0,68	-0,6	-0,62	-0,61	-0,8
$\Delta i_{1t-1}$	-	0,09	-	-	-
$\Delta i_{2t-1}$	-	-	0,27	-	-
$\Delta i_{3t-1}$	-	-	-	0,36	-
$\Delta i_{3t-2}$	-	-	-	0,26	-
$\Delta i_{3t-3}$	-	-	-	0,31	-
$\Delta i_{4t-1}$	-	-	-	-	0,07
upravené $R^2$	0,86	0,87	0,88	0,88	0,87
MAPE in sample	0,87	0,58	0,55	0,54	0,55

Pozn. Všetky koeficienty sú významné na hladine významnosti 5%.

Zdroj: IFP

Z porovnania výsledkov vyplýva, že najlepší a najspoľahlivejší model miery nezamestnanosti spomedzi použitých je ten, ktorý obsahuje internetové vyhľadávania „profesia“.

Graf 5: Chyba (v počte ľudí) „baseline“ a „profesia“ modelu oproti skutočnému počtu nezamestnaných



Zdroj: UPSVAR, vlastné výpočty

**Predikcia na mesiac dopredu odhaduje nárast zamestnanosti na 13,74%** (Graf 4). Klasický „baseline“ model odhaduje mierne nižší nárast na 13,65%.

S rastúcim použitím a čoraz väčším rozšírením internetu vzniká zaujímavá databáza údajov, ktorá môže pomôcť lepšie vysvetliť to, čo sa momentálne deje v ekonomike. Aj napriek existujúcim limitáciám internetových dát - nevieme zistiť skutočný počet vyhľadávaní a časové rady sú zatiaľ relatívne krátke - dostávame zaujímavú alternatívu k tradičným doteraz používaným dátam, ktoré sú zverejňované s výrazným oneskorením. Internetové vyhľadávania môžu pomôcť vysvetliť maloobchodné tržby, predaj áut, bytov ale aj zmeny v nezamestnanosti. Zo zahraničných inštitúcií ich pravidelne monitoruje napríklad anglická centrálna banka.

**Heslo „profesia“ sa zaradilo medzi desať najrýchlejšie rastúcich internetových vyhľadávaní v roku 2011. Naša jednoduchá analýza ukázala, že krátkodobá prognóza miery nezamestnanosti je o 36% lepšia, ak do modelu zaradíme práve toto vyhľadávanie.** V najbližších dňoch uvidíme, či novembrová miera nezamestnanosti bude naozaj mierne vyššie, ako tá, ktorú prognózuje klasický „baseline“ model.

Ukázalo sa teda, že štatistiky internetových vyhľadávaní dokážu byť v slovenských podmienkach užitočné v prípade trhu práce. **Určite by bolo zaujímavé overiť ich použitie napríklad v prípade krátkodobého výhľadu spotreby domácností alebo či majú internetové dáta potenciál identifikovať bod obratu.**

#### **Upozornenie**

*Materiál prezentuje názory autorov a Inštitútu finančnej politiky, ktoré nemusia nutne odzrkadľovať oficiálne názory Ministerstva financií SR. Cieľom publikovania komentárov Inštitútu finančnej politiky (IFP) je podnecovať a zlepšovať odbornú a verejnú diskusiu na aktuálne ekonomické témy. Citácie textu by sa preto mali odkazovať na IFP (a nie MF SR), ako autora týchto názorov.*