

Caiete de studii
Nr. 57

Marcel Antonio Sandu ▪ Mădălin Viziniuc

Prognoze în timp real ale creșterii economice

CAIETE DE STUDII

Nr. 57

Mai 2021

NOTĂ

Opiniile prezentate în această lucrare aparţin în întregime autorilor
şi nu implică sau angajează în vreun fel Banca Naţională a României.

Toate drepturile rezervate. Reproducerea informaţiilor este permisă
numai în scopuri educative şi necomerciale şi numai cu indicarea sursei.

ISSN 1584-0883 (versiune *online*)
ISSN 1584-0883 (versiune e-Pub)

Prognoze în timp real ale creșterii economice

Marcel Antonio Sandu*

Mădălin Viziniuc*

* Direcția Studii Economice, Banca Națională a României.

Autorii doresc să mulțumească Elenei Iorga, Mădălinei Movileanu, lui Csaba Bálint, Wilhelm Salater, Andrei Tănase, precum și altor colegi din cadrul BNR pentru sugestiile și observațiile formulate pe parcursul realizării lucrării.

Cuprins

Rezumat 7

1. Introducere 9

2. Aspecte teoretice 11

2.1. Modele punte 11
2.2. Modele MIDAS 13
2.3. Modele cu factori dinamici cu frecvențe mixte 15
2.4. Modele VAR Bayesiene la frecvențe diferite 18

3. Baza de date și designul exercițiilor de prognoză 21

4. Performanța de prognoză 24

4.1. Modele punte 25
4.2. Modele MIDAS 27
4.3. Modele cu factori dinamici cu frecvențe mixte 31
4.4. Modele VAR Bayesiene cu frecvențe mixte 32

5. Combinarea prognozelor 34

5.1. Generarea și combinarea densităților de repartiție a prognozelor 36
5.2. Performanța de prognoză 39
5.3. Robustețea rezultatelor 44

Concluzii și direcții viitoare de cercetare 46

Anexă 49
Bibliografie 52

Rezumat

Lucrarea de față propune un cadru de prognoză pe termen scurt a evoluției trimestriale a PIB care încorporează eficient fluxul de informații oferit de calendarul de publicare a indicatorilor cu frecvență lunară. Metodologia presupune combinarea densităților de repartiție generate de o paletă largă de abordări econometrice (modele punte, MIDAS, precum și modele cu factori comuni dinamici și VAR Bayesian, ambele cu frecvențe mixte), strategia fiind potrivită pentru a fi aplicată la nivelul economiilor emergente, unde activitatea este afectată deopotrivă de modificări structurale și de o dimensiune relativ redusă a setului de informații statistice și a perioadei acoperite de acestea. Procesul de evaluare, care a avut în vedere exerciții succesive de prognoză începând cu primul trimestru din 2013 (utilizând versiunea de conturi naționale disponibilă la acel moment), relevă o performanță sporită a cadrului de a anticipa traiectoria creșterii PIB, acesta prezentând, totodată, și o capacitate ridicată de a prelua schimbările de traiectorie ale indicatorilor lunari, trăsătură care s-a dovedit esențială pentru proiecția PIB pe parcursul contracției economice determinate de criza sanitară în 2020. În același timp, se remarcă un grad ridicat de complementaritate a modelelor econometrice utilizate, excluderea oricăruia dintre ele conducând, în linii mari, la o deteriorare a acurateței de prognoză.

Cuvinte-cheie: prognoză pe termen scurt, prognoză combinată, densitate de repartiție, frecvență mixtă, componentă principală, MIDAS, factor comun dinamic, Bayesian VAR

Coduri de clasificare JEL: C52, C53, E17, E37

1. Introducere

Surprinderea cât mai rapidă a modificărilor intervenite în traiectoria activității economice reprezintă un element important al implementării eficiente a politicilor economice. De regulă, primele evaluări cu privire la PIB, indicatorul-cheie care prezintă evoluția economiei în ansamblu, sunt publicate cu o întârziere de aproximativ 45 de zile de la încheierea trimestrului, în cazul estimărilor *semnal* și aproximativ 70 de zile în cazul datelor *provizoriu I*, în acest interval de timp devenind disponibili, treptat, o serie de indicatori cu frecvență lunară. De pildă, primele informații cu privire la activitatea industrială sunt diseminate cu aproximativ 15 zile înaintea finalizării trimestrului, a doua lună din trimestru este publicată în aproximativ 15 zile de la încheierea acestuia, iar cea de-a treia este furnizată de către Institutul de Național de Statistică (INS) cu câteva zile înaintea estimărilor *semnal* ale PIB.

Datele cu frecvență ridicată conțin informații importante, îmbunătățirea modului de utilizare a acestora având potențialul de a aduce câștiguri notabile la nivelul acurateței proiecțiilor pe termen scurt. Cu toate că sunt disponibile numeroase metodologii statistice care pot încorpora în procesul de prognoză cele mai recente informații (de văzut *inter alia* Camacho, 2013; Foroni și Marcellino, 2013), niciuna dintre acestea nu a dovedit a avea cele mai bune performanțe pe parcursul tuturor fazelor ciclului economic, discuția fiind mult mai aprinsă în cazul economiilor emergente, unde, suplimentar, modificările mai frecvente de structură pot afecta decisiv stabilitatea în timp a performanței de prognoză a unei anumite metode (Clements și Hendry, 1996; Hendry și Clements, 2003; Clark și McCracken, 2010; Rossi, 2013). În acest context, cea mai accesibilă soluție este aceea de a combina proiecțiile generate de mai multe metodologii, fiecare dintre ele prezentând diferite niveluri de sensibilitate la eventuale modificări ale legăturilor dintre indicatori. Astfel, în funcție de performanța de prognoză, acestea ar urma să primească o pondere actualizată în procesul de agregare (de văzut *inter alia* Jore *et al.*, 2010), strategia fiind totodată utilă și în situația în care modelele folosite prezintă diferite grade de eroare la nivelul specificației (Timmermann, 2006).

Prin urmare, lucrarea de față propune un cadru de prognoză pe termen scurt care îmbină *output*-ul generat de mai multe metodologii într-o prognoză unitară, fiind preferată combinarea întregii distribuții (care oferă, de altfel, și o evaluare riguroasă a incertitudinii asociate proiecțiilor) în detrimentul opțiunii clasice, unde sunt vizate doar momentele de ordin I. Metodele considerate acoperă un spectru larg de abordări, începând cu unele relativ simple, care se bazează pe ecuații liniare estimate cu factori comuni (evaluați pe baza metodei componentei principale statice, Stock și Watson, 2002), și continuând cu modele MIDAS (Ghysels *et al.*, 2004; Clements și Galvão, 2009), în care se recurge la funcții polinomiale pentru limitarea numărului de parametri care trebuie estimați. Totodată, sunt incluse și modele cu factori dinamici care utilizează filtrul Kalman (MF-DFM) pentru evaluarea și prelungirea acestora în linie cu orizontul de proiecție ales (Giannone *et al.*, 2008), precum și metode mai complexe care folosesc tehnici Bayesiene de estimare, aplicate într-un context în care sunt prezente multiple variabile neobservabile – modele Bayesian VAR care utilizează seturi de date cu frecvență mixtă (Schorfheide și Song, 2015) și prezintă erori variabile în timp (Carriero *et al.*, 2016). Toate aceste metode sunt aplicate utilizând un set

de indicatori compus din 95 de variabile lunare, care acoperă evoluția majorității ramurilor economice (producție industrială, cifră de afaceri din comerț și servicii, numărul de angajați, câștigul salarial), precum și indicatori financiari (credite, indice bursier) și de conjunctură (încrederea în economie).

Strategia de combinare a densităților de repartiție oferă performanțe solide de prognoză, în multiple situații prezentând cea mai ridicată acuratețe. De menționat că pe parcursul perioadei de evaluare (cuprinsă între T1 2013 și T3 2020), distribuția combinată a prezentat o calitate inferioară a prognozelor, semnificativă din punct de vedere statistic, doar pentru orizontul T+2 și în situația în care creșterea economică a marcat o evoluție extremă în comparație cu media istorică; în acest context, metodologia MF-DFM a consemnat o performanță superioară. În același timp, cadrul rămâne robust la modificarea modalității de calcul a ponderilor de agregare, puterea sa predictivă înrăutățindu-se, în linii mari, doar în situația în care, rând pe rând, câte o metodologie este înlăturată. Totodată, în condițiile în care volatilitatea ridicată a indicilor de volum ai VAB în agricultură constituie o particularitate a economiei locale, excluderea acestui sector din PIB conduce la o îmbunătățire simțitoare a capacității metodologiei de a surprinde tendința activității economice.

Dincolo de aceste aspecte, cadrul de monitorizare în timp real prezintă o relevanță mai ridicată în momentele imediat următoare materializării unor șocuri asupra activității economice. În cel mai recent episod de acest fel, cauzat de pandemia SARS-CoV-2, pe baza evoluțiilor aferente lunii martie (unde erau vizibile primele efecte negative), metodologia de prognoză a actualizat valorile pentru creșterea economică de la o rată de creștere consistentă până la una apropiată de valoarea nulă, concomitent cu o modificare a asimetriei distribuției către stânga acesteia, ceea ce sugerează riscuri de supraestimare – datele publicate ulterior de către INS au confirmat o evoluție modestă a activității economice. Mai departe, în cazul primelor prognoze pentru T2 2020, informațiile încă deficitare cu privire la evoluția sectoarelor economice au condus la o plajă foarte largă de prognoze individuale ale modelelor, însă procedura de combinare a acestora a generat o valoare apropiată de cea publicată ulterior de către INS.

La nivelul economiilor avansate, metodologiile de monitorizare în timp real a activității economice care se bazează pe date publicate de către institutele de statistică au deja un istoric destul de bogat de aplicare în cadrul instituțiilor financiar-monetare. De pildă, Fed utilizează în procesul de proiecție a creșterii economice metoda propusă de Giannone *et al.*, 2008, FMI, pe cea prezentată de Matheson, 2014, în timp ce la nivel european, exemple de astfel de metodologii sunt ilustrate de Altissimo *et al.*, 2010; Bańbura *et al.*, 2010; Marcellino și Schumacher, 2010; Bańbura *et al.*, 2013. În schimb, lucrările care includ astfel de metodologii sunt mai rare la nivelul țărilor central și est-europene (CEE). În cazul acestora, procesul de tranziție către stadiul de economie de piață funcțională a fost marcat de multiple și ample modificări structurale, care au afectat inclusiv calitatea datelor statistice. Astfel, setul de indicatori disponibili care pot fi folosiți într-un proces de prognoză în timp real prezintă anumite carențe, de ordin calitativ, dar și cantitativ, aceste impedimente conducând la concluzia că modelele simple autoregresive de ordinul I au o performanță superioară față de unele mai complexe care includ un set mai larg de informații (Rünstler *et al.*,

2009). Odată cu intensificarea procesului de convergență economică și majorarea disponibilității cronologice a seriilor de timp, lucrări mai recente au dovedit fiabilitatea metodelor de prognoză în timp real și pentru țările CEE, mai multe metode teoretice cu grade diferite de complexitate fiind aplicate cu relativ succes economiilor din regiune (Tănase, 2014; Feldkircher *et al.*, 2015; Havrlant *et al.*, 2016; Rusnák, 2016; Armeanu *et al.*, 2017; Tóth, 2017; Riedl și Wörz, 2018; Gălățescu și Labhard, 2019).

Lucrarea este structurată astfel: în Secțiunea 2 fiecare dintre metodele de prognoză aminte anterior este prezentată în detaliu; Secțiunea 3 trece în revistă baza de date utilizată, precum și maniera în care a fost construită procedura de evaluare *out-of-sample*, în timp ce în Secțiunea 4 este prezentată performanța de prognoză individuală a modelelor. Secțiunea 5 prezintă modalitatea de combinare a densităților de repartiție alături de evaluarea acurateței prognozei și a unor analize de robustețe. Ultima secțiune a lucrării expune concluziile și direcții viitoare de cercetare.

2. Aspecte teoretice

Disponibilitatea tot mai ridicată a datelor cu frecvențe mixte a prezentat o motivație importantă pentru dezvoltarea metodologiilor econometrice care tratează, într-o manieră adecvată, această proprietate, seriile de timp publicate mai frecvent conținând informații importante, care pot îmbunătăți proiecțiile variabilelor care prezintă un interval trimestrial de publicare; în cazul de față, primele date oficiale cu privire la creșterea economică sunt furnizate după aproximativ 45 de zile de la finalizarea trimestrului, interval de timp în care multipli indicatori cu frecvență lunară, care sintetizează evoluția principalelor sectoare economice, devin disponibili. În această lucrare sunt considerate patru categorii de modele, care prezintă strategii diferite de a include cei mai recenți indicatori lunari: (i) modele punte (engl. *bridge equations*), (ii) modele MIDAS (engl. *MIxed DAta Sample*), (iii) modele cu factori dinamici cu frecvențe mixte (MF-DFM), precum și (iv) modele VAR Bayesiene, de asemenea, cu frecvențe mixte (MF-BVAR).

