

dr. sc. Tihana Škrinjarčić

Ekonomski fakultet Zagreb, Sveučilište u Zagrebu, Zagreb, Republika Hrvatska
tskrinjar@net.efzg.hr

MOGU LI GOOGLE TREND PODACI POBOLJŠATI PROGNOZIRANJE PRINOSA NA ZAGREBAČKOJ BURZI?

Primljen: 19. listopada 2018.

Prihvaćen: 11. prosinca 2018.

Prethodno priopćenje

Sažetak

Uspješan menadžment portfelja danas predstavlja velik izazov. Danas je investitorima na raspolaganju značajno veći broj investicijskih mogućnosti, kao i informacija, matematičkih i statističkih alata, modela i metoda. U ovome istraživanju nastoji se prikazati kako uključivanje volumena online pretraživanja na najpopularnijoj tražilici Google može doprinijeti modeliranju i prognoziranju prinosa i rizika na Zagrebačkoj burzi. Temeljem mjesečnih podataka za razdoblje siječanj 2004. – rujan 2018. godine, procijenjeno je nekoliko specifikacija ARMA-GARCH modela bez i s uključenim volumenom pretraživanja vezanim uz dionice i Zagrebačku burzu. Rezultati analize su pokazali da su u prognoziranju uspješniji modeli s uključenjem volumena pretraživanja. Dodatno su rezultati potvrđeni simulacijama investicijskih strategija van uzorka, gdje su portfelji temeljeni na prognozama uz uključen Google volumen pretraživanja rezultirali s većim vrijednostima uloženog novca, i uz uključene transakcijske troškove. Međutim, postoje još brojne mogućnosti istraživanja u ovome području kako bi se investicijski ciljevi u budućnosti ostvarili još brže i uspješnije.

Ključne riječi: *Google tražilica, predviđanje prinosa, Zagrebačka burza, volumen pretraživanja*

JEL: C58, D53, G1

1. UVOD

Kontinuirano unapređenje performansi investicijskih portfelja u centru je pažnje akademika i investitora već dugu niz desetljeća. Danas je investitorima na raspolaganju značajno veći broj investicijskih mogućnosti, kao i informacija, matematičkih i statističkih alata, modela i metoda. Dodatno, sve se brže razvija i tehnologija koja omogućuje praćenje kretanja na različitim tržištima, kao i sve složeniju analizu istih. Internet omogućava gotovo istovremenu dostupnost novih informacija važnih za analizu portfelja i dinamičko restrukturiranje istog u svrhu ostvarivanja investitorovih ciljeva. Posljednjih nekoliko godina sve se više pažnje pridodaje povezanosti traženja pomoću pretraživača *Google* i predviđanja prinosa i rizika dionica i općenito financijske imovine (Choi i Varian 2009, 2012, Preis i dr. 2010, Da i dr. 2011, Saavedra i dr. 2011, Bollen i dr. 2011, itd.). Ukoliko se ustanovi da je moguće predviđati kretanja cijene ili rizika dionice nekog poduzeća, dolazi do narušavanja (slabog) oblika hipoteze efikasnog tržišta (Fama 1965, 1970). Predvidljivost prinosa dionica nije novost; već se niz desetljeća uočavaju brojne anomalije u kretanjima cijena i prinosa, s obzirom na ona ponašanja koja se predviđaju u okviru racionalnih financijskih modela. Tako su poznate anomalije vezane uz kalendarske učinke (Škrinjarić 2012), anomalije vezane uz momentum (Cremens i Pareek 2015), itd. Posebice na značaju u posljednja dva desetljeća dobivaju bihevioralni ekonomisti (Vishny, Kanheman, Tversky, itd.; vidjeti Shiller 2003), jer se sve više naglašavaju psihološki i bihevioralni čimbenici kao determinante kretanja cijena dionica, i sve više vjeruju da se buduće cijene dionica donekle mogu predvidjeti temeljem prošlih uzoraka cjenovnih kretanja, kao i što treba inkorporirati psihološke čimbenike u postojeće modele koji ih do sada nisu razmatrali.

Cilj ovoga istraživanja dati je sažet pregled literature vezane uz mogućnost predviđanja cijena dionica na tržištima uz fokus na provođenje empirijskog istraživanja na Zagrebačkoj burzi. Kako se u trenutku pisanja ovoga rada ne pronalaze slična istraživanja ne samo za hrvatsko, već i šire Balkansko područje, postoji nada da se ovim istraživanjem pokriva praznina u postojećoj literaturi te da će se time potaknuti daljnji razvoj literature u ovome pravcu. Fokus će se izvršiti na mogućnost predviđanja pomoću volumena pretraživanja na tražilici *Google*, s obzirom da udio ove tražilice u ukupnom broju svjetskih pretraživanja danas čini gotovo 80% (u kolovozu 2018., vidjeti Netmarketshare 2018; dok je 2013. taj iznos bio oko 67%, Lee 2013). Glavne istraživačke hipoteze su sljedeće. Postoji velik obujam empirijske literature koja nastoji istražiti vezu između Internet pretraživanja, posebice pomoću tražilice *Google* i cijena (prinosa) financijske imovine. Druga hipoteza odnosi se na rezultate empirijskog istraživanja nad hrvatskim dioničkim tržištem: rezultati pretraživanja pomoću Interneta doprinose prognoziранju budućeg prinosa i rizika na Zagrebačkoj burzi i time omogućavaju brže

ostvarivanje investitorovih ciljeva glede prinosa i rizika. Na taj način će se dobiti osnovni uvidi u mogućnosti daljnjeg analiziranja tržišta dionica u ovome pravcu na CEE (*Central and Eastern European*) tržištima, kao i na tržištima Jugoistočne Europe. Ako se ukažu mogućnosti budućeg predviđanja cijena, prinosa i/ili rizika na dioničkim tržištima, postavljaju se daljnje mogućnosti iskorištavanja rezultata takvih analiza.

Struktura rada je sljedeća. Drugo poglavlje daje pregled relevantnije literature vezane uz istraživanje provedeno u ovome radu. Potom se u trećem istraživanju opisuje metodologija empirijske analize u četvrtome poglavlju. Posljednje, peto poglavlje zaključuje članak s preporukama za daljnja istraživanja.

2. PREGLED LITERATURE

Promatrajući postojeća istraživanja koja se usmjeravaju na prognoziranja pomoću volumena pretraživanja *online*, može se donijeti nekoliko zaključaka. Većina interesa u ovome području javlja se tek u posljednjem desetljeću, što znači da postoje još brojna neodgovorena pitanja, razvoj teorije, te empirijska istraživanja. Nadalje, ponovno je većina interesa usmjerena na razvijenije zemlje i tržišta (poput SAD-a, Njemačke, UK-a, Francuske i slično). Stoga postoji jaz u literaturi vezano uz zemlje i tržišta poput Hrvatske. Interes za uključivanjem *online* pretraživanja u predviđanje (makro)ekonomskih varijabli je sve veći i širi se ne samo na tržišta dionica i predviđanja prinosa i rizika, već i na mogućnosti kratkoročnog predviđanja dolazaka broja turista, stope nezaposlenosti, količine maloprodaje specifičnih proizvoda (automobili i slično), itd. Metodologija koja se primjenjuje u istraživanjima je doista raznolika, u ovisnosti s kakvim podacima i njihovim frekvencijama se raspolože. Tako se razmatraju regresijski, GARCH, (S)ARMA¹ modeli, VAR² i VEC³ modeli te PCA⁴ analiza, testira se Grangerova uzročnost, Kalmanov filter, diskriminacijska analiza, itd. Osnovna ideja u većini istraživanja jest kako kvantificirati varijablu koja će doprinijeti kvaliteti modeliranja te hoće li kratkoročno prognoziranje prinosa i/ili rizika biti uspješnije uz uključivanje takve varijable. Većina radova koristi normalizirane vrijednosti *Google* pretraživanja na stranici *Google Trends*, pri čemu se uspoređuje kvaliteta modela bez i s tom varijablom. Kako je literatura sve obilnija, ovdje se fokus vrši na povezana istraživanja koja se usmjeravaju na predviđanje cijena/prinosa/rizika dionica. Na kraju poglavlja kratko su prikazane analize koje za druge ekonomske varijable

¹ Sezonske ARMA modele.

² *Vector Autoregression*.

³ *Vector Error Correction*.

⁴ *Principal Component Analysis*.

razmatraju volumen pretraživanja kao značajni faktor prilikom modeliranja i posebice prognoziranja.

Mondria i dr. (2010) nisu razmatrali tipični pristup uključivanja varijable pretraživanja u neki od modela predviđanja prinosa ili rizika, već su razvili model predviđanja ponašanja ljudi u *online* pretraživanju vezanim uz držanje međunarodnog portfelja. Dodatno su koristili AOL bazu pretraživanja gdje je tada bilo dostupno 21 milijun pretraživanih pitanja od strane preko 657 tisuća korisnika. U rezultatima analize pronalaze da ljudi pretražuju općenito o onim zemljama više u kojima drže određenu financijsku imovinu, kao i da više ulažu u one zemlje o kojima su pretražili više informacija. Da i dr. (2010) također rade dodatan korak, tako da su konstruirali varijablu koju su nazvali *FEARS* indeks, kojom su mjerili sentiment investitora s obzirom na njihovo *online* pretraživanje pozitivnih i negativnih vijesti. Rezultati provedenog istraživanja za razdoblje srpanj – prosinac 2009. godine ukazali su da *FEARS* indeks uspješno predviđa kratkoročne preokrete (engl. *reversals*) u prinosima dionica, kao i privremena povećanja volatilnosti. Da i dr. (2011) su na uzorku od Russell 3000 dionica za razdoblje od 2004. do 2008. godine (tjedni podaci) utvrdili da volumen pretraživanja na tražilici *Google* pozitivno koreliraju s cijenama dionica i pažnjom investitora na dioničkim tržištima. Volumen pretraživanja čak do dva tjedna uspješno predviđa kretanja cijena dionica, kao i što doprinosi nadprosječno velikom prinosu prilikom inicijalne javne ponude dionica poduzeća. Ovo je veoma opsežno istraživanje na koje se zainteresirani čitatelji upućuju za daljnje izučavanje, s obzirom da razmatra i teorijsku stranu priče detaljno.

Yang (2011) razmatra zasebno cijene sektora tehnologije, energije te financija s *Yahoo! Finance* stranice te kako pretraživanja vezana uz recesiju utječu na njihove volatilnosti, za razdoblje od siječnja 2004. do studenog 2011. godine. Metodološki gledano, procjenjuje logističku regresiju i Gausijsku diskriminacijsku analizu, i pokazuje uspješnost modeliranja uz uključivanje pretraživanja u samu analizu. Smith (2012) uključuje volumen pretraživanja u GARCH⁵ modele predviđanja rizika tečaja američkog i kanadskog dolara, eura te japanskog jedna (siječanj 2004. do prosinac 2010. godine je promatran raspon u analizi), pri čemu su rezultati analize robusni. Preis i dr. (2013) nad podacima DJIA indeksa analiziraju pretraživanja vezana uz dug u SAD-u. Rezultati empirijske analize za razdoblje siječanj 2004. do veljače 2011. godine ukazali su na prediktivnu moć pretraživanja vezanih uz pretraživanja o dugu i smanjenja cijena dionica te autori zaključuju kako je moguće konstruirati investicijske strategije koje će moći zarađivati temeljem ovakvih rezultata. Balakrishnan i Dixit (2013) nad podacima za S&P500 Indeks (za vremenski raspon siječanj 2012. – studeni 2015. godine) pronalaze da uključivanje pretraživanja vezanih uz 14 najčešćih riječi vezanih uz

⁵ *Generalized AutoRegressive Conditional Heteroskedasticity.*

financije doprinose prognoziranju volatilnosti na tržištu. Brochado (2016) želi definirati pozornost investitora (engl. *investor attention*) pomoću *online* pretraživanja na portugalskom tržištu u razdoblju ožujak 2006. do veljača 2014. godine. Kao i prethodna istraživanja, i ovo ukazuje na korisnost uključivanja varijable pretraživanja u analizu, kao i robusnost samih rezultata. Perlin i dr. (2017) se fokusiraju na četiri zemlje (UK, SAD, Australiju i Kanadu) kako bi ispitali učinke pretraživanja ključnih riječi koje koriste u sebi riječ *dionica* te dodatno uspoređuju različite investicijske strategije temeljene na rezultatima procijenjenih modela. Autori nalaze kao i većina istraživanja do sada da modeliranje uz uključivanje rezultata pretraživanja doprinosi boljem prognoziranju kretanja na financijskim tržištima, kao i uspješnijem menadžmentu portfelja. Kako su istraživanja sve brojnija, a rezultati i zaključci su slični, ovdje se u nastavku samo navode još neki radovi za zainteresirane čitatelje: Bank i dr. (2011) istražuju njemačko tržište, Dzielinski (2012) američko, australsko, kanadsko, njemačko, englesko i japansko tržište, Zhang i dr. (2013) šangajsko tržište, Aouadi i dr. (2013) francusko tržište, Takeda i Wakao (2014) japansko tržište, itd.

Osim *Google* rezultata pretraživanja, koriste se i druge *online* platforme, poput *Twitter*-a. Tako Bollen i dr. (2011) temeljem javno dostupnih objava na toj platformi vezanih uz dionice formiraju varijable kojima se mjeri sentiment javnosti vezan uz dionice. Za razdoblje od veljače do prosinca 2008. godine u predviđanju kretanja DJIA indeksa uključuju sentiment javnosti te pokazuju kako 3-4 dana unaprijed taj sentiment nagoviješta kretanja na burzi. Detaljniji pregled sličnih istraživanja moguće je naći u Nisar i Yeung (2018). Čak se razmatraju predviđanja i drugih (makro)varijabli s obzirom na rezultate pretraživanja *online*, s obzirom na korisnu svrhu takvih pretraživanja u području maloprodaje, predviđanja nezaposlenosti i slično. Tako Askitas i Zimmermann (2009) na primjeru njemačkih podataka nezaposlenosti (siječanj 2004. do travanj 2009. godine) kako određene ključne riječi vezane uz nezaposlenost doprinose predviđanju buduće nezaposlenosti ako se dodaju u SARIMA modele. Sve je veći obujam literature vezan uz predviđanje dolazaka turista, s obzirom na značaj za BDP neke zemlje. Primjerice, Onder (2017) koristi rezultate tražilice *Google* ne samo vezano uz riječi, već i uz slike za prognoziranje dolaska turista u Beč i Barcelonu, te Austriju i Belgiju. Predviđanje makroekonomskih statistika razmatrano je u Ettredge i dr. (2005); inflacije u Guzman (2011); odlazak u kino i prodaju kino ulaznica u Judge i Hand (2010); kupovina automobila u Carriere-Swallow i Labbe (2013), predviđanje tečaja uz makroekonomske fundamente i uključivanje *Google Trend* podataka u Yaganti i Manpuria (2018), itd. S obzirom na brojnost pozitivnih rezultata glede uključivanja takvih varijabli u modeliranje, za očekivati je i u nastavku ovog istraživanja kako će volumen pretraživanja doprinijeti kvalitetnijem modeliranju i prognoziranju prinosa i rizika na Zagrebačkoj burzi. U nastavku su prikazani metodologija i rezultati empirijskog istraživanja.

3. METODOLOGIJA

U svrhe osnovnog predviđanja prinosa i rizika dionica na Zagrebačkoj burzi, u ovome radu će se koristiti jednostavnija metodologija uobičajena za prognoziranje prinosa i rizika. S obzirom na turbulencije na dioničkim tržištima, kratkoročne prognoze su korisnije u odnosu na dugoročne, pa su odabrani modeli koji upravo služe u tu svrhu. Radi se o ARMA-GARCH modelima (engl. *AutoRegressive Moving Average, Generalized AutoRegressive Conditional Heteroskedasticity*) koji su razvijeni u svrhu kratkoročnog predviđanja. Općenit zapis ARMA(p, q) modela za stacionaran vremenski niz y_t definira se kao (Alexander 2009):

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (1)$$

gdje p predstavlja broj pomaka varijable y_t , a q broj pomaka slučajne varijable ε_t . Kompaktno se model dan u (1) može zapisati pomoću operatora pomaka

$\phi(B) = (1 - \sum_{i=1}^p \phi_i B^i)$, $B y_t = y_{t-1}$, $\theta(B) = (1 - \sum_{i=1}^q \theta_i B^i)$, $B \varepsilon_t = \varepsilon_{t-1}$ na sljedeći način:

$$\phi(B) y_t = \theta(B) \varepsilon_t. \quad (2)$$

Ukoliko je model u (2) stacionaran i invertibilan, moguće je predviđati buduća kretanja vremenskog niza y_t . Generalni zapis GARCH(p, q) modela prikazan je u (3):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2, \quad (3)$$

gdje se ε_{t-i}^2 odnose na šokove s tržišta u prethodnim razdobljima, dok σ_{t-i}^2 se odnose na autoregresivne vrijednosti prethodne volatilnosti. Naravno, uobičajene pretpostavke na parametre u modelu (3) vrijede za striktnu pozitivnost uvjetne varijance ($\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0 \forall i$). Pronalazak optimalnog modela kojim će se prinos i rizik dionice modelirati pomoću (2) i (3) uobičajeno se provodi usporedbom vrijednosti optimalne vrijednosti funkcije vjerodostojnosti (engl. *Log Likelihood*), informacijskih kriterija (AIC, SIC, HQC) te provjerom ostale dijagnostike modela.

Varijabla koja će se uključiti u modeliranje i prognoziranje prinosa i rizika dionica jest volumen pretraživanja na *Google* tražilici (engl. *Google Search Volume*). Radi se o broju pretraživanja određene riječi na određenoj geografskoj lokaciji u danom vremenskom rasponu. Rezultati koji se mogu pronaći na *Google Trends* stranici su normalizirane vrijednosti u rasponu [0, 100], pri čemu veća vrijednost znači da je određena riječ pretraživana više u promatranom rasponu i geografskoj

lokaciji u odnosu na druge. Stoga će se modeli (2) i (3) proširiti za varijablu rezultata volumena pretraživanja⁶ ($search_t$) i njenim pomacima:

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^r \gamma_i search_{t-i} + \varepsilon_t \quad (4)$$

i

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2 + \sum_{i=1}^r \delta_i search_{t-i} . \quad (5)$$

4. REZULTATI EMPIRIJSKE ANALIZE

Za potrebe empirijske analize sa web stranica *Google Trends* (2018) prikupljeni su podaci o ukupnom volumenu traženja ključnih riječi vezanih uz Zagrebačku burzu i indeks CROBEX. Odabran je najdulji mogući vremenski raspon, stoga su u analizu uključeni mjesečni podaci za razdoblje od siječnja 2004. do rujna 2018. godine. *Google Trends* prikuplja podatke o volumenu traženja određenih riječi kroz neko vremensko razdoblje te se volumen traženja normalizira za određeno geografsko razdoblje i vremenski raspon. Vrijednosti su normalizirane u raspon [0, 100], pri čemu veća vrijednost označava veću relativnu važnost traženja određene riječi. U trenutku pretraživanja koje su ključne riječi najčešće tražene preko tražilice *Google* vezane uz Zagrebačku burzu radilo se o sljedećim riječima:

Tablica 1. Koeficijenti korelacije

	burza	crobex	crobex_ dionice	dionice	fondovi	hrportfolio	limun	portfolio	indeks
burza	1								
crobex	0,474***	1							
crobex_ dionice	0,064	0,099	1						
dionice	0,698***	0,348***	-0,023	1					
fondovi	0,306***	0,292***	-0,078	0,670***	1				
hrportfolio	0,748***	0,607***	-0,094	0,720***	0,344***	1			
limun	0,574***	0,750***	-0,108	0,589***	0,551***	0,368***	1		
portfolio	0,739***	0,556***	-0,033	0,761***	0,813***	0,932***	0,818***	1	
indeks	0,669***	0,333***	-0,098	0,643***	0,376***	0,370***	0,648***	0,778***	1

Napomena: *** označava statističku značajnost na razini 1%.

Izvor: izračun autorice

⁶ Odnosno promjene volumena pretraživanja ako je originalna vremenska serija nestacionarna.

⁷ Radi se o portalu koji se bavio financijskim analizama i vijestima do kraja 2017. godine, a nakon toga je integriran u web stranice Raiffeisen banke.

Za daljnju analizu prikupljene su mjesečne vrijednosti indeksa CROBEX te su izračunati kontinuirani prinosi. Prije same procjene odabranih modela prognozi- ranja prinosa i rizika, proveden je test jediničnog korijena (ADF) nad logaritmiranim vrijednostima indeksa CROBEX te u prvim diferencijama, kao i za varijablu *search*. Rezultati su prikazani u tablici 2, gdje je vidljivo kako su oba vremenska niza integrirani prvoga reda. Stoga se daljnja analiza nastavila s vrijednostima u prvim diferencijama.

Tablica 2. Rezultati testa jediničnog korijena (ADF) za varijable u razinama i prvim diferencijama

Varijabla	Razina, kon- stanta i trend	Razina, konstanta	Prva diferencija, konstanta	Prva diferencija
Log(CROBEX _t)	-2,679 (0,247)	-2,504 (0,116)	-10,628*** (0,000)	-10,632*** (0,000)
Log(SEARCH _t)	-3,322 (0,066)	-2,692 (0,077)	-15,705*** (0,000)	-15,750*** (0,000)

Napomena: *** označava statističku značajnost na razini 1%. *p*-vrijednosti dane su u zagradama.

Izvor: izračun autorice

U prvome koraku procijenjeno je nekoliko specifikacija ARMA-GARCH modela do zaključno prosinca 2017. godine, pri čemu se razmatrao princip parsimonije. Rezultati su detaljno prikazani u tablici 3. Može se uočiti da je najbolji model ARMA(1,1)-ARCH(1) s obzirom na većinu pokazatelja. Stoga se u idućem koraku modeliranja pristupilo uključivanju varijable volumena pretraživanja na način da se u spomenuti model uključila trenutna vrijednost varijable pretraživanja i u jednadžbu prinosa i rizika te s jednim pomakom. Rezultati tog modela (M1) prikazani su u prvome stupcu tablice 4. Kako volumen pretraživanja nije bio statistički značajan s prvim pomakom u jednadžbi varijance, u modelu (M2) je isključen taj pomak. Sada je u modelu (M2) postala i trenutna vrijednost volumena pretraživanja neznačajna, kao i lagirana vrijednost u jednadžbi prinosa. Konačno se procijenio model (M3) samo s trenutnom vrijednosti volumena pretraživanja u jednadžbi prinosa. Kako je dijagnostika trećeg modela pokazana najboljom pri usporedbi s prva dva, odabran je treći kao najbolji. Ono što se može uočiti u sva tri modela jest pozitivan učinak volumena pretraživanja na prinose dionica na Zagrebačkoj burzi. Dakle, porast volumena pretraživanja vodi do povećanja prinosa. Unatoč statističkoj neznačajnosti te varijable u jednadžbi varijance, predznak parametra uz volumen pretraživanja je pozitivan, što znači da ipak dolazi i do povećanja volatilnosti dionica kao posljedica povećanog pretraživanja *online*. Rezultati su u skladu s očekivanjem.

Tablica 3. Rezultati procjena modela za prognoziranje prinosa i rizika bez *Google trends* variable

Parametar/ dijagnostika	AR(1)- ARCH(1)	MA(1)- ARCH(1)	ARMA(1,1) -ARCH(1)	ARMA(1,1)- GARCH(1,1)	AR(1) - GARCH(1,1)	MA(1)- GARCH(1,1)
$\hat{\mu}$	0,004 (0,224)	0,004 (0,233)	0,003 (0,781)	0,001 (0,897)	0,004 (0,225)	0,004 (0,246)
$\hat{\phi}_1$	0,111 (0,192)	–	0,981 (0,000)	0,984 (0,000)	0,114 (0,183)	–
$\hat{\theta}_1$	–	–0,072 (0,403)	–0,942 (0,000)	–0,962 (0,000)	–	0,095 (0,276)
$\hat{\alpha}_0$	0,003 (0,070)	0,003 (0,069)	0,002 (0,038)	0,001 (0,161)	0,001 (0,232)	0,001 (0,245)
$\hat{\alpha}_1$	0,662 (0,130)	0,718 (0,125)	0,791 (0,056)	0,492 (0,099)	0,433 (0,167)	0,437 (0,179)
$\hat{\beta}_1$	–	–	–	0,389 (0,066)	0,487 (0,008)	0,506 (0,004)
Dof t	2,959 (0,002)	2,965 (0,002)	3,241 (0,004)	3,160 (0,002)	2,993 (0,003)	2,948 (0,003)
Log L	269,867	272,079	276,136	278,491	274,026	276,561
AIC	–3,0270	–3,0350	–3,0873	–3,1028	–3,0632	–3,0746
SIC	–2,9366	–2,9449	–2,9788	–2,9762	–2,9546	–2,9665
HQC	–2,9904	–2,9984	–3,0433	–3,0514	–3,0191	–3,0307
Q–stat (6)	15,112 (0,010)	16,926 (0,005)	9,343 (0,053)	8,918 (0,063)	13,835 (0,017)	14,840 (0,011)
Q–stat ² (6)	1,219 (0,943)	1,391 (0,925)	0,623 (0,960)	2,005 (0,735)	3,458 (0,628)	3,780 (0,582)

Napomena: Dof t odnosi se na broj stupnjeva slobode procijenjenih za Studentovu distribuciju, Log L na maksimalnu vrijednost funkcije vjerodstoynosti, AIC, SIC i HQC na Akaikeov, Schwartzov i Hannan-Quinn informacijski kriterij, Q-stat (6) test autokorelacije reziduala do zaključno 6. pomaka te Q-stat² (6) na test heteroskedastičnosti reziduala do zaključno 6. pomaka.

Izvor: izračun autorice

Tablica 4. Rezultati procjena modela za prognoziranje prinosa i rizika uz uključenje *Google* volumena pretraživanja

Parametar/ dijagnostika	M1	M2	M3
$\hat{\mu}$	0,009 (0,273)	0,008 (0,018)	0,005 (0,546)
$\hat{\phi}_1$	0,943 (0,000)	-0,458 (0,143)	0,978 (0,000)
$\hat{\theta}_1$	-0,880 (0,000)	0,519 (0,094)	-0,953 (0,000)
$\hat{\gamma}_1$	0,048 (0,022)	0,070 (0,000)	0,050 (0,004)
$\hat{\gamma}_2$	0,036 (0,021)	0,038 (0,036)	–
$\hat{\alpha}_0$	0,001 (0,028)	0,001 (0,443)	0,001 (0,182)
$\hat{\alpha}_1$	0,240 (0,010)	0,416 (0,408)	0,487 (0,143)
$\hat{\beta}_1$	0,477 (0,004)	0,535 (0,005)	0,367 (0,095)
$\hat{\delta}_1$	0,003 (0,001)	0,005 (0,448)	–
$\hat{\delta}_2$	0,0003 (0,727)	–	–
Dof t	6,334 (0,022)	2,527 (0,002)	2,948 (0,002)
Log L	260,931	259,982	260,395
AIC	-3,0295	-3,0300	-3,0409
SIC	-2,8224	-2,8419	-2,8909
HQC	-2,9454	-2,9537	-2,9800
Q-stat (6)	5,405 (0,248)	10,899 (0,028)	8,386 (0,088)
Q-stat ² (6)	5,241 (0,263)	6,034 (0,197)	2,300 (0,681)

Napomena: Dof t odnosi se na broj stupnjeva slobode procijenjenih za Studentovu distribuciju, Log L na maksimalnu vrijednost funkcije vjerodostojnosti, AIC, SIC i HQC na Akaikeov, Schwartzov i Hannan-Quinn informacijski kriterij, Q-stat (6) test autokorelacije reziduala do zaključno 6. pomaka te Q-stat² (6) na test heteroskedastičnosti reziduala do zaključno 6. pomaka

Izvor: izračun autorice

Kako bi se ispitala hipoteza o poboljšanju prognoziranja prinosa i/ili rizika uključivanjem volumena pretraživanja na *Google* tražilici, provedene su prognoze van uzorka, za razdoblje od siječnja do rujna 2018. godine. Na taj način se dobivaju uvidi u mogućnosti korištenja ovakve varijable u dinamičkoj optimizaciji portfelja i predviđanju budućih kretanja na burzi. Temeljem rezultata u tablici 5, gdje su uspoređene standardne mjere usporedbe točnosti prognoze, uočava se da je u 7 od 8 slučajeva model s uključenim volumenom pretraživanja točniji u prognoziranju. Dakle, u promatranom razdoblju se čini da uključivanje volumena pretraživanja doprinosi točnijim prognozama. Dodatno se grafički usporedilo kretanje prognoziranje prinosa CROBEX-a temeljem originalnog ARMA(1,1)-ARCH(1) modela (oznaka *prinos_f_1*) te pomoću modela (M3) (oznaka *prinos_f_2*) na slici 1. Može se uočiti kako prognoziranje uz uključivanje varijable volumena pretraživanja predviđa puno uspješnije kretanje stvarnog prinosa (oznaka *prinos*), kao i smjera kretanja (predznaci).

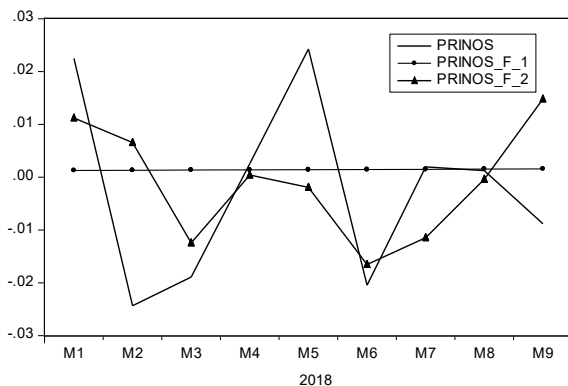
Tablica 5. Usporedba prognozi za originalni model i model uz uključen volumen pretraživanja, prognoziranje van uzorka

Mjera	Originalni model	Model uz <i>Google</i> volumen pretraživanja
RMSE	0,01709	0,01690
MAE	0,01380	0,01329
MAPE	79,9094	168,4198
Theileov indeks nejednakosti	0,93646	0,62302
Udio pristranosti	0,04533	0,00454
Udio varijance	0,94328	0,14651
Udio kovarijance	0,01139	0,84895

Napomena: RMSE, MAE, MAPE označavaju sljedeće mjere: korijen srednje kvadratne pogreške, srednja apsolutna pogreška, srednja apsolutna postotna pogreška (engl. *Root Mean Squared Error*, *Mean Absolute Error*, *Mean Absolute Percentage Error*). Udio pristranosti, varijance i kovarijance mjere koliko je prosjek prognoze udaljen od stvarnog prosjeka, varijanca prognoze i preostala udaljena stvarne varijance i nesistematična greška prognoze koja je preostala. Za uspješno prognoziranje udio kovarijance treba biti najveći.

Izvor: izračun autorice

Grafikon 1. Kretanje vrijednosti prognoziranih prinosa za originalni model te uz uključen volumen pretraživanja

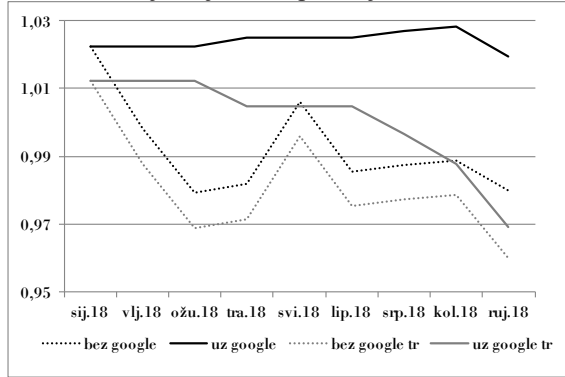


Izvor: izračun autorice

Kako bi se uvidjele potencijalne koristi za investitore, simulirano je nekoliko strategija trgovanja u razdoblju van uzorka temeljem prognoza sa grafikona 1. U svakoj strategiji se pretpostavilo da investitor na početku razdoblja ulaže jednu novčanu jedinicu. Pretpostavka u svakoj strategiji je da investitor kupuje indeks ako je prognozirani prinos u mjesecu negativan jer time kupuje indeks pri manjim cijenama dionica, a da prodaje kad se prognozira pozitivan prinos kako bi prodao portfelj po većim cijenama. Dodatno, ako investitor već raspolaže s CROBEX-om, te prognozira negativan prinos, držat će i dalje takav portfelj dok se ne ostvari pozitivan prinos. Konačno, simulirane su strategije s uključenim transakcijskim troškovima, tako da je od vrijednosti portfelja pri svakoj strategiji oduzeto 1% vrijednosti portfelja. Razmatraju se 4 strategije naziva: *bez Google* (strategija temeljem prognoza ARMA(1,1)-ARCH(1) modela); *uz Google* (strategija modela M3 uz volumen pretraživanja); te obje uz uključene transakcijske troškove: *bez Google tr* i *uz Google tr*. Kretanja vrijednosti svih portfelja predočene su na grafikonu 2. Ako se usporede prve dvije spomenute strategije, može se uočiti da je u svakom mjesecu vrijednost portfelja uz uključen volumen pretraživanja (strategija *uz Google*) veća u odnosu na početnu strategiju temeljenu samo na ARMA-GARCH modeliranju. Radi se o izuzetno pozitivnoj vijesti za investitore. Dodatno, ovom strategijom je ostvarena veća vrijednost portfelja u odnosu na inicijalno uloženu jednu jedinicu, dok je strategija bez volumena pretraživanja rezultirala s gubitkom (vidjeti krivulju *bez Google*). Ako se uključe transakcijski troškovi u analizu, sada obje strategije ostvaruju gubitak u odnosu na inicijalno uloženi novac, ali je manji gubitak ostvaren u slučaju s uključenim volumenom trgovanja (*uz Google tr*). Dakle, u svakome slučaju, uključivanje volumena pretraživanja u prognoziranje prinosa smanjuje gubitke kada tržište

„pada“ i sačuva vrijednost portfelja više u odnosu na strategije bez uključivanja *Google* pretraživanja.

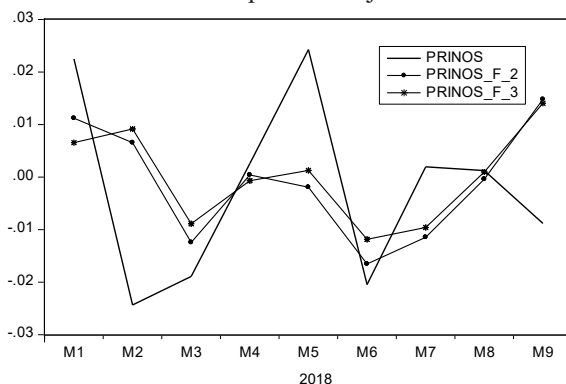
Grafikon 2. Kretanje vrijednosti portfelja za odabrane strategije



Izvor: izračun autorice

Konačno, kako su procijenjeni modeli (M1) i (M2) gdje je volumen pretraživanja bio uključen i u modeliranje rizika, obavljene su prognoze van uzorka i u slučaju ta dva modela, što je prikazano na grafikonu 3. Može se uočiti kako i u ovome slučaju dolazi do relativno dobrog prognoziranja barem predznaka kretanja prinosa. Kako je gledajući ukupnu dijagnostiku bio bolji model (M2), on je odabran za usporedbu strategija investiranja u odnosu na model *uz Google* s prethodne slike.

Grafikon 3. Kretanje vrijednosti prognoziranih prinosa za originalni model te uz uključen volumen pretraživanja i u rizik

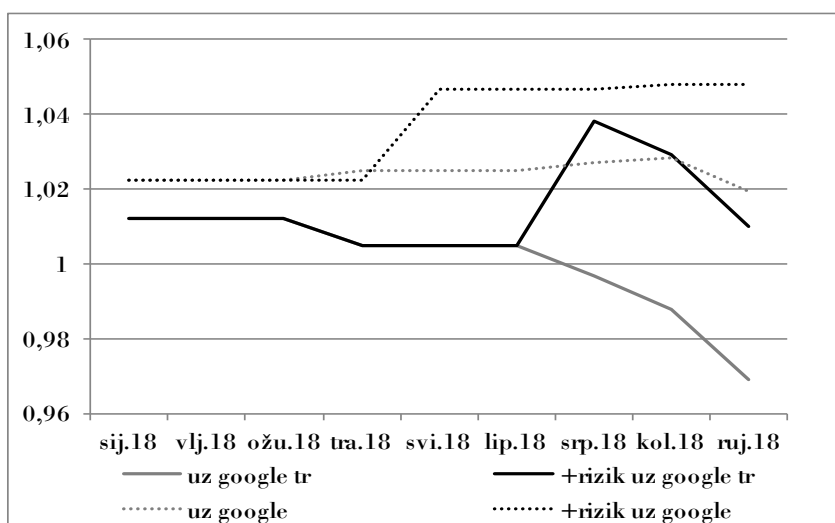


Izvor: izračun autorice

Sada su rezultati usporedbe prikazani na grafikonu 4. Ako se najprije usporede vrijednosti portfelja bez uključenih transakcijskih troškova, uočava se da od travnja 2018. godine do kraja promatranog razdoblja portfelj za model (M2) ostvaruje

veću vrijednost. Dakle, može se zaključiti da unatoč statističkoj neznačajnosti varijable volumena pretraživanja u procjeni samoga modela (M2), korisno je ostaviti ovu varijablu jer očito doprinosi uspješnijem menadžmentu portfelja. Sličan zaključak vrijedi i za uključivanje transakcijskih troškova u analizu. Iako se vrijednosti oba portfelja u tom slučaju podudaraju do lipnja 2018. godine zbog veoma sličnih prognoza za to razdoblje, nakon toga se ponovno uočava da je vrijednost portfelja za model (M3) veća do kraja razdoblja. I ovo je jedini model u kojemu i uz uključivanje transakcijskih troškova dolazi do povećanja vrijednosti portfelja na kraju promatranog razdoblja.

Grafikon 4. Kretanje vrijednosti portfelja za odabrane strategije



Izvor: izračun autorice

Temeljem provedene analize može se iznijeti nekoliko zaključaka. Najprije, s obzirom na današnje računalne mogućnosti, uključivanje dodatnih informacija u postojeće financijske i ekonometrijske modele ne bi smio istraživačima i investitorima predstavljati problem. Kako su prethodna empirijska literatura te rezultati u ovome istraživanju pokazali korisnost uključivanja volumena pretraživanja na tražilici *Google*, može se preporučiti uključivanje ovakve varijable u analizama u svrhu uspješnijeg menadžmenta portfelja. Nadalje, rezultati simuliranih strategija ukazuju da se može ostvariti povećanje vrijednosti portfelja na kraju razdoblja držanja ako se u prognoziranje prinosa i rizika uključi volumen pretraživanja na *online* tražilicama.

5. ZAKLJUČAK

Uspješan menadžment portfelja danas predstavlja veoma velik izazov. Jedan od novijih aspekata poboljšanja postojeće prakse u posljednjih nekoliko godina odnosi se na uključivanje volumena *online* pretraživanja u području financija, investiranja i dioničke burze u predviđanje prinosa i rizika dionica. Stoga ne čudi što se u posljednjem desetljeću značajno povećao broj empirijskih istraživanja koja nastoje konstruirati varijable pretraživanja i sentimenta investitora kako bi se što uspješnije daljnja analiza mogla provesti. U ovome istraživanju nastojao se dati kritički osvrt postojeće literature kako bi zainteresirani čitatelji mogli dobiti predodžbu o mogućnostima primjena nad stvarnim podacima, s obzirom da ne postoje slična istraživanja na širem Balkanskom području. Rezultati empirijske analize ukazuju da se vrijednost portfelja može povećati temeljem uključivanja volumena pretraživanja kao jedne od varijable prognoziranja prinosa i rizika. Odabran pristup modeliranja je jednostavan, međutim, ARMA-GARCH modeli predstavljaju osnovicu u modeliranju i prognoziranju prinosa i rizika dionica veoma korisnim za kratkoročne prognoze.

Nedostaci provedenog istraživanja očituju se u sljedećem. Nedostupnost podataka je čest problem u empirijskim analizama za tržišta poput Hrvatskog. Stoga su se analizirali mjesečni podaci, a investitorima bi od veće koristi u svrhu dinamičke optimizacije portfelja vjerojatno značili tjedni podaci. Dodatno, razmatrani su osnovni modeli prognoziranja prinosa i rizika, bez razmatranja drugih mogućnosti ulaganja u drugačije oblike financijske imovine. Međutim, u radu se nastojalo utvrditi ima li uopće pretraživanje pomoću tražilice *Google* prediktivnu moć na hrvatskom tržištu dionica. Stoga je temelj postavljen za buduće analize koje će nastojati ukloniti nedostatke ove. Nadalje, buduća istraživanja razmotrit će i druge modele u okviru kojih se ovakva varijabla može uključiti u modeliranje kako bi se dobili još robusniji i pravovremeni rezultati. Dodatno, nastojat će se istraživanja i usmjeriti na detaljnije formiranje i razmatranje investicijskih strategija s uključenjem volumena pretraživanja kao jednog od prediktora prinosa i rizika kako bi investitori ostvarili svoje ciljeve još brže i pouzdanije, te kao bi se potakla zamrla investicijska aktivnost na Zagrebačkoj burzi, ali i okolnim tržištima.

LITERATURA

1. Alexander, C. (2009), *Market risk analysis*, Volume 2, UK: John Wiley & Sons.
2. Aouadi, A., Aroudi, M., Teulon, F. (2013), *Investor attention and stock market activity: evidence from France*, *Economic Modeling*, 35, str. 674-681.

3. Askita, N., Zimmermann, K. F. (2009), Google Econometrics and Unemployment Forecasting, *Applied Economics Quarterly*, 55(2), str. 107-120.
4. Balakrishnan, A., Dixit, K. (2013), Predicting Market Volatility Using Semantic Vectors and Google Trends, dostupno na http://cs229.stanford.edu/proj2015/200_report.pdf (pristupljeno 15. rujna 2018.)
5. Bank, M., Larch, M., Peter, G. (2011), Google search volume and its influence on liquidity and returns of German stocks, *Financial Markets and Portfolio Management*, 25(3), str. 239-264.
6. Bollen, J., Mao, H., Zeng, X. (2011), Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, 2(1), str. 1-8.
7. Carriere-Swallow, Y., Labbe, F. (2013), Nowcasting with Google Trends in an Emerging Market, *Journal of Forecasting*, 32, str. 289-298.
8. Choi, H., Varian, H. (2009), Predicting the present with Google Trends, Working paper, Google Inc.
9. Choi, H., Varian, H. (2012), Predicting the Present with Google Trends, *Economic Record*, 88, str. 2-9.
10. Cremens, M., Pareek, A. (2015), Short-Term Trading and Stock Return Anomalies: Momentum, Reversal, and Share Issuance, *Review of Finance*, 19(4), 1 str. 1649–1701.
11. Da, Z., Engelberg, J., Gao, P. (2010), The Sum of All Fears: Investor Sentiment and Asset Prices, SSRN eLibrary.
12. Da, Z., Engelberg, J., Gao, P. (2011), In Search of Attention, *The Journal of Finance*, 66(5), str. 1461-1499.
13. Dzielinski, M. (2012), Measuring economic uncertainty and its impact on the stock market, *Finance Research Letters*, 9, str. 167-175.
14. Ettredge, M., Gerdes, J., Karuga, G. (2005), Using web-based search data to predict macroeconomic statistics, *Communications of the ACM*, 48(11), str. 87-92.
15. Fama, E. F. (1965), Random Walks In Stock Market Prices, *Financial Analysts Journal*, 21(5), str. 55–59.
16. Fama, E. F. (1970), Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), str. 383-417.
17. Google trends (2018), dostupno na <https://trends.google.com/trends> (pristupljeno 15. rujna 2018.).
18. Guzman, G. (2011), Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations, *The Journal of Economic and Social Measurement*, 36(3), str. 119-167.

19. Judge, G., Hand, C. (2010), Searching for the picture: forecasting UK cinema admissions making use of Google Trends data, Department of Economics Discussion Paper No. 162.
20. Lee, J. (2013), Google's Search Market Share Shoots Back to 67%. dostupno na <http://searchenginewatch.com/article/2289560/Googles-Search-Market-Share-Shoots-Back-to-67> (pristupljeno 15. rujna 2018.)
21. Mondria, Jordi, Thomas Wu, and Yi Zhang, 2010, The determinants of international investment and attention allocation: using internet search query data, *Journal of International Economics*, 82, str. 85–95.
22. Netmarketshare (2018), dostupno na <https://netmarketshare.com/search-engine-market-share.aspx> (pristupljeno 15. rujna 2018.)
23. Nisar, T. M., Yeung, M. (2018), Twitter as a tool for forecasting stock market movements: A short-window event study, *The Journal of Finance and Data Science*, 4(2), str. 101-111.
24. Onder, I. (2017), Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities, *International Journal of Tourism Research*, 19, str. 648–660.
25. Perlin, M. S., Caldeira, J. F., Santos, A. A. P. (2017), Can We Predict the Financial Markets Based on Google's Search Queries?, *Journal of Forecasting*, 36(4), str. 454-467.
26. Preis T, Reith D, Stanley H E (2010), Complex dynamics of our economic life on different scales: insights from search engine query data, *Philosophical Transactions of the Royal Society A*, 368, str. 5707–5719.
27. Preis, T., Moat, H. S., Stanley, H. E. (2013), Quantifying Trading Behavior in Financial Markets Using Google Trends, *Scientific Reports*, 3, str. 1-6.
28. Saavedra S, Hagerty K, Uzzi B (2011), Synchronicity, instant messaging, and performance among financial traders, *Proceedings of the National Academy of Sciences*, 108, str. 5296–5301.
29. Shiller, R. J. (2003), From Efficient Markets Theory to Behavioral Finance, *Journal of Economic Perspectives*, 17(1), str. 83-104.
30. Smith, G. P. (2012), Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency, *Finance Research Letters*, 9(2), str. 103-110.
31. Škrinjaric, T. (2012), Kalendarski učinci u prinosima dionica, *Ekonomski pregled*, 63(11), str. 651-678.
32. Takeda, F., Wakao, T. (2014), Google search intensit and its relationship with returns and trading volume of Japanese stocks, *Pacific-Basin Finance Journal*, 27, str. 1-18.

33. Yaganti, H., Manpuria, Y. (2018), Exchange Rate Forecast Enhancement Using Sentiments from Google Trends, *IUP Journal of Financial Risk Management*, 15(2), str. 7-21.
34. Yang, Y. (2011), Learning Stock Volatility Using Keyword Search Volume, dostupno na <https://pdfs.semanticscholar.org/fc38/ce277fe3912be3b6ebf52e2478eb1032ab34.pdf> (pristupljeno 15. rujna 2018.)
35. Zagrebačka burza (2018), dostupno na <https://www.zse.hr> (pristupljeno 15. rujna 2018.)
36. Zhang, W., Shen, D., Zhang, Y., Xiong, X. (2013), Open source information, investor attention and asset pricing, *Economic Modelling*, 33, str. 613-619.

Tihana Škrinjarić, PhD

Faculty of Economics and Business Zagreb, University of Zagreb, Zagreb, Croatia
tskrinjar@net.efzg.hr

CAN GOOGLE TRENDS DATA ENHANCE FORECASTING RETURNS ON ZAGREB STOCK EXCHANGE?

Received: September 19, 2018

Accepted: December 11, 2018

Preliminary communication

Abstract

Successful portfolio management presents a huge challenge nowadays. Today, investors have available a greater amount of investment possibilities, as well as information, mathematical and statistical tools, models and methods. This research seeks to give an overview how to include online search volume on the most popular search engine Google in modelling and forecasting return and risk on the Zagreb Stock Exchange. Based upon monthly data from January 2004 until September 2018, several specifications of ARMA-GARCH models have been estimated: without and with included search volume regarding stocks and Zagreb Stock Exchange. The results of the analysis have shown that models with included search volume are more successful in forecasting. Moreover, the results are confirmed by simulations of investment strategies out of sample, where portfolios based on forecasting with included Google search volume resulted with greater portfolio values. This was confirmed with the inclusion of transaction costs as well. However, there still remain many research possibilities in this area in order to achieve investment goals more quickly and successfully.

Keywords: *Google search engine, return forecasting, Zagreb Stock Exchange, search volume*

JEL: C58, D53, G1