



NÁRODNÁ BANKA SLOVENSKA
EUROSYSTEM

TESTOVANIE PRESNOSTI PREDIKCIÍ

MARIÁN VÁVRA

NETECHNICKÉ
ZHRNUTIE

5/2015



© Národná banka Slovenska
www.nbs.sk
Imricha Karvaša 1
813 25 Bratislava

research@nbs.sk

Júl 2015

ISSN 1337-5830

Práca neprešla jazykovou úpravou.

Prezentované názory a výsledky v tejto štúdií sú názormi autora a nevyjadrujú oficiálne stanovisko Národnej banky Slovenska.

Všetky práva vyhradené.

Krátke časti textu, nie viac ako dva odseky, môžu byť citované bez predchádzajúceho súhlasu autorov, pokiaľ bude úplne uvedený zdroj.



Testovanie presnosti predikcií

Výskumná štúdia NBS

Marián Vávra¹

Abstrakt

V predkladanej štúdii je navrhnutá modifikácia testu Diebolda a Mariana používaného na testovanie presnosti predikčných metód či modelov. V štúdii je demonštrované, že navrhnutá testová štatistika sa dá veľmi jednoducho vypočítať a jej štatistické vlastnosti sú oveľa lepšie, ako má pôvodný test. Z týchto dôvodov je tento nový test veľmi užitočný pre testovanie presnosti predikcií v centrálnych bankách. Obidva testy (t.j. pôvodný aj modifikovaný) sú použité na hodnotenie predikčných chýb 7 kľúčových ukazovateľov ekonomiky USA. Predikcie pochádzajú z dvoch rôznych zdrojov – z jednoduchých modelov časových radov odhadnutých vo Federal Reserve Bank of Philadelphia a z prieskumu prognóz amerických makroekonomických analytikov (SPF). Z empirických výsledkov prekvapivo vyplýva, že analytici nie sú schopní prognózovať kľúčové makroekonomické, ale ani finančné, ukazovatele presnejšie ako jednoduché štatistické metódy.

JEL klasifikácia: C12, C15, C32, C53

Kľúčové slová: Diebold-Marian test, bootstrap, presnosť predikcií

Voľne prístupné na www.nbs.sk/sk/publikacie/vyskumne-studie

¹ Marián Vávra, Odbor výskumu, NBS.



1. NETECHNICKÉ ZHRNUTIE

V súčasnosti niet pochýb o tom, že menová politika musí byť vpred hľadiaca, aby mohla efektívne vplývať na ekonomiku a dosiahnuť svoje ciele. Centrálné banky používajú na prognózovanie kľúčových veličín spravidla celú sadu modelov. Preto vyhodnotenie presnosti prognóz jednotlivých modelov je pre centrálné banky veľmi dôležité. A to nie len pre výber „najlepších“ modelov, ale aj pre ďalší vývoj progresívnych prognostických metód.

V literatúre sa venuje špeciálna pozornosť testovej štatistike Diebolda a Mariana (DM). Aj keď je koncepčne jednoduchá, nenáročná na výpočet a je veľmi populárna vo finančných a makroekonomických aplikáciách, jej štatistické vlastnosti v malých vzorkách sú nepostačujúce. Z toho vyplýva, že DM test príliš často vyvodzuje nesprávne závery z dát, čo v podstate výrazne obmedzuje jeho spoľahlivosť pre empirické aplikácie. Táto štúdia prispieva do literatúry návrhom bootstrapovej verzie DM štatistiky (BDM). Ukazuje sa, že BDM štatistiku sa dá veľmi jednoducho vypočítať a má výborné štatistické vlastnosti aj v malých vzorkách. To znamená, že je spoľahlivým testom pre empirické aplikácie.

Predkladaná štúdia má nasledujúce ciele:

- I) Modifikovať testovú štatistiku, pomocou ktorej sa dá otestovať, ktorý z modelov poskytuje presnejšie predikcie kľúčových ekonomických ukazovateľov.
- II) Porovnať štatistické vlastnosti navrhnutého testu s vlastnosťami testu, ktorý sa v literatúre najčastejšie používa.
- III) Porovnať presnosť predikcií finančných analytikov a jednoduchých modelov časových radov pre 7 kľúčových ukazovateľov ekonomiky USA.



I. Modifikovaný test Diebolda a Mariana

Nech $\{X_{1t}, X_{2t}: t \in \mathbb{Z}\}$ je párom kovariačne stacionárnych a prierezovo korelovaných predikčných chýb pochádzajúcich z dvoch alternatívnych modelov alebo prístupov. Diebold a Mariano (1995)² navrhujú koncepčne jednoduchú štatistiku na testovanie rovnosti predikčnej presnosti dvoch alternatívnych modelov. Pomocou DM štatistiky sa testuje, či populačný priemer stratového diferenciálu $E(d) = E(G(X_{1t}) - G(X_{2t}))$ sa rovná nule, kde $G(x)$ je nejaká stratová funkcia. Autori navrhujú použiť štandardný t -test na testovanie nulovej hypotézy, že $E(d) = 0$, to znamená

$$D = \sqrt{n} \left(\frac{\bar{d}}{\hat{\sigma}} \right) \xrightarrow{d} N(0,1) \text{ pri } n \rightarrow \infty,$$

kde \bar{d} predstavuje vzorkovú analógiu stratového diferenciálu a $\hat{\sigma}$ je rovná dlhodobému rozptylu.

Za poznamenanie stojí fakt, že štatistické vlastnosti DM testu v malých vzorkách (napr. pri $n=50$ pozorovaní) sú slabé. Test trpí významným skreslením v malých vzorkách,³ ktoré sa zvyrazňuje s perzistenciou predikčných chýb a s dĺžkou horizontu prognóz. Hlavným problémom DM testu je, že konzistencia $\hat{\sigma}$ je síce v literatúre dobre podložená, jej hodnota je však vysoko nepresná (vychýlená smerom nadol) v prípade perzistentných stochastických procesov v malých vzorkách. V dôsledku toho konvergencia empirického rozdelenia DM štatistiky do jej limitnej podoby normálneho rozdelenia je pomerne pomalá. Inými slovami, limitné normálne rozdelenie nie je dobrou aproximáciou v malých vzorkách. Prirodzeným riešením tohto problému je technika bootstrapu pre výpočet dlhodobého rozptylu. Detaily o výpočte kritických hodnôt modifikovaného DM testu pomocou bootstrapu (označeného ako BDM) nájde čitateľ v priloženej výskumnej štúdií.

² Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, 13, 253–263.

³ Toto skreslenie znamená, že test nesprávne zamietá nulovú hypotézu príliš často a preto dáva zavádzajúce závery o kvalite prognóz.



II. Štatistické vlastnosti BDM testu

Štatistické vlastnosti navrhnutého BDM testu sú otestované pomocou Monte Carlo experimentov. Výsledky sú porovnané s pôvodným DM testom. Zo simulačných výsledkov vyplývajú nasledujúce závery:

(i) Pre všetky Monte Carlo experimenty platí, že navrhovaný BDM test má empirické úrovne (tzv. size) blízke nominálnym hodnotám 0.10, kým pôvodný DM test vykazuje významné skreslenie aj v pomerne veľkých vzorkách.⁴ Toto skreslenie DM testu spôsobuje najmä perzistencia dát, kým prierezová korelácia medzi predikčnými chybami hrá iba menšiu úlohu.

(ii) Ako sa dalo očakávať, určité skreslenie je pozorovateľné v prípade sily BDM testu (tzv. power) v porovnaní s pôvodným DM testom.⁵ Avšak štatistické vlastnosti BDM testu sa rýchlo zlepšujú s rastúcou veľkosťou vzorky. V každom prípade pozorovaná strata štatistickej sily nemá taký rozsah, aby sa BDM test nestal atraktívnym pre empirické aplikácie.

III. Aplikácia

V tejto časti sa stručne diskutuje otázka, či prognózy makroekonomických analytikov sú presnejšie než jednoduché (autoregresné) modely časových radov. Táto otázka je obzvlášť dôležitá pre centrálnu banku, keďže konsenzus analytikov sa často porovnáva s oficiálnymi prognózami menových autorít. Vychádzame z konsenzuálnej prognózy podľa amerického prieskumu Survey of Professional Forecasters (SPF). Presnosť týchto prognóz porovnáваме s výsledkami lineárnych autoregresných modelov použitím navrhnutého BDM testu.

⁴ Nominálna hodnota je v podstate pravdepodobnosť, že nesprávne zamietneme nulovú hypotézu o rovnosti predikčnej presnosti dvoch metód.

⁵ Sila testu je naopak pravdepodobnosť správneho zamietnutia nulovej hypotézy o rovnosti predikčnej presnosti dvoch metód.



Nulová hypotéza o rovnosti (stredných kvadratických chýb, MSE)

$H_0: MSE(SPF) = MSE(AR)$ oproti $H_1: MSE(SPF) \neq MSE(AR)$,

sa testuje pre nasledujúci súbor predikčných chýb (s horizontom predikcie 1 až 4 štvrťroky dopredu): krátkodobá úroková sadzba, výnos korporátnych dlhopisov, tempo rastu reálneho HDP, tempo rastu deflátoru HDP, tempo rastu priemyselnej produkcie, miera nezamestnanosti a stavebné povolenia.

Medzi závermi BDM a DM testov sú dramatické rozdiely ohľadom zamietnutia nulovej hypotézy (t.j. H_0). Pôvodný DM test nesprávne zamietá nulovú hypotézu zhodnej presnosti dvoch prognóz dva až trikrát častejšie než BDM test. Výsledky podľa spoľahlivejšieho BDM testu naznačujú, že prognózy analytikov sú systematicky presnejšie než predpovede autoregresných modelov iba pre jednu ekonomickú veličinu, ktorou je miera nezamestnanosti. Tento záver nemusí byť príliš prekvapivý z pohľadu vysokej asymetrickej miery nezamestnanosti v priebehu hospodárskych cyklov, čo môže byť lepšie zachytené v prognózach analytikov (podľa predchádzajúcej skúsenosti) než v jednoduchých lineárnych modeloch. Navyše, nenašli sa štatistické dôkazy ani o tom, že prognózy úrokových sadzieb podľa analytikov by boli systematicky lepšie než predpovede jednoduchých modelov. Tento výsledok je pomerne prekvapivý, keďže úrokové sadzby sú kľúčovými veličinami z pohľadu analytikov z finančného sektora.