2.1. MODELE PUNTE

Modelele punte reprezintă una dintre primele abordări econometrice și, de altfel, una dintre cele mai puțin complexe metode de tratare a eșantioanelor de date cu frecvență mixtă. Printre primele contribuții teoretice se numără cea al lui Klein și Sojo, 1989, metodologia fiind ulterior dezvoltată de Baffigi *et al.*, 2004; Golinelli și Parigi, 2007; Diron, 2008. Aceste tipuri de modele sunt intens utilizate în cadrul exercițiilor de prognoză a activității economice întrucât oferă un cadru statistic simplificat care permite modificarea cu ușurință a setului de variabile utilizat. Metodologia aparține clasei de modele, denumite *distributed lag,* care presupune specificarea unui număr

ridicat de *lag*-uri pentru toate variabilele considerate, abordarea reprezentând, practic, punctul de plecare a majorității metodelor care studiază efectele intertemporale dintre variabile. Un model econometric de acest fel poate fi scris:

$$y_t = \alpha + \sum_{s=1}^{Q} \beta_s \, x_{it-s} + u_t \qquad (1)$$

unde y_t este variabila de interes, cu x_{it} este notată variabila explicativă *i*, *Q*, este numărul maxim de *lag*-uri, β_s sunt coeficienții *lag*-urilor (sau *lag weights)*, iar u_t sunt erorile modelului, presupuse a fi staționare. Această ecuație poate fi estimată doar după ce indicatorii cu frecvență înaltă sunt transformați din punctul de vedere al dimensiunii temporale, pentru a se potrivi cu cei disponibili la frecvență joasă. În ceea ce privește specificarea regresiei, de cele mai multe ori variabilele sunt incluse (alături de structura autoregresivă) pe baza criteriilor informaționale, care surprind capacitatea modelului statistic de a explica varianța variabilei de interes. O altă abordare se referă la utilizarea procedurii *Bayesian Model Averaging* care realizează o combinare a mai multor specificații (estimate pe baza unui set extins de indicatori), în funcție de performanța istorică (*in-sample*) a acestora (v. Fernández *et al.*, 2001; Bencivelli *et al.*, 2012).

În ceea ce privește procesul de uniformizare a frecvențelor variabilelor incluse în modelul statistic, o alegere facilă este dată de agregarea observațiilor prin aplicarea operatorului de medie aritmetică $X_{it} = \frac{1}{3}(X_{it}^{(m)} + X_{it}^{(m-1)} + X_{it}^{(m-2)})$, procedură care se realizează, de regulă, înaintea oricărei transformări care vizează staționarizarea datelor, cum ar fi, operatorul de primă diferență $x_{it} = \Delta \ln X_{it}$. De menționat că există posibilitatea de a lucra direct cu variații lunare, funcția de agregare având astfel o formă mai complexă (Mariano și Murasawa, 2003). Ecuația (1) poate fi rescrisă astfel (similar cu Armesto *et al.*, 2010):

$$y_t = \alpha + \sum_{s=1}^{Q} \beta_s \, (L^s)\overline{x}_{t-s} + u_t \qquad (2)$$

unde L^s este operatorul de *lag* trimestrial (de exemplu, *s* = 2 prin urmare), $(L^2)\overline{x}_t = \overline{x}_{t-2}$, $\overline{x}_t = \Delta \ln \overline{X}_t = \Delta \ln \left(\frac{1}{m} \sum_{h=1}^{m} (L_h^s) X_{it}^{(m)} \right)$ unde $X_{it}^{(m)}$ este indicatorul i la frecvență lunară, iar L_h^s operatorul de *lag* lunar. În cazul de față, unde sunt considerate observații trimestriale și lunare, *m* = 3.

În ceea ce privește modalitatea de generare a prognozelor, se poate folosi metoda directă, situație în care nu sunt necesare proiecții suplimentare pentru variabilele lunare, *lag*-uri aferente acestora fiind specificate începând cu orizontul *t* – *h*:

$$y_t = \alpha + \sum_{s=1}^{Q} \beta_s \, (L^s)\overline{x}_{t-h-s} + u_t \qquad (3)$$

Totuși, în exercițiile de prognoză în timp real (*nowcasting*), este preferabilă utilizarea celor mai recente informații, multe dintre acestea fiind disponibile după orizontul *t*. În această situație, se poate folosi metoda indirectă: variabilele lunare nu mai sunt

specificate cu *lag*-uri începând cu *t* – *h*, ci se păstrează relația de contemporaneitate, valorile lipsă pentru a completa informațiile din trimestru, necesare pentru aplicarea diverselor funcții pentru transformarea frecvențelor, obținându-se cu ajutorul unor modelele econometrice auxiliare – Foroni și Marcellino, 2013 discută pe larg avantajele și dezavantajele fiecărei abordări.

În situația în care se dorește utilizarea în continuare a metodei directe de prognoză, alături de folosirea celor mai recente informații disponibile, este necesară renunțarea la funcția auxiliară de transformare a variabilelor. Așadar, ecuația (2) presupune că fiecare coeficient estimat, care poate fi asociat observațiilor lunare $x_{it}^{(m)}$, este același, respectiv $\beta_s \frac{1}{m}$. O relaxare a acestei ipoteze presupune că fiecare parametru $\gamma_s = \beta_s \frac{1}{m_s}$, care poate fi „urmărit" până la nivelul observațiilor lunare, este unic. În aceste condiții, ecuația de mai sus devine:

$$y_t = \alpha + \sum_{s=1}^{Qm} \gamma_s (L_h^s) x_{it-s}^{(m)} + u_t \qquad (4)$$

$x_{it}^{(m)}$ este observația lunară *m* dintr-un trimestru – fiecărei variabile trimestriale îi corespund trei variabile lunare. În această formulare, numărul coeficienților care trebuie estimați este: $\gamma_s = \beta_s m$, unde *m* este factorul de conversie între observațiile lunare și cele trimestriale, respectiv 3 la 1. Prin urmare, una dintre problemele ridicate de această formă este dată de creșterea alertă a parametrilor care trebuie estimați atunci când sunt considerate mai multe variabile și/sau *lag*-uri. De exemplu, într-un context în care sunt incluse două variabile lunare, fiecare cu două *lag*-uri trimestriale, numărul parametrilor care trebuie estimați este: $(i \times Q) \times m = 12$. Mai multe modalități prin intermediul cărora se poate reduce acest număr sunt prezentate în secțiunea următoare, varianta utilizării variabilelor lunare netransformate fiind preferabilă, întrucât orice ajustare a frecvențelor introduce un anumit grad de eroare la nivelul specificației (Andreou *et al.*, 2012).

2.2. MODELE MIDAS

Modelele MIDAS, care aparțin, la rândul lor, clasei *distributed lag* ilustrate de ecuația (4) presupun introducerea unor restricții la nivelul coeficienților, astfel încât numărul de parametri care trebuie estimați să se reducă, fără a ignora însă unele legături relevante care apar la nivelul variabilelor considerate. Dezvoltată inițial pentru a fi folosită în domeniul financiar (Ghysels *et al.*, 2004; León *et al.*, 2007), iar apoi aplicată și în cazul analizei macroeconomice (Clements și Galvão, 2009; Armesto *et al.*, 2010; Kuzin *et al.*, 2011; Bai *et al.*, 2013), metodologia MIDAS permite introducerea în aceeași regresie a unor serii de timp care pot prezenta frecvențe multiple (zilnice, săptămânale, lunare etc.), fără a fi necesare transformări prealabile ale datelor. Principalul avantaj al acestei metodologii îl reprezintă posibilitatea de a folosi în procesul de prognoză abordarea directă, fără a renunța la unele informații mai recente. În schimb, metodologia MIDAS reclamă utilizarea unor tehnici neliniare de estimare, în contextul în care reducerea numărului de parametri se realizează prin impunerea unor funcții polinomiale, fapt care

complică semnificativ procesul de evaluare a acestor modele. Discuții mai aprofundate cu privire la metodologia MIDAS, precum și unele comparații cu modelele punte sunt disponibile în Foroni și Marcellino, 2013, respectiv în Schumacher, 2016.

Forma generică a regresiei MIDAS este următoare:

$$y_t = \alpha + \beta B(L_m, \theta) x_t^{(m)} + u_t \quad (5)$$

unde y_t este variabila cu frecvență trimestrială, $x_t^{(m)}$ este variabila lunară, iar $B(L_m, \theta)$ este un polinom de *lag* care dictează importanța fiecărui *lag* lunar L_m asupra variabilei de interes: $B(L_m, \theta) = \sum_{s=0}^{Q} c(s, \theta) L_m^s$, unde $L_m^s x_t^{(m)} = x_{t-s}^{(m)}$ reprezintă operatorul de *lag* de frecvență înaltă (lunar), în timp ce $c(s, \theta)$ este funcția care denotă coeficienții de ponderare, Q reprezentând numărul maxim de *lag*-uri. Cadrul teoretic al utilizării polinoamelor de *lag* se bazează pe teorema *Weierstrass,* potrivit căreia orice funcție într-un interval continuu finit poate fi aproximată printr-un polinom de gradul i (Lutkepohl, 1980).

În literatură sunt prezente multiple specificații pentru forma funcțională a funcției de ponderare $c(s, \theta)$, dezideratul fiind acela de a formula o legătură între parametrii θ și importanța asociată fiecărui *lag* lunar, în contextul unui număr potențial ridicat de *lag*-uri ale indicatorilor lunari. Totodată, precizia cu care parametrii θ sunt estimați este direct corelată cu numărul de *lag*-uri specificat (Kuzin *et al.*, 2011). Una dintre cele mai frecvent utilizate forme funcționale pentru funcția de ponderare este dată de polinomul *Exponential Almon*:

$$c(s, \theta) = \frac{\exp\left(\theta_1 s + \cdots + \theta_Q s^Q\right)}{\sum_{s=1}^{Q} \exp\left(\theta_1 s + \cdots + \theta_Q s^Q\right)} \quad (6)$$

Ghysels *et al.*, 2005 și Andreou *et al.*, 2012, de pildă, folosesc o specificație cu doi parametri θ; pentru $\theta_1 = \theta_2 = 0$, polinomul distribuie ponderi egale *lag*-urilor. În funcție de valorile parametrilor θ, funcția poate prezenta multiple forme; de exemplu, parametrii $\theta_1 = 0{,}01$ și $\theta_2 = -0{,}03$ determină o pantă descrescătoare abruptă (fiind astfel favorizate cele mai recente *lag*-uri), în timp ce o formă de cupolă este obținută dacă $\theta_1 = 0{,}5$, iar $\theta_2 = -0{,}05$ (Grafic 1).

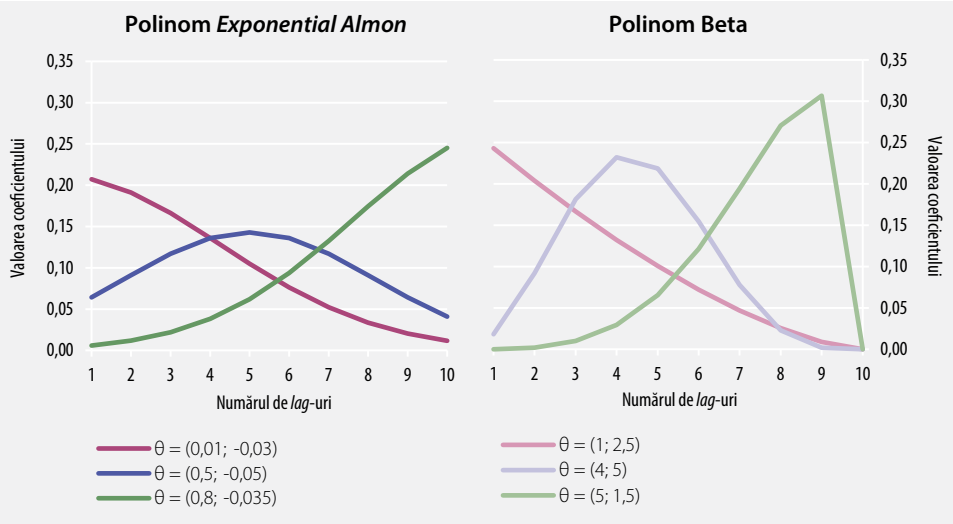
O altă abordare cu privire la specificația polinomului este denumită *Beta lag*, care presupune estimarea a doi parametri θ_1, θ_2 care imprimă forma funcției de ponderare:

$$c(k, \theta_1, \theta_2) = \frac{f(k/Q, \theta_1, \theta_2)}{\sum_{k=1}^{Q} f(k/Q, \theta_1, \theta_2)} \quad (7)$$

unde $f(x, \theta_1, \theta_2) = \frac{x^{\theta_1 - 1}(1-x)^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}$, iar $\Gamma(\theta) = \int_0^\infty e^{-x} x^{\theta - 1}\, \mathrm{d}x$.
Similar, în funcție de valoarea parametrilor θ, polinomul de ponderare poate avea o formă descrescătoare, de cupolă, etc., o pondere egală pentru toate *lag*-urile fiind obținută când $\theta_1 = \theta_2 = 1$. Întrucât influența *lag*-urilor lunare asupra variabilei de

interes, indiferent de numărul acestora, este evaluată cu ajutorul unui număr restrâns de parametri, în cadrul regresiei MIDAS pot fi incluse mai multe variabile lunare, fără ca numărul de grade de libertate să se reducă masiv; alte tipuri de funcții polinomiale sunt prezentate în Ghysels *et al.*, 2009; Andreou *et al.*, 2012 sau Schumacher, 2016.

Grafic 1. Parametrizări ale polinoamelor de *lag*



Metodologia MIDAS permite, de asemenea, includerea unui termen autoregresiv la nivelul regresiei (γy_{t-1}), după cum se obișnuiește în aplicații utilizând variabile macroeconomice. În acest caz, procedura de estimare devine mai anevoioasă, întrucât introducerea unei componente autoregresive în ecuația (5) aduce printre variabilele explicative una specificată la o frecvență inferioară celorlalte deja prezente în model. În aceste condiții, rescrierea modelului în formă redusă generează o componentă a polinomului de *lag* care ia valori doar odată la 3 observații lunare, ceea ce creează o formă de sezonalitate la nivelul modelului (Ghysels *et al.*, 2007; Foroni și Marcellino, 2013). Pentru a elimina această componentă (care afectează inclusiv acuratețea prognozei), se utilizează o procedură de estimare în mai multe etape, unde mai întâi este evaluat modelul fără componentă autoregresivă, urmând ca pentru aceasta să se obțină o valoarea inițială pe baza reziduurilor ecuației. Mai departe, variabilele din model sunt ajustate ținând cont de estimarea inițială a parametrului autoregresiv, după care procesul de estimare este reluat – Clements și Galvão, 2009 redau în detaliu această procedură.

2.3. MODELE CU FACTORI DINAMICI CU FRECVENȚE MIXTE

Modelele cu factori dinamici (DFM) presupun că informațiile prezente la nivelul unui set de indicatori pot fi sintetizate în două componente ortogonale neobservate: prima include traiectoriile comune tuturor indicatorilor, iar cea de-a doua (numită componentă idiosincratică) grupează evoluțiile specifice ale acestora, precum și unele erori de măsurare (Stock și Watson, 2002). Evoluțiile comune sunt grupate apoi într-o serie de vectori, numiți factori comuni, aceștia încorporând o parte din covariațele

dintre indicatorii incluși în baza de date utilizată de model. Trebuie menționat că datorită modului în care factorii comuni au fost construiți evoluția acestora nu are o interpretare economică clară (în sensul că nu poate fi asociată unui anumit sector de activitate), în cazul de față plecându-se de la ipoteza că aceștia reflectă forțele motrice ale economiei.

Descompunerea unei serii lunare x_{it}staționare este următoarea:

$$x_{it} = \mu_i + \lambda_{i1} f_{1t} + \cdots + \lambda_{iR} f_{Rt} + \xi_{it} \quad (8)$$

unde $i = 1 \ldots n$, *n* fiind numărul total de variabile lunare. f_{rt} reprezintă factorul comun r , iar ξ_{it} este componenta idiosincratică. În formă matriceală, ecuația de mai sus poate fi scrisă astfel:

$$X_t = \Lambda F_t + \varepsilon_t \quad (9)$$

unde $X_t = \left(x_{1t} \ldots x_{nt}\right)'$, $\varepsilon_t = \left(\xi_{1t} \ldots \xi_{nt}\right)'$, $F_t = (f_{1t} \ldots f_{Rt})'$, unde *R* reprezintă numărul total de factori cu $R \leq n$. Λ este matricea coeficienților λ_{ir} (engl. *factor loadings),* de dimensiune $n \times R$ care reprezintă cantitatea de informații asociată variabilei x_{it} inclusă în factorul comun f_{rt} . De regulă, datele utilizate pentru estimarea factorilor sunt standardizate, astfel încât nu mai este necesară specificarea unui termen constant μ_i.

Factorii comuni neobservabili F_t pot fi estimați într-o manieră robustă utilizând metoda componentelor principale (prezentată de Stock și Watson, 2002). În cazul de față, având în vedere că metodologia este aplicată cu scopul de a realiza proiecții pentru orizontul $T + h$, este preferabilă construirea unei structuri autoregresive pentru factorii comuni. În caz contrar, ar fi necesară realizarea unor proiecții auxiliare pentru setul de indicatori X_t, pentru întreg orizontul de proiecție, factorii comuni putând fi astfel extrași pe baza setului extins de indicatori. Această abordare comportă unele discuții cu privire la calitatea prognozelor auxiliare, întrucât eventuale erori la acest nivel pot fi transmise în proiecțiile finale. De menționat că această procedură este aplicată în cazul modelelor punte (prezentate în subsecțiunea 2.1) care utilizează factori comuni.

Prin urmare, în modele DFM componenta dinamică este dată de o structură autoregresivă de ordin p pentru factorii comuni. O reprezentare de ordinul I, în formă matriceală, este următoarea:

$$F_t = AF_{t-1} + Bu_t \quad (10)$$

unde termenul de eroare este distribuit $u_t \sim \mathcal{N}(0, I_Q)$, *A* este o matrice de ordin $(R \times R)$, iar *B* o matrice de ordin $(R \times Q)$, unde Q este numărul șocurilor comune incluse pentru a surprinde anumite evoluții de-a lungul ciclurilor economice (mai multe detalii se găsesc în Forni *et al.*, 2009).

Modelul cu factori comuni dinamici, ilustrat de ecuația (8) și ecuația (9), reclamă ca toate variabilele lunare x_{it} să aibă același număr de observații. În acest context, eventualele serii de date care devin disponibile pot fi folosite doar atunci când

toți indicatorii din baza de date sunt publicați. Totuși, pentru a putea include noile observații apărute, Doz *et al.*, 2011 propun utilizarea în conjuncție a filtrului Kalman și a metodologiei componentelor principale pentru a balansa setul lunar de observații și a estima factorii comuni. În prima etapă a acestei proceduri, matricea Λ și factorii comuni F_t sunt evaluați utilizând metoda componentelor principale aplicată setului de informații $\bar{X}_t$, care anterior a fost echilibrat prin renunțarea la observațiile incomplete, standardizat și prelucrat pentru eliminarea valorilor extreme (mai multe detalii cu privire la metoda componentelor principale se regăsesc în Giannone *et al.*, 2008). În cea de-a doua etapă, se utilizează filtrul Kalman, mai precis partea de netezire (engl. *smoothing*; mai multe detalii se regăsesc în Durbin și Koopman, 2012), care se bazează pe structura autoregresivă a modelului DFM pentru a estima observațiile lipsă și pentru a evalua factorii comuni, avându-i ca punct de plecare pe cei estimați în prima etapă. O altă strategie de tratare a observațiilor lipsă este reprezentată de algoritmul *Expectation Maximization* (propus în a fi utilizat în astfel de aplicații de către Stock și Watson, 2002; alte exemple sunt Bańbura și Rünstler, 2011 și Angelini *et al.*, 2006), care presupune un proces de maximizare a funcției de *likelihood* a setului de indicatori utilizând pentru observațiile lipsă valori care sunt reactualizate având ca reper nivelul așteptat al acestora din iterația anterioară.

Până în acest moment, discuția s-a concentrat asupra procedurii de modelare a factorilor comuni lunari, în literatură conturându-se două abordări cu privire la legarea acestora de o variabilă trimestrială. Prima dintre acestea presupune că o observație trimestrială este compusă din variabile lunare neobservabile, estimarea acestora realizându-se odată cu cea a factorilor comuni prin intermediul filtrului Kalman (Bańbura și Rünstler, 2011). De menționat că această strategie este utilizată și în cadrul modelelor VAR Bayesiene cu frecvență mixtă care vor fi prezentate în secțiunea următoare. A doua metodă (propusă de Giannone *et al.*, 2008), care este, de altfel, cea preferată în această lucrare, presupune transformarea frecvenței factorilor într-o manieră similară modelelor punte:

$$y_{qt} = \alpha + \beta' \hat{F}_{qt} \tag{11}$$

unde y_{qt} este variabila trimestrială, α și β' sunt coeficienții care trebuie estimați, iar $\hat{F}_{qt}$ sunt factorii transformați la frecvență trimestrială, ținând cont că aceștia au fost estimați pe baza dinamicilor lunare (procedeul de transformare este similar celui propus de Mariano și Murasawa, 2003):

$$\hat{f}_{qrt} = \frac{1}{3}[f_{rt} + 2f_{rt-1} + 3f_{rt-2} + 2f_{rt-3} + f_{rt-4}] \tag{12}$$

În formă matriceală, ecuația de mai sus poate fi scrisă astfel:

$$\hat{F}_{qt} = \Lambda_Q F_t{}' + \phi_Q \xi_t \tag{13}$$

unde $\Lambda_Q = \left[\Lambda_Q \; 2\Lambda_Q \; 3\Lambda_Q \; 2\Lambda_Q \; \Lambda_Q\right]$, $F_t = [F_t \; \ldots \; F_{t-4}]'$, $\phi_Q = [1\; 2\; 3\; 2\; 1]$, iar ξ_t este termenul de eroare.

2.4. MODELE VAR BAYESIENE LA FRECVENȚE DIFERITE

Modelele VAR (Vectori Auto Regresivi) au devenit în ultimii ani o metodologie des folosită, atât pentru analiza diverselor fenomene economice, cât și în cadrul exercițiilor de prognoză. Un obstacol major în utilizarea în practică a acestei metode este dat de creșterea alertă a numărului de parametri odată cu extinderea variabilelor și/sau a *lag*-urilor, fapt care erodează nivelul de precizie a estimărilor. O soluție de limitare a efectelor negative amintite anterior (propusă inițial de Litterman, 1986) constă în utilizarea tehnicilor Bayesiene de estimare, care presupun, în esență, combinarea informațiilor cuprinse în setul de date cu unele ad-hoc.

De pildă, o ipoteză des folosită presupune că cei mai importanți parametri sunt cei aferenți primului *lag*, informația fiind inclusă în procesul de estimare prin specificarea unei distribuții *a priori* pentru matricea parametrilor, centrate la valoarea unitate pentru primul *lag*, parametrii *lag*-urilor superioare având o importanță mai redusă, sau, în unele cazuri, putând fi complet ignorați (similar covarianțelor dintre variabile). Intensitatea cu care această informație este preluată la nivelul rezultatelor finale depinde de varianța asociată mediei *a priori*, care poate fi reglată cu ajutorul unui parametru denumit *tightness*. Această abordare poartă denumirea de *Minnesota prior* – mai multe detalii cu privire la tehnicile utilizate în econometria Bayesiană oferă Koop, 2003; Geweke, 2005.

Un model VAR de ordin p are următoarea formă:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t \qquad (14)$$

cu $x_t = \left[x_{1,t} \ldots x_{n,t}\right]'$ sunt notate variabilele incluse în model, ϕ_1 sunt coeficienții asociați primului *lag*, iar $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ este termenul de eroare.

Până în acest moment nu s-a făcut nicio presupunere cu privire la frecvența variabilelor incluse în modelul VAR, singura constrângere fiind aceea ca numărul de observații al fiecărei variabile să fie același. În continuare, se presupune că vectorul x_t conține variabile observabile la frecvență lunară $x_{m,t}$, precum și unele care sunt disponibile doar trimestrial, dar care prezintă totuși un proces lunar de generare a datelor, ale cărui realizări sunt neobservabile la această periodicitate – $x_{q,t}$ (similar cu Schorfheide și Song, 2015). Prin urmare $x_t = \left[x_{m,t}{}', x_{q,t}{}'\right]'$.

Întrucât vectorul x_t conține variabile neobservabile (latente), respectiv $x_{q,t}$, evaluarea acestuia se realizează, de regulă, utilizând filtrul Kalman. Pentru aplicarea acestuia este necesar ca modelul VAR, ecuația (14), să fie scris sub forma unui model *state-space* unde se presupune un sistem cu două ecuații: o ecuație a stărilor care modelează evoluția în timp a variabilelor neobservabile și o ecuație de măsurare, care include relațiile dintre indicatorii observați și cei latenți – mai multe detalii în această privință sunt disponibile în Hamilton, 1994; Durbin și Koopman, 2012.

Prin urmare, rescriind în formă matriceală ecuația (14) se obține o formulă care poate fi interpretată ca o ecuație de dinamică pentru componenta neobservată a modelului:

$$Z_t = \pi + \Pi Z_{t-1} + \epsilon_t \tag{15}$$

unde $\varepsilon_t \sim \mathcal{N}(0, \Omega)$, iar $Z_t = \left(x_t', x_{t-1}', \dots, x_{t-p+1}'\right)'$, de dimensiune $np \times 1$ primele n rânduri reproducând ecuația (14), iar celelalte fiind identități $x_{t-1}' = x_{t-1}'$. Ecuația (15) conține și variabile lunare, care în situația de față sunt observate. De menționat că această formă a modelului majorează semnificativ timpul asociat procesului de estimare, matricea Π având o dimensiune nejustificat de mare comparativ cu numărul variabilelor latente (Schorfheide și Song, 2015 propun o rescriere alternativă a modelului care reduce considerabil dimensiunea matricei care vizează evoluția stărilor).

În continuare, vectorul care conține variabilele observate este notat $y_t = \left(y_{m,t}', y_{q,t}'\right)'$, unde observațiile lunare sunt $y_{m,t} = x_{m,t}$. Cu $y_{q,t}$ sunt notate observațiile trimestriale. În cazul de față, indicatorii considerați sunt introduși în model în nivel (i.e. variabilele sunt transformate prin aplicarea funcției de logaritm natural, mai puțin cele care au unitatea de măsură specificată în procente), iar relația dintre variabilele observate și cele latente este următoarea:

$$y_{iq,t} = \frac{1}{3}\left(x_{iq,t} + x_{iq,t-1} + x_{iq,t-2}\right) \tag{16}$$

unde $y_{iq,t}$ este o valoare trimestrială observată pentru indicatorul i, iar $x_{iq,t}$ este nivelul estimat de model pentru ultima lună din trimestru. Pentru a uniformiza numărul de observații între variabilele lunare $y_{m,t}$ și cele trimestriale $y_{q,t}$, fără a afecta relațiile temporale, vectorul $y_{q,t}$ este completat astfel:

$$y_{q,t} = \begin{cases} \frac{1}{3}\left(x_{q,t} + x_{q,t-1} + x_{q,t-2}\right), & t \in \{\text{mar, iun, sep, dec}\} \\ \emptyset, & \text{altfel} \end{cases} \tag{17}$$

Ecuația de măsurare poate fi scrisă după cum urmează:

$$Y_t = M_t \Lambda Z_{\mathrm{t}} \tag{18}$$

unde M_t este o matrice de selecție, care în cazul variabilelor trimestriale $M_{q,t}$ are valoarea unitate dacă t corespunde ultimei observații din trimestru. Componenta asociată variabilei lunare $M_{m,t}$ este completată cu zero numai pentru observațiile lipsă aferente variabilei lunare i – de exemplu, dacă seria de date lunare τ prezintă observații până în luna martie, iar indicatorul i doar până în ianuarie, cele două poziții din $M_{m,t}$ care sunt lipsă pentru i (față de seria τ) sunt completate cu 0. În acest mod, în ecuația (18) se păstrează relațiile de identitate între seriile de date, fiind în același timp identificate observațiile lipsă la nivelul indicatorilor lunari. Matricea Λ este completată cu 1/3 în cazul observațiilor trimestriale și cu valoarea unitate în cazul celor lunare. Prin urmare, modelul format din ecuația (18) și din ecuația (15) reprezintă forma *state-space* a VAR care poate fi folosită pentru aplicarea filtrului Kalman pentru a obține concomitent o estimare pentru variabilele latente $x_{q,t}$ și pentru parametrii modelului incluși în matricea Π și pentru vectorul π, în care sunt prezenți termenii constanți (lucrări mai recente care utilizează astfel de modele în exerciții de prognoză sunt Kuzin *et al.*, 2011; Brave *et al.*, 2019; Ankargren și Jonéus, 2020).

Trecând la procesul de estimare, acesta se bazează pe tehnicile Bayesiene, care presupun combinarea funcției de *likelihood* a modelului condiționată de setul de indicatori cu distribuția parametrilor presupusă *a priori*, pentru a obține astfel o estimare a coeficienților; aceștia sunt, de fapt, o medie ponderată a celor două seturi de informații. În cazul de față, unde modelul considerat este de tipul *state-space*, rezultatul final are în vedere și distribuția factorilor latenți (a se vedea Schorfheide și Song, 2015). Prin urmare, procedura de estimare este mai complexă, pentru fiecare extragere din distribuția parametrilor fiind necesară aplicarea filtrului Kalman pentru a evalua variabilele neobservate – de menționat că etapa de *updating* a filtrului se realizează odată la trei perioade, după ce noile informații trimestriale sunt disponibile; Durbin și Koopman, 2012 tratează pe larg acest subiect.

În modelul VAR utilizat în această lucrare, specificația informațiilor *a priori* despre parametri este în linie cu abordarea *Minnesota prior* (Litterman, 1986). Prin urmare, se presupune că parametrii sunt distribuiți normal cu media $\widetilde{\Pi}_0$ și varianță $\widetilde{\Omega}_0$:

$$p(\widetilde{\Pi}) \sim \mathcal{N}(\widetilde{\Pi}_0, \widetilde{\Omega}_0) \tag{19}$$

unde $\widetilde{\Pi} = [\pi' \ \Pi']$, iar $\widetilde{\Pi}_0$ conține ipoteza cu privire la valoarea coeficientului autoregresiv de ordinul l:

$$\widetilde{\Omega}_0 = \begin{pmatrix} \gamma \ \mathrm{diag}(n) & \cdots & 0_{n\times[(p-1)+1]} \\ \vdots & \ddots & \vdots \\ 0_{n\times[(p-1)+1]} & \cdots & 0_{n\times[(p-1)+1]} \end{pmatrix} \tag{20}$$

În ceea ce privește matricea de varianță covarianță, $\widetilde{\Omega}_0$, informațiile *a priori* sunt completate astfel:

$$\widetilde{\Omega}_0 = \begin{pmatrix} \tilde{\Sigma}_0 & \cdots & 0_{n\times[(p-1)+1]} \\ \vdots & \ddots & \vdots \\ 0_{n\times[(p-1)+1]} & \cdots & 0_{n\times[(p-1)+1]} \end{pmatrix} \tag{21}$$

$$\tilde{\Sigma}_{0_l}^{(i,j)} = \begin{cases} \dfrac{\lambda^2}{(l^{\lambda_3})^2}, & i = j \\ \dfrac{\lambda^2\lambda_2^2}{(l^{\lambda_3})^2}\dfrac{s_i^2}{s_j^2}, & \text{în caz contrar} \end{cases} \tag{22}$$

$\tilde{\Sigma}_0^{(i,j)}$ semnalează elementul de pe linia i coloana j a matricei de varianță-covarianță $\tilde{\Sigma}_0$ asociată *lag*-ului *l*. λ este coeficientul de *tightness* care determină gradul de încredere asociat informațiilor *a priori*, o valoare mai ridicată a acestuia (care echivalează cu varianță mai înaltă pentru coeficientul autoregresiv γ) semnalând un nivel mai redus. λ_3 este parametrul care determină viteza de apropiere către zero a *lag*-urilor superioare (în cazul de față este fixat la 1), iar λ_2 este parametrul care dictează importanța covarianțelor între variabilele modelului, presupus a fi egal cu 0,5. s_i^2/s_j^2 este un termen de scalare calculat utilizând deviația standard a fiecărei ecuații.

În ceea ce privește matricea de varianță-covarianță a erorilor, o primă abordare utilizată în această lucrare este aceea de a o considera constantă în timp, cu o

distribuție *Inverse Wishart* (aceasta este, de altfel, abordarea standard în literatură; a se vedea *inter alia* Bańbura *et al.*, 2010):

$$p(\Omega_0) \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \tag{23}$$

unde $\underline{S} = (\underline{\nu} - n - 1)\text{diag}(s_1^2, \dots, s_n^2)$, s_i^2 fiind varianțele reziduurilor asociate variabilei (ecuației) x_i, iar $\underline{\nu}$ sunt gradele de libertate $\underline{\nu} = n + 2$.

Pe lângă această formă generală pentru evoluția erorilor, a doua abordare este aceea de a introduce un factor comun de scalare la nivelul matricei de varianță-covarianță a modelului, care prezintă variabilitate în timp. Astfel:

$$\epsilon_t \epsilon_t' = f_t \Omega \tag{24}$$

unde $f_t = f_{t-1} + v_t$ are o reprezentare *Random Walk*, v_t fiind termenul de eroare. Întrucât factorul comun de scalare este neobservabil, ecuația care descrie evoluția lui în timp suplimentează modelul *state-space* dat de ecuația (18), respectiv ecuația (15), traiectoria fiind evaluată odată cu parametrii modelului și cu celelalte componente latente; mai multe detalii se găsesc în Carriero *et al.*, 2016.

3. Baza de date și designul exercițiilor de prognoză

Baza de date utilizată pentru evaluarea capacității de prognoză a modelelor prezentate în Secțiunea 2 este compusă dintr-un set de 95 de indicatori cu frecvență lunară, pentru o perioadă cuprinsă între ianuarie 2000 și septembrie 2020 – datele sunt furnizate, în principal, de către INS, Eurostat și BNR, informații suplimentare fiind prezentate în Anexă. Strategia de construire a bazei de date a avut în vedere, printre altele, asigurarea unei reprezentativități adecvate a sectoarelor economice, în funcție de importanța acestora la formarea PIB (Tabel 1).

Tabel 1. Reprezentativitatea indicatorilor din baza de date

Sector economic	Secțiune CAEN	% în total VAB, 2019	% din numărul de indicatori din baza de date
Agricultură	A	4,5	2,5
Industrie	B-E	24,1	24,1
Construcții	F	7,1	7,6
Comerț și servicii recreative	G, I, R	15,3	16,5
Transport și depozitare	H	6,8	7,6
Servicii financiare	K	2,7	6,3
Servicii IT&C	J	6,1	7,6
Activități profesionale și de suport	L-N	16,7	17,7
Educație, administrație publică și alte servicii	O-T	16,7	10,1

Evoluția sectorului industrial este surprinsă prin indicele producției industriale și subcomponente ale acestuia din perspectiva marilor grupe industriale (MGI), volumul cifrei de afaceri, precum și indicatori care conțin anumite informații predictive, cum ar comenzile noi adresate sectorului manufacturier sau indicatori furnizați de S*ondajul de conjunctură DG ECFIN*. În general, aceste variabile (altele decât cele din sondaj) sunt disponibile după aproximativ 45 de zile de la încheierea lunii de referință – de pildă, pentru luna decembrie, indicatorii se publică în a doua săptămână din luna februarie a anului următor.

Trecând la sectorul comercial și cel al serviciilor recreative, setul de indicatori include indici de volum ai cifrei de afaceri pentru comerțul cu amănuntul (bunuri și carburanți), vânzările de autovehicule (inclusiv piese, motociclete și alte vehicule ușoare), precum și volumul activității pe segmentul hotelurilor și al restaurantelor, al serviciilor recreative și al agențiilor de turism. De asemenea, sunt incluși și indici de volum ai comerțului cu ridicata, în acest caz fiind necesară deflatarea seriei utilizând IPC exclusiv impactul modificării cotei TVA, indicatorii aferenți acestui sector fiind disponibili pentru întreaga perioadă considerată, doar în termeni nominali. Aceste informații sunt publicate de către INS după aproximativ 45-50 de zile de la finalizarea lunii de referință.

Una dintre cele mai importante provocări în alcătuirea setului de indicatori a fost reprezentată de estimarea unor indici de preț lunari pentru sectorul serviciilor prestate întreprinderilor pentru întreg intervalul de timp; indici de volum pentru cifra de afaceri sunt disponibili doar începând cu 2010, în timp ce informații cu privire la valoarea lunară a încasărilor sunt furnizate de Eurostat începând cu anul 2000 și pentru un set mai larg de secțiuni CAEN. Totuși, INS publică serii de date cu privire la evoluția prețurilor din sectorul serviciilor prestate întreprinderilor, însă acestea au frecvență trimestrială și, de asemenea, nu sunt disponibile decât din anul 2010. Indicii de preț trimestriali reprezintă punctul de plecare pentru aproximarea celor lunari: pentru perioada ulterioară anului 2010, evoluțiile sunt interpolate utilizând metodologia MF-BVAR, iar pentru cea anterioară, sunt vizate legăturile econometrice dintre indicii de preț și alți indicatori lunari (remunerarea salariaților din fiecare sector de activitate, unele componente din IPC servicii, precum și cursul de schimb nominal, cel din urmă prezentând o relevanță sporită în cazul serviciilor telecom și al transporturilor de mărfuri). Prin urmare, baza de date include volumul cifrei de afaceri pentru sectorul serviciilor de transport, informații și comunicații (IT&C), activități profesionale și științifice, precum și componente ale acestora (la nivel de diviziune CAEN). De regulă, acești indicatori sunt disponibili cu o întârziere de aproximativ o săptămână după publicarea celor aferenți sectorului industrial.

Un alt set de informații prezent în baza de date vizează piața muncii, fiind în esență disponibile serii de date cu privire la evoluția numărului de angajați (inclusiv date privind șomajul), precum și a veniturilor salariale nominale. Toți acești indicatori sunt publicați în formă brută, utilizarea lor în exerciții econometrice necesitând aplicarea unor tehnici de ajustare sezonieră (a fost preferată metodologia *Tramo-Seats*, similar celei folosite în alte situații de către INS). În ceea ce privește câștigurile salariale, în baza de date sunt incluse doar cele nete și în termeni reali, deflatarea acestora realizându-se pe baza IPC. Totodată, pe lângă variabile care reflectă evoluția la nivelul

întregii economii, sunt introduși în setul de indicatori și cei aferenți principalelor sectoare economice, definite la nivel de diviziune CAEN. Informații cu privire la evoluția câștigurilor salariale sunt disponibile odată cu publicarea de către INS a producției industriale, în timp ce statistici privind efectivul de salariați din economie sunt furnizate, de regulă, cu o întârziere de o săptămână.

Ceilalți indicatori cuprinși în baza de date se referă la competitivitatea prin preț a economiei locale, REER, respectiv NEER, aceștia fiind preluați de la BIS și BCE, indici ai volumului exporturilor și importurilor de bunuri, conform clasificării BEC, lucrările de construcții (rezidențiale și inginerești), autorizațiile de construire (indici pe baza suprafeței utile), precum și indicatori din *Sondajul de conjunctură DG ECFIN* (încrederea în economie, industrie, comerț și construcții).

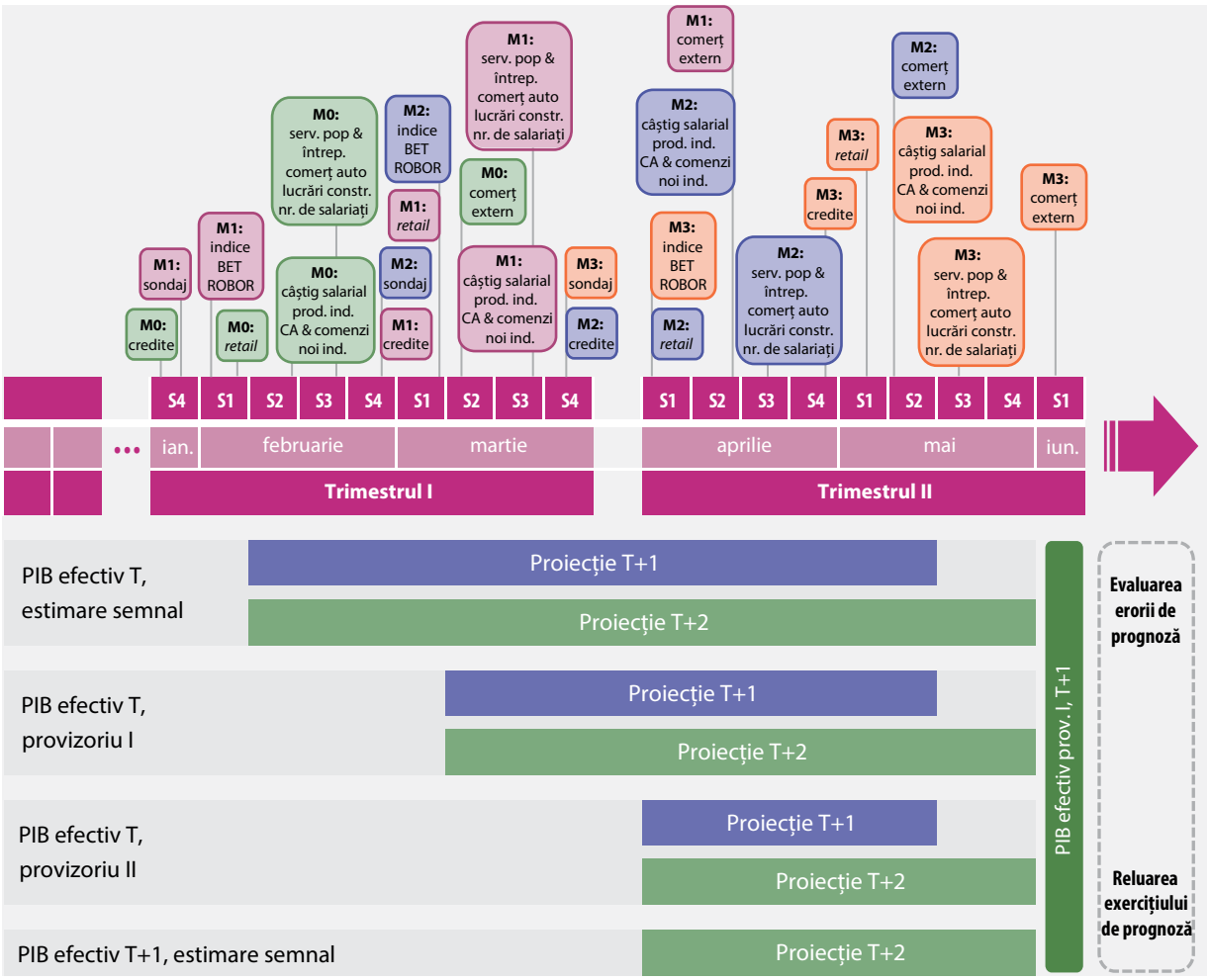
O absență notabilă în ceea ce privește acoperirea activității economice de indicatorii incluși în baza de date vizează sectorul agricol. Singurele variabile lunare disponibile se referă la câștigul salarial și la efectivul de salariați din sector, însă niciuna dintre acestea nu are capacitatea de a oferi indicii relevante cu privire la valoarea adăugată brută a sectorului, dependentă în mare măsură de condițiile meteorologice. Prin urmare, chiar dacă în ultimii 20 de ani ponderea agriculturii în PIB s-a redus semnificativ (până la circa 4 la sută în 2019), variațiile ample ale producției afectează simțitor dinamica trimestrială a PIB, sectorul reprezentând, de asemenea, o sursă importantă a revizuirilor istorice ale creșterii economice, odată ce INS dispune de informații actualizate privind producția agricolă (de regulă, abia după finalizarea anului calendaristic).

În ceea ce privește calendarul de publicare a conturilor naționale, primele informații legate de evoluția PIB sunt diseminate în aproximativ 45 de zile de la încheierea trimestrului de referință. Trebuie menționat că aceste seturi de date includ doar dinamica trimestrială a PIB real și presupun un grad ridicat de estimare, ceea ce poate conduce la ample revizuiri ulterioare. Serii de date care prezintă o imagine mai fidelă cu privire la activitatea economică sunt publicate după aproximativ 70 de zile de la finele trimestrului (provizoriu I), următoarele informații disponibile fiind după alte 30 de zile (provizoriu II); acestea încorporează mai multe variabile care conduc la o evaluare mai robustă a componentelor PIB (un exemplu relevant este reprezentat de disponibilitatea unor noi date cu privire la evoluția comerțului exterior cu servicii).

Trecând la procedura de evaluare a performanței de prognoză, pentru a replica cât mai fidel posibil setul de informații asociat calendarului de publicare a variabilelor incluse în baza de date (ilustrat în Graficul 2), exercițiile de prognoză sunt realizate succesiv, prin adăugarea de noi observații lunare. Astfel, într-o primă fază, setul de informații nu include nicio observație lunară aferentă indicatorilor de activitate sectorială pentru trimestrul *nowcasting* (T+1), situație marcată cu „M0". Apoi, iterativ, câte o nouă observație lunară este inclusă pentru fiecare indicator lunar din baza de date, fapt marcat cu „M1", „M2", respectiv „M3". Totodată, în cadrul procesului de evaluare se ține cont de revizuirile seriei PIB, fiecare exercițiu de prognoză utilizând setul de date provizoriu I disponibil la acel moment – folosirea celei mai recente ediții a datelor de Conturi naționale într-un astfel de exercițiu ar presupune că revizuirile pot fi prevăzute de către agenți (Faust *et al.*, 2005), fapt invalidat în practică, în

condițiile în care există indicii consistente că revizuirile PIB pe plan local nu sunt autocorelate (Nalban, 2014). Ideal ar fi fost să se utilizeze în cazul tuturor indicatorilor din baza de date versiunea seriei disponibilă în momentul realizării proiecției, însă aceste informații nu sunt disponibile. Totuși, trebuie menționat că indicii care se referă la evoluția producției și a cifrei de afaceri (denumiți „indicatori pe termen-scurt") prezintă revizuiri minore, în general, și doar pentru observația anterioară. Alte tipuri de modificări pot fi generate de modelul de ajustare sezonieră, însă amplitudinea acestora este suficient de restrânsă pentru a nu afecta decisiv rezultatele obținute. Perioada de evaluare *out-of-sample* începe în T1 2013, fiind adăugat, succesiv, câte un nou trimestru, până în T3 2020 (engl. *expanding window)*.

Grafic 2. Calendarul de publicare a indicatorilor din setul de date



4. Performanța de prognoză

În secțiunile anterioare au fost prezentate câteva detalii teoretice cu privire la metodologiile considerate, împreună cu designul exercițiului de prognoză. Cu toate că metodele utilizate în această lucrare sunt intens folosite în mediul academic și

la nivelul instituțiilor financiar-monetare, până în acest moment nu s-a cristalizat o abordare standard cu privire la modalitatea în care fiecare dintre ele poate fi operaționalizată, performanțele acestora fiind strâns legate de caracteristicile economiei unde sunt aplicate – un număr redus de observații, de pildă, favorizează specificații mai suple, dar care nu permit o reprezentare riguroasă a legăturilor dintre variabile. Prin urmare, în această secțiune se are în vedere explorarea diferitelor specificații ale metodologiilor de prognoză prezentate în Secțiunea 2, demers relevant pentru stabilirea celor mai potrivite abordări pentru prognoza PIB în cazul economiei locale, acestea urmând a fi incluse în cadrul final de prognoză. Metodologiile sunt evaluate pe o perioadă cuprinsă între T1 2013 (T2 în cazul orizontului de prognoză T+2) și T1 2020, interval care oferă o imagine cu privire la capacitatea modelelor de a surprinde evoluția activității economice în perioade normale (din perspectiva mediei istorice) – o discuție mai amplă cu privire la performanța modelelor de prognoză în contextul pandemiei cu virusul SARS-CoV-2 va fi prezentată în subsecțiunea 5.2. Evaluările sunt realizate prin prisma erorii medii pătratice (*Root Mean Square Forecast Error* – RMSFE), care prezintă unele avantaje computaționale într-un context caracterizat de un număr ridicat de specificații; strategia va fi generalizată în Secțiunea 5, unde sunt evaluate densitățile de repartiție ale prognozelor.

4.1. MODELE PUNTE

Prima categorie vizată în cadrul analizei performanței de prognoză este cea compusă din modele de tip punte, care presupun, în esență, o legătură liniară simplă între variabila de interes și un set de indicatori construit prin aplicarea unei funcții care permite obținerea unei frecvențe comune. În cazul de față, unde variabilele utilizate prezintă două frecvențe (trimestrială și, respectiv, lunară), compatibilizarea indicatorilor se realizează prin aplicarea unei medii aritmetice simple observațiilor lunare aferente unui trimestru (alte transformări des întâlnite sunt reprezentate de utilizarea mediei geometrice sau de considerarea exclusivă a ultimei observații lunare din trimestru).

Principalul avantaj al acestei metodologii este dat de structura statistică flexibilă, în sensul că parametrii care guvernează legătura dintre variabile sunt într-un număr redus și nu sunt supuși vreunei restricții. Această caracteristică permite metodei să surprindă rapid schimbările recente produse în traiectoria indicatorilor lunari, ceea ce poate conduce la o performanță sporită a prognozei, în special pe termen scurt și, mai ales, dacă setul de informații este suficient de extins pentru a surprinde o bună parte din factorii care determină evoluția variabilei de interes. Acest din urmă aspect reprezintă și cea mai importantă provocare asociată metodei, întrucât numărul de indicatori care pot fi utilizați în cadrul regresiei este limitat, mai ales în situația de față, unde numărul de observații trimestriale este scăzut. Pentru a depăși această problemă sunt luate în considerare două strategii. Prima dintre acestea se referă la explorarea bazei de date prin estimarea unor specificații care conțin câte un singur indicator, prognoza finală fiind dată de combinarea proiecțiilor individuale ale modelelor în funcție de performanța anterioară de prognoză.

A doua strategie presupune extragerea traiectoriilor comune prezente la nivelul întregului eşantion de date, utilizând metoda componentei principale statice (PC), procedeul permiţând reducerea semnificativă a numărului de variabile. Această abordare este potrivită în contextul de faţă întrucât traiectoria PIB poate fi rareori surprinsă de un set restrâns de variabile sectoriale, aceasta fiind mai degrabă determinată de raportul net al evoluţiilor fiecărui sector din economie. În acest context, identificarea uneia sau a mai multor traiectorii comune prezente la nivelul setului de indicatori (factori comuni), care să sintetizeze raportul net de forţe între sectoarele economice, contribuie la reducerea semnificativă a dimensiunii modelelor, îmbunătăţind astfel nivelul de precizie a estimărilor. Exerciţiul empiric presupune estimarea mai multor specificaţii pentru modelele punte, într-o primă fază fiind evaluate ecuaţii bivariate pe baza primilor cinci factori comuni, urmând ca apoi setul de modele să fie extins prin adăugarea unor specificaţii care conţin, succesiv, până la şapte factori comuni.

În ceea ce priveşte metoda de prognoză, în situaţia în care se doreşte exploatarea informaţiilor lunare apărute pe parcursul trimestrului de *nowcasting* (T+1), se impune utilizarea metodei indirecte, care presupune mai întâi realizarea unor proiecţii pentru variabilele lunare, astfel încât observaţiile lipsă din cadrul trimestrului să fie completate. De regulă, proiecţiile variabilelor lunare se realizează cu modele autoregresive simple.

Tabel 2. Performanţa de prognoză (RMSFE) a modelelor punte

Orizont de prognoză		T+1				T+2
		M0	M1	M2	M3	
Modele punte bivariate		0,746	0,715	0,709	0,708	0,749
Modele punte bivariate cu termen AR		0,798	0,827	0,784	0,778	0,878
Modele bivariate cu factori comuni		0,747	0,709	0,706	0,697	0,751
Modele punte cu factori comuni	F=2	0,857	0,789	0,734	0,700	0,861
	F=3	0,861	0,803	0,743	0,709	0,864
	F=5	0,853	0,800	0,735	0,717	0,854
	F=7	0,871	0,844	0,771	0,748	0,873

Notă: În tabel este prezentată performanţa medie de prognoză (RMSFE) calculată începând cu T1 2013 (T2 pentru orizontul T+2) până în T1 2020. Pentru referinţă, un model *Random Walk* generează un RMSFE de aproximativ 1 pentru T+1, respectiv 1,1 pentru T+2. „F" se referă la numărul de factori comuni incluşi într-un model. Coloana notată cu „M0" indică performanţa de prognoză în situaţia în care nicio observaţie lunară aferentă trimestrului de *nowcasting* nu este disponibilă, iar în coloanele notate cu „M1", „M2", respectiv „M3" setul de informaţii este extins succesiv cu câte o observaţie lunară. Modelele bivariate presupun ecuaţii care includ un singur indicator lunar prezent în baza de date, iar cele cu factori bivariate sunt evaluate pe baza primilor cinci factori extraşi utilizând metoda componentei principale statice. Proiecţiile sunt realizate indirect, indicatorii lunari fiind astfel prelungiţi utilizând un model ARMA ales pe baza criteriilor informaţionale. Prognoza pentru T+2 este condiţionată de setul de informaţii disponibil în M3, T+1.

În cadrul exerciţiului de evaluare a capacităţii de prognoză *out-of-sample* (Tabel 2), modelele punte oferă o performanţă de prognoză satisfăcătoare, eroarea medie fiind în jur de 0,7 – de pildă, un model *Random Walk* generează un RMSFE de aproximativ 1 pentru T+1, respectiv 1,1 pentru T+2. Totodată, în medie, modelele bivariate estimate pe baza fiecărui indicator din baza de date prezintă o performanţă similară celor evaluate pe baza primilor cinci factori comuni. Totuşi, trebuie menţionat că în ambele situaţii proiecţiile generate de aceste specificaţii prezintă o variabilitate redusă, fiind adesea în jurul creşterii trimestriale medii a PIB, fapt care sugerează

că setul de informații inclus este insuficient pentru a surprinde, atunci când este necesar, deviațiile semnificative față de evoluția istorică. De exemplu, pentru T1 2020, proiecțiile anterioare deteriorării indicatorilor lunari în contextul izbucnirii pandemiei cu virusul SARS-CoV-2, se plasau între 0,9-1,1 la sută, noile informații, aferente lunii martie 2020, determinând actualizarea proiecției doar până la 0,8 la sută, variație trimestrială. În schimb, modelele estimate folosind multipli factori comuni au fost capabile să surprindă mai exact inversarea tendinței activității economice, acestea actualizând proiecția de la aproximativ 1,5 la sută (obținută pe baza datelor aferente lunilor ianuarie-februarie 2020) până la valori ușor negative, atunci când setul de informații a fost extins prin includerea lunii martie 2020 – datele publicate ulterior de către INS au indicat o creștere minoră a PIB în T1 2020.

4.2. MODELE MIDAS

După cum a fost prezentat anterior, metodologia MIDAS permite utilizarea într-un singur model a seriilor de timp care prezintă frecvențe mixte, fără a fi necesare filtrări suplimentare ale datelor – în cazul modelelor punte, descrise în secțiunea anterioară, numărul de observații al variabilelor este uniformizat, în cazul celor cu frecvență ridicată, prin aplicarea operatorului de medie, în timp ce metodele care se bazează pe filtrul Kalman (a căror performanță de prognoză va fi prezentată în subsecțiunea 4.4) multiplică frecvența cea mai redusă. Regresia MIDAS abordează această problemă prin rearanjarea (spargerea) variabilelor care au o frecvență ridicată astfel încât noile serii de date obținute să prezinte un număr de observații similar variabilei cu frecvență joasă. În cazul de față, în care metodologia este folosită pentru prognoza creșterii trimestriale a PIB pe baza indicatorilor cu frecvență lunară, pentru a putea aplica tehnicile uzuale de estimare a parametrilor regresiei, cei din urmă sunt rearanjați în trei variabile auxiliare (vectori) care prezintă același număr de observații precum variabila cu frecvență redusă. În urma acestei transformări, unei variabile trimestriale îi corespund două *lag*-uri lunare: m_t este valoarea contemporană, cronologic fiind ultima observație asociată trimestrului, m_{t-1} este *lag*-ul de ordin I și reprezintă penultima observație din trimestru, iar m_{t-2} este cel de-al doilea *lag*.

De cele mai multe ori, pentru a surprinde și, ulterior, a prognoza evoluția unei variabile economice într-o manieră satisfăcătoare (în cazul de față PIB) este necesară identificarea unei legături între variabile care persistă pe o perioadă mai lungă de timp. Acest deziderat este obținut în cadrul regresiei MIDAS prin specificarea unui număr mai ridicat de *lag*-uri pentru variabila lunară și/sau prin includerea mai multor serii de timp în model. Astfel, utilizând transformarea prezentată în paragraful anterior, includerea observațiilor care echivalează cu două *lag*-uri trimestriale presupune estimarea a șase parametri aferenți variabilelor lunare rearanjate. Adăugarea unui alt indicator dublează, practic, numărul de parametri care trebuie evaluați, ceea ce conduce la o erodare simțitoare a preciziei procesului de estimare.

Pentru a limita creșterea semnificativă a numărului de parametri, practica uzuală din literatura de specialitate presupune specificarea unor polinoame care dictează relevanța fiecărui *lag* al indicatorului lunar pentru variabila de interes. Două astfel

de funcții folosite pe parcursul evaluării performanței de prognoză a metodologiei MIDAS poartă denumirea de polinom *Exponential Almon*, respectiv de polinom Beta – ambele au fost prezentate pe larg, din punct de vedere teoretic, în subsecțiunea 2.2. În acest fel, în cadrul regresiei MIDAS pot fi incluse mai multe variabile lunare fără ca numărul de grade de libertate să se reducă masiv, întrucât influența *lag*-urilor (indiferent de numărul acestora) asupra variabilei de interes este aproximată cu ajutorul a numai doi parametri pentru fiecare specificație a polinomului (Foroni și Marcellino, 2013 și Ghysels, 2016 prezintă o discuție legată de proprietățile regresiei MIDAS multivariate).

Cu toate acestea, numărul de variabile care pot fi incluse într-un astfel de model este redus, fapt care alimentează unele îngrijorări cu privire la modul de selecție a indicatorilor. În acest sens, pentru a limita efectele nefaste a unor specificații eronate, în această secțiune sunt evaluate: (i) modele bivariate care sunt estimate folosind iterativ toți indicatorii din baza de date, (ii) ecuații cu trei variabile, unde indicatorii incluși sunt extrași aleatoriu din baza de date, avându-se în vedere ca fiecare dintre ei să fie uniform reprezentați în modelele selectate, precum și (iii) specificații care includ cinci variabile. În acest din urmă caz, indicatorii considerați au fost extrași aleatoriu din cinci clustere unde au fost grupați cei din sectorul (a) industrial, (b) comerț și servicii de piață prestate populației, (c) construcții, sector financiar și public, (d) servicii prestate întreprinderilor, precum și (e) unii indicatori care descriu poziția ciclică a economiei (rata șomajului, ROBOR, indicatori de încredere, indicele BET etc). În toate aceste cazuri, proiecția rezultată este agregată pe baza performanței istorice de prognoză a fiecărui model.

O altă modalitate de a limita efectele nefavorabile asociate selecției eronate a indicatorilor este dată de utilizarea componentelor principale, acestea având capacitatea de a sintetiza eficient evoluțiile comune prezente la nivelul setului de indicatori (Marcellino și Schumacher, 2010). Această strategie este evaluată prin utilizarea unei specificații MIDAS nerestricționate, unde sunt estimate modele pe baza primilor zece factori comuni, rezultatul final fiind, de asemenea, agregat pe baza performanței anterioare de prognoză.

În ceea ce privește procedura de obținere a proiecției, trebuie menționat că aceasta este una directă: evoluțiile preconizate pentru PIB nu se bazează pe proiecții ale variabilelor lunare, informațiile necesare fiind acoperite prin modificarea *lag*-urilor indicatorilor lunari incluși în model. Așadar, pentru fiecare orizont de prognoză, după identificarea numărul necesar de *lag*-uri, modelele sunt (re)estimate. Abordarea alternativă constă într-un proces în două etape: inițial, indicatorii lunari sunt prognozați astfel încât să fie obținut numărul necesar de observații infratrimestriale, iar ulterior, seriile de timp prelungite sunt utilizate în modelul MIDAS pentru proiecția PIB. În literatură nu sunt suficiente argumente care să încline balanța în favoarea abordării directe sau a celei alternative, performanța empirică superioară fiind asociată mai de grabă unor factori conjuncturali – de pildă, dacă modelul utilizat pentru prognoza variabilelor lunare are o performanță bună, este posibil ca abordarea în două etape să aibă o performanță empirică superioară; o discuție detaliată este disponibilă în Schumacher, 2016. De menționat că în cazul modelelor punte prezentate în secțiunea anterioară este preferată procedura în două etape, întrucât

utilizarea metodei directe presupune renunțarea la unele informații infratrimestriale care se pot dovedi utile în exercițiile de *nowcasting*.

Tabel 3. Performanța de prognoză (RMSFE) a metodologiei MIDAS

Orizont de prognoză		T+1				T+2
		M0	M1	M2	M3	
Modele bivariate						
MIDAS Almon	lm = 2	0,740	0,729	0,712	0,728	0,739
	lm = 5	0,725	0,708	0,706	0,706	0,744
	lm = 8	0,735	0,716	0,706	0,715	0,744
	lm = 11	0,738	0,732	0,719	0,720	0,749
MIDAS Beta	lm = 8	0,743	0,724	0,734	0,739	0,754
	lm = 11	0,742	0,722	0,736	0,738	0,752
Modele cu trei variabile						
MIDAS Almon	lm = 2	0,768	0,733	0,716	0,734	0,772
	lm = 5	0,762	0,731	0,739	0,728	0,763
	lm = 8	0,772	0,712	0,725	0,716	0,763
	lm = 11	0,764	0,736	0,742	0,729	0,771
MIDAS Beta	lm = 8	0,759	0,715	0,730	0,740	0,767
	lm = 11	0,758	0,714	0,734	0,747	0,767
Modele cu cinci variabile						
MIDAS Almon	lm = 2	0,776	0,737	0,695	0,747	0.776
	lm = 5	0,774	0,710	0,740	0,720	0.761
	lm = 8	0,808	0,738	0,731	0,719	0,776
	lm = 11	0,785	0,742	0,743	0,741	0,779
MIDAS Beta	lm = 8	0,799	0,701	0,743	0,770	0,793
	lm = 11	0,782	0,705	0,739	0,772	0.796
Modele bivariate nerestricționate						
UMIDAS	lm = 2	0,742	0,725	0,724	0,704	0,746
	lm = 5	0,741	0,718	0,710	0,719	0,750
	lm = 8	0,754	0,739	0,732	0,723	0,777
AR UMIDAS	lm = 2	0,804	0,791	0,795	0,778	0,822
	lm = 5	0,813	0,791	0,785	0,781	0,838
	lm = 8	0,820	0,805	0,802	0,787	0,850
FA-UMIDAS	lm = 2	0,745	0,726	0,714	0,700	0,748
	lm = 5	0,749	0,725	0,724	0,720	0,737
	lm = 8	0,798	0,769	0,724	0,715	0,814

Notă: În tabel este prezentată performanța medie de prognoză (RMSFE), calculată începând cu T1 2013 (T2 în cazul orizontului T+2), până în T1 2020. Pentru referință, un model *Random Walk* generează un RMSFE de aproximativ 1 pentru T+1, respectiv 1,1 pentru T+2. „lm" se referă la numărul de *lag*-uri specificat pentru variabila lunară. Coloana notată cu „M0" indică performanța de prognoză în situația în care nicio observație lunară aferentă trimestrului de *nowcasting* nu este disponibilă, iar în coloanele notate cu „M1", „M2", respectiv „M3" setul de informații este extins succesiv cu câte o observație lunară. „MIDAS Almon" presupune utilizarea unui polinom de ponderare *Exponential Almon* cu doi parametri. „AR-UMIDAS" sunt modele bivariate nerestricționate care prezintă un termen autoregresiv de ordin I, iar „FA-UMIDAS" sunt modele bivariate estimate care utilizează primii 10 factori comuni extrași cu ajutorul metodei componentei principale statice. Prognoza pentru T+2 este condiționată de setul de informații disponibil în M3, T+1.

În ceea ce privește performanța de prognoză (Tabel 3), modelele MIDAS care utilizează o funcție polinomială de tipul *Exponential Almon* (cu doi parametri) sunt, în medie, superioare celor estimate cu ajutorul polinomului Beta. O posibilă explicație pentru acest rezultat este dată de gradul mai ridicat de flexibilitate pe care îl prezintă funcția *Almon*, care favorizează o eficiență sporită a procesului de estimare în contextul unui număr redus de observații. În cazul polinomului Beta, identificarea precisă a celor

doi parametri este mai anevoioasă în condițiile unui set restrâns de observații, forma funcțională a acestuia fiind similară distribuției Beta, unde sunt prezente multiple neliniarități.

La o primă vedere, din perspectiva erorii medii de prognoză (RMSFE), modelele MIDAS *Almon* prezintă o performanță de prognoză apropiată, indiferent de numărul de variabile incluse în specificație. Similar celor discutate în secțiunea referitoare la modelele punte, specificațiile MIDAS bivariate au dificultăți în a surprinde unele modificări importante ale traiectoriei indicatorilor lunari, setul de informații introdus în regresie fiind insuficient pentru a genera, în situația în care indicatorii lunari favorizează o astfel de evoluție, o prognoză a creșterii economice în disonanță cu media pe termen lung. Pe baza exemplului folosit în subsecțiunea 4.1, cea mai pesimistă prognoză a modelelor bivariate pentru T1 2020, utilizând informații lunare până în luna martie 2020 (care includ deteriorarea activității economice cauzată de pandemia cu virusul SARS-CoV-2), este în jur de 0,7 la sută, în timp ce regresia MIDAS *Almon* cu trei variabile, precum și cea cu cinci variabile lunare generează proiecții semnificativ mai apropiate de valoarea nulă, în linie cu cea publicată ulterior de către INS.

Actualizarea setului de informații prin includerea observațiilor aferente primului orizont de prognoză (trimestrul de *nowcasting*) aduce câștiguri mai pregnante în termeni de acuratețe a prognozei în cazul modelelor multivariate, în timp ce în unele situații completarea trimestrului cu toate observațiile lunare determină o înrăutățire a calității prognozelor. O posibilă explicație pentru acest fenomen derivă din faptul că, prin funcția polinomială, metodologia favorizează într-o anumită măsură informațiile cele mai recente, care pot fi în disonanță cu evoluția înregistrată pe întreg parcursul trimestrului. Această tendință este mai pronunțată în contextul unui număr redus de *lag*-uri pentru variabilele lunare și/sau în cazul modelelor bivariate. Acordarea unei importanțe mai ridicate ultimelor observații este totuși preferabilă în contextul unui exercițiu de *nowcasting*, eventualele efecte negative ale unor evoluții punctuale contrastante putând fi limitate prin includerea mai multor variabile lunare.

Mergând mai departe la modelele nerestricționate, atât specificațiile simple bivariate, cât și cele cu factori comuni au, în linii mari, o performanță similară cu cele restricționate utilizând polinomul *Almon*. În schimb, introducerea unui termen autoregresiv pentru variabila dependentă conduce la o deteriorare semnificativă a performanței de prognoză, evoluția putând fi explicată de volatilitatea relativ înaltă a dinamicii PIB real în economia locală (care determină un grad redus de autocorelare). În economii unde traiectoria PIB este mai stabilă, introducerea unui termen autoregresiv conduce la o îmbunătățire consistentă a performanței de prognoză (Clements și Galvão, 2009).

4.3. MODELE CU FACTORI DINAMICI CU FRECVENȚE MIXTE

Demersul de evaluare a metodologiilor de prognoză continuă cu ilustrarea performanței modelelor cu factori dinamici cu frecvențe diferite (MF-DFM). În esență, această abordare presupune parcurgerea a două etape. Prima dintre acestea constă în estimarea factorilor comuni, pe baza setului de indicatori lunari și a unui model dinamic (bazat pe filtrul Kalman), cadrul permițând în același timp evaluarea observațiilor lipsă, precum și estimarea și prelungirea factorilor pentru a completa setul de informații necesar, în linie cu orizontul de prognoză ales. În cea de-a doua etapă, factorii comuni sunt puși în relație cu dinamica PIB, o observație trimestrială a acestora fiind dată de o medie ponderată a cinci observații lunare succesive (ecuația (12)).

Așadar, implementarea metodologiei se rezumă la alegerea modalității prin care observațiile lipsă pot fi interpolate, a numărului de factori considerați în regresia finală, precum și a structurii autoregresive a modelului pe baza căruia sunt evaluați factorii comuni. În cazul de față s-a optat pentru utilizarea unui model *state-space* pentru estimarea valorilor lipsă a indicatorilor și a factorilor comuni (similar cu Giannone *et al.*, 2008); o altă abordare, des întâlnită în literatură, este dată de algoritmul *Expectation-Maximisation*, utilizat de Bańbura și Rünstler, 2011. Trebuie menționat faptul că nu sunt disponibile indicii consistente care să releve superioritatea uneia dintre cele două metode în cadrul exercițiilor de prognoză, opțiunea pentru cea dintâi fiind susținută de costul computațional semnificativ mai redus. Performanța de prognoză a acestei metodologii este prezentată în Tabelul 4.

Tabel 4. Performanța de prognoză (RMSFE) a modelelor MF-DFM

Orizont de prognoză		T+1				T+2
		M0	M1	M2	M3	
Ordinul autoregresiv al modelului cu factori $p = 2$						
Număr de factori comuni	F = 2	0,796	0,845	0,776	0.733	0,787
	F = 3	0,787	0,827	0,761	0,723	0,776
	F = 5	0,819	0,776	0,728	0,726	0,815
	F = 7	0,815	0,796	0,739	0,738	0,813
Modele cu cinci factori comuni						
Ordin autoregresiv	p = 1	0,786	0,763	0,737	0,726	0,779
	p = 3	0,822	0,766	0,727	0,722	0,810
	p = 4	0,810	0,763	0,725	0,721	0,801

Notă: În tabel este prezentată performanța medie de prognoză (RMSFE) calculată începând cu T1 2013 (T2 în cazul orizontului T+2), până în T1 2020. Pentru referință, un model *Random Walk* generează un RMSFE de aproximativ 1 pentru T+1, respectiv 1,1 pentru T+2. „F" se referă la numărul de factori comuni utilizați, iar „p" reprezintă ordinul autoregresiv al modelului *state-space* utilizat pentru extragerea factorilor comuni. Coloana notată cu „M0" indică performanța de prognoză în situația în care nicio observație lunară aferentă trimestrului de *nowcasting* nu este disponibilă, iar în coloanele notate cu „M1", „M2", respectiv „M3" setul de informații este extins succesiv cu câte o observație lunară. Prognoza pentru T+2 este condiționată de setul de informații disponibil în M3, T+1.

Din punctul de vedere al erorii medii (RMSFE), modelul cu factori dinamici prezintă o performanță de prognoză relativ similară celei a metodelor prezentate anterior. Îmbogățirea setului informațional prin majorarea numărului de factori comuni introduși în model conduce la ușoare îmbunătățiri ale capacității de prognoză, punctul de inflexiune (pentru orizontul T+1) fiind oferit de specificația cu cinci

factori comuni. Totodată, se remarcă reziliența metodologiei la schimbarea ordinului autoregresiv al modelului pe baza căruia sunt estimați factorii comuni, modificările la acest nivel având doar impact minor asupra acurateței prognozei.

4.4. MODELE VAR BAYESIENE CU FRECVENȚE MIXTE

Ultima categorie de modele avută în vedere se referă la cele VAR Bayesiene, estimate pe baza unui set de indicatori care prezintă frecvențe diferite de eșantionare. În general, modelele VAR sunt intens utilizate în exerciții de prognoză, întrucât structura autoregresivă bine evidențiată de care dispun permite decelarea unor relații stabile prezente la nivelul seriilor de date, fapt ce conduce la o performanță bună, în special pe orizonturi mai lungi. O provocare des întâlnită în utilizarea în practică a acestei metodologii decurge din identificarea unui echilibru între numărul de variabile și cel al *lag*-urilor modelului, o extindere a oricăreia dintre cele două dimensiuni conducând la o creștere importantă a numărului de parametri care trebuie estimați. Soluția propusă în literatura de specialitate pentru a include un număr mai ridicat de indicatori, alături de un ordin autoregresiv suficient de mare, este aceea de a utiliza tehnici Bayesiene de estimare, care presupun că fiecare parametru al modelului aparține unei distribuții cunoscute, iar eventualele informații *a priori* despre forma acesteia permit identificarea mai precisă a parametrului respectiv în setul de date.

Posibilitatea de a include informații *ad-hoc* în procesul de estimare a stimulat literatura de specialitate în a căuta diverse modalități pentru a îmbunătăți capacitatea modelelor de a surprinde evoluțiile prezente în date. O abordare frecvent întâlnită constă în a considera că variabilele modelului evoluează în linie cu un proces autoregresiv de ordinul 1 (al cărui coeficient de inerție este notat cu γ), intensitatea cu care această ipoteză este impusă fiind dată de varianța specificată *a priori* pentru parametri. Mecanismul este util mai ales în contextul unui număr mare de variabile incluse în model, întrucât o restricționare *a priori* a spațiului parametrilor modelului face ca doar aceia care pot fi identificați cu precizie în setul de date să își modifice valoarea, limitându-se astfel riscul de apariție a unor legături false. Gradul de încredere acordat informațiilor *a priori* este ajustat cu ajutorul coeficientului de *tightness*, notat cu λ, o valoare mică a acestuia semnalând o relevanță mai ridicată a distribuției *a priori* – mai multe detalii se regăsesc în subsecțiunea 2.4.

Totodată, tehnicile Bayesiene permit introducerea de informații cu privire la forma matricei de varianță-covarianță a erorilor. Una dintre cele mai des întâlnite abordări este aceea de a considera că aceasta este constantă în timp și are o distribuție *Inverse Wishart* (Bańbura și Rünstler, 2011; Karlsson, 2013). Totuși, în anumite situații, cum ar fi, perioadele în care activitatea economică prezintă o volatilitate ridicată, este de preferat ca o parte din aceasta să fie preluată de erorile modelului, neforțând astfel parametrii să-și schimbe radical legăturile care prevalează pe termen mai lung. Această caracteristică este introdusă în modelul VAR prin considerarea unui factor comun neobservabil și variabil în timp, care afectează matricea de varianță-covarianță a erorilor (Carriero *et al.*, 2016). O astfel de specificație se poate dovedi utilă mai ales

în situația în care modificarea poziției economiei este surprinsă de un număr redus de observații.

Trecând la provocările legate de selecția indicatorilor care ar trebui incluși în model, pentru a limita efectele negative ale omiterii unor variabile relevante pentru proiecția PIB, abordarea preferată în această lucrare este aceea de a estima o suită de modele (cinci, în total) care explorează principalii indicatori din baza de date – mai multe detalii cu privire la variabilele incluse în modelele VAR sunt prezentate în Anexă. Fiecare model avut în vedere este estimat pe baza a 11 indicatori, prognoza finală fiind o agregare în funcție de performanța anterioară de prognoză. Totodată, trebuie menționat că, spre deosebire de celelalte metode prezentate în secțiunile anterioare, explorarea tuturor indicatorilor din baza de date este extrem de costisitoare din punct de vedere al duratei în timp a procedurii de estimare, datorită complexității modelului și a procesului de estimare.

Tabel 5. Performanța de prognoză (RMSFE) a modelelor MF-BVAR

Orizont de prognoză			T+1				T+2
			M0	M1	M2	M3	
Coeficientul de *tightness* $\lambda = 0{,}1$; Coeficient autoregresiv $\gamma = 0$							
Numărul de *lag*-uri lunare	CSV	lm = 3	0,816	0,740	0,730	0,710	0,829
		lm = 6	0,826	0,760	0,745	0,721	0,838
		lm = 9	0,857	0,769	0,750	0,733	0,856
	IW	lm = 3	1,034	0,859	0,769	0,740	0,947
		lm = 6	0,996	0,818	0,748	0,710	0,917
		lm = 9	1,025	0,836	0,765	0,721	0,923
Numărul de *lag*-uri lunare lm = 6; Coeficient autoregresiv $\gamma = 0$							
Coeficientul de *tightness*	CSV	$\lambda = 0{,}05$	0,827	0,771	0,743	0,708	0,833
		$\lambda = 0{,}15$	0,825	0,762	0,748	0,719	0,837
		$\lambda = 0{,}30$	0,825	0,769	0,750	0,715	0,847
	IW	$\lambda = 0{,}05$	1,242	0,938	0,837	0,788	1,101
		$\lambda = 0{,}15$	0,950	0,804	0,731	0,693	0,906
		$\lambda = 0{,}30$	0,944	0,829	0,768	0,733	0,935
Numărul de *lag*-uri lunare lm = 6; Coeficientul de *tightness* $\lambda = 0{,}1$							
Coeficient autoregresiv	CSV	$\gamma = 0{,}5$	0,817	0,755	0,724	0,707	0,832
		$\gamma = 1{,}0$	0,948	0,896	0,878	0,836	0,931
	IW	$\gamma = 0{,}5$	0,953	0,804	0,723	0,686	0,919
		$\gamma = 1{,}0$	0,986	0,901	0,855	0,812	0,950

Notă: În tabel este prezentată performanța medie de prognoză (RMSFE), calculată începând cu T1 2013 (T2 în cazul orizontului T+2), până în T1 2020. Pentru referință, un model *Random Walk* generează un RMSFE de aproximativ 1 pentru T+1, respectiv 1,1 pentru T+2. „lm" se referă la numărul de *lag*-uri specificat pentru modelul VAR lunar, iar λ este coeficientul de *tightness*, o valoarea mai mică a acestuia semnalând o importanță mai ridicată asociată acesteia în distribuția *a posteriori* a modelului. „CSV" (engl. *common stochastic volatility*) se referă la modele care prezintă un factor stohastic la nivelul matricei de varianță-covarianță, în timp ce „IW" semnalează o matrice a erorilor constantă în timp, distribuită *Inverse Wishart*. Coloana notată cu „M0" consemnează performanța de prognoză în situația în care nicio observație lunară aferentă trimestrului de *nowcasting* nu este disponibilă, iar în coloanele notate cu „M1", „M2", respectiv „M3", setul de informații este extins succesiv cu câte o observație lunară. Prognoza pentru T+2 este condiționată de setul de informații disponibil în M3, T+1.

Performanța de prognoză a modelelor MF-BVAR (Tabel 5) prezintă anumite modificări în funcție de ipotezele asociate matricei de varianță-covarianță a erorilor. Astfel, forma mai flexibilă adusă de specificarea unui factor stohastic comun contribuie la o îmbunătățire a performanței de prognoză a metodologiei MF-BVAR, tendință care

se menține și în cazul în care noi observații lunare sunt incluse în setul de informații, dar și în contextul în care ceilalți parametri relevanți pentru procesul de estimare sunt modificați. Totodată, trebuie menționat că pe măsură ce noi observații lunare infratrimestriale devin disponibile, modelele care prezintă o structură constantă a matricei de varianță-covarianță încep să surprindă din ce în ce mai bine impactul evoluțiilor indicatorilor lunari, reușind astfel să recupereze din ecartul prezent între cele două metodologii în prima fază a exercițiului de prognoză. În anumite situații, de pildă atunci când coeficientul de *tightness* este fixat la $\lambda = 0{,}1$, viteza de îmbunătățire a prognozelor este suficient de ridicată pentru ca modelele cu erori constante în timp să prezinte performanțe superioare celeilalte abordări atunci când toate evoluțiile lunare din cadrul trimestrului de *nowcasting* sunt disponibile. O posibilă explicație pentru acest fenomen este dată de faptul că, pe măsură ce setul de informații este suficient de extins, parametrii se ajustează în consecință și nespecificarea unei structuri variabile în timp (care adaugă simțitor la numărul de parametri ce trebuie evaluați) conduce la o majorare a gradelor de libertate, iar mai departe la o precizie de estimare mai înaltă. Așadar, în situația în care nu intervin modificări importante de dinamică la nivelul observațiilor lunare, beneficiile includerii unei structuri a erorilor variabile în timp par a se reduce. Cu toate acestea, trebuie menționat că intervalul pe baza căruia au fost calculate erorile (circa 30 de observații) nu este unul suficient de lung pentru a permite atenuarea unor eventuale performanțe conjuncturale ale modelelor.

Totodată, exercițiul de prognoză arată o sensibilitate relativ redusă a metodologiei față de modificarea structurii autoregresive, erorile generate fiind, în linii mari, similare, indiferent dacă sunt considerate 3, 6, sau 9 *lag*-uri pentru modelul VAR lunar – diferențele nu sunt suficient de mari pentru a conchide că o anumită specificație este superioară pe toate planurile. O concluzie similară se cristalizează și în cazul modificării coeficientul de *tightness*, în special în situația în care se optează pentru includerea de erori care prezintă o varianță variabilă în timp, abordarea care presupune erori constante arătând o sensibilitate ușor mai pronunțată la modificarea acestui parametru. În ceea ce privește valoarea termenului autoregresiv asociat distribuției *a priori*, utilizarea unei valori de $\gamma = 0{,}5$ are ca rezultat o ameliorare a erorii de prognoză, în timp ce valoarea unitară $\gamma = 1$ favorizează o inerție mai ridicată a procesului autoregresiv, ceea ce conduce la o erodare a puterii de predicție.

5. Combinarea prognozelor

Metodologiile prezentate în secțiunile anterioare au o capacitate ridicată de a încorpora un set consistent de informații în procesul de prognoză a creșterii economice. Însă, modelele econometrice reprezintă doar aproximări ale procesului de generare a datelor, omisiunile inerente la nivelul setului de informații utilizat conducând astfel la un anumit grad de eroare la nivelul specificației, cu efecte nefavorabile asupra acurateței prognozelor. De asemenea, unele caracteristici ale seriilor de timp – cum ar fi, persistența unui anumit grad de nestaționaritate (în pofida aplicării unor transformări la nivelul datelor care ar trebui să elimine astfel de efecte) sau un număr

limitat de observații – afectează semnificativ calitatea procesului de estimare (Larsson *et al.*, 2001; Hendry și Clements, 2003).

O altă dificultate des întâlnită în practică se referă la modul în care sunt tratate rupturile structurale prezente la nivelul setului de indicatori. Exemple în acest sens sunt furnizate de modificări la nivelul metodologiilor statistice, care nu sunt implementate la nivelul întregului set de observații, sau de tranziția economiei către un nou model de creștere, care reduce impactul unui anumit sector economic; în acest din urmă caz, dificultăți suplimentare sunt aduse de o anumită lentoare a procesului transformațional, care împiedică o datare precisă. Prin urmare, în funcție de momentul în care aceste modificări sunt vizibile la nivelul setului de indicatori sunt posibile mai multe abordări. Dacă, de exemplu, acestea intervin în prima parte a eșantionului de date, o opțiune ar fi excluderea observațiilor până la acel moment, strategia fiind viabilă doar în situația în care setul de informații rămas este suficient de extins pentru a asigura o estimare robustă. O alternativă poate fi reprezentată de utilizarea unor metodologii care prezintă parametri și/sau structură variabilă în timp a erorilor, aceasta oferind rezultate favorabile și în situația în care modificările structurale sunt fie multiple, fie se petrec mai recent, iar setul de informații este suficient de consistent pentru a asigura evaluarea adecvată a noului regim. Din punct de vedere al acurateței prognozelor, performanța acestei abordări se reduce simțitor dacă legăturile dintre variabile se deteriorează în perioada foarte recentă (de pildă, pe parcursul ultimului an al eșantionului), caz în care informațiile pe baza cărora noul regim poate fi evaluat sunt insuficiente (Pesaran și Timmermann, 2007 prezintă mai multe detalii cu privire la acest aspect). În oricare dintre situațiile menționate anterior, combinarea proiecțiilor generate de multiple metodologii reprezintă o soluție viabilă pentru menținerea pe un palier mai stabil al performanței de prognoză pe un orizont mai lung de timp, indiferent dacă este preferată agregarea unei singure valori sau a întregii distribuții a prognozelor. Hendry și Clements, 2004; Timmermann, 2006; Clark și McCracken, 2010 prezintă pe larg avantajele combinării prognozelor.

Din suita de metode considerate în această lucrare, modelele VAR manifestă cea mai mare sensibilitate la modificarea legăturilor dintre variabilele economice. Această metodologie presupune utilizarea unei structuri autoregresive complexe care reclamă estimarea unui număr ridicat de parametri, aceștia necesitând informații consistente despre eventuale schimbări de structură la nivelul economiei pentru a se actualiza. Mai precis, numărul de observații al noului echilibru trebuie să fie suficient de ridicat pentru a conduce la o modificare semnificativă a parametrilor, media acestora fiind influențată decisiv de observațiile aferente vechiului regim; Cogley și Sargent, 2005 prezintă o discuție detaliată în acest sens. Această sensibilitate a metodologiei a reprezentat un motiv suplimentar pentru care în subsecțiunea 4.4 evaluarea performanței de prognoză a modelelor de acest tip (chiar dacă s-a considerat și o structură variabilă în timp pentru erori) s-a realizat pe baza a cinci specificații care acoperă uniform cei mai importanți indicatori din baza de date. În același timp, eventualele schimbări structurale afectează și modul în care sunt extrași factori comuni, însă impactul asupra robusteții estimărilor, și mai departe asupra prognozelor, este sensibil mai redus (Bates *et al.*, 2013; Baltagi *et al.*, 2021).

5.1. GENERAREA ȘI COMBINAREA DENSITĂȚILOR DE REPARTIȚIE A PROGNOZELOR

De cele mai multe ori rezultatul exercițiilor de prognoză este reprezentat de o valoare medie (sau mediană) generată de metodologiile utilizate. Prin această alegere se presupune implicit o funcție de pierdere simetrică, deviații ale valorii efective în oricare dintre cele două direcții prezentând pentru agenți un cost similar (asociat unor eventuale decizii nefavorabile cauzate de prognoze inexacte); o discuție mai detaliată este prezentată în Capitolul 3 din Elliott și Timmermann, 2016. Abordări mai complexe, care aproximează mai fidel eventualele costuri întâmpinate în practică, presupun specificarea unei funcții de pierdere asimetrice. De pildă, o subestimare a creșterii economice poate avea un impact social mai ridicat (presiunile inflaționiste care însoțesc expansiunea economiei nu au fost identificate la timp) comparativ cu o supraestimare. În acest regim, procesul obținerii proiecției optime (i.e. valoarea care minimizează funcția de pierdere) poate fi extrem de complicat, o soluție analitică fiind de puține ori disponibilă – Granger, 1999; Elliott *et al.*, 2005. Totodată, prezența unor preferințe asimetrice pune sub semnul întrebării rezultatul standard din literatura de combinare a proiecțiilor (v. Stock și Watson, 1999), care indică schemele simple de ponderare ca fiind printre cele mai performante abordări din perspectiva acurateței proiecției (Elliott și Timmermann, 2004 prezintă mai multe detalii în acest sens).

Utilizarea densităților de repartiție a prognozelor (care reprezintă, în esență, o estimare cu privire la realizările viitoare ale unei variabile aleatoare, condiționată de setul de informații din prezent) elimină necesitatea specificării unei funcții de pierdere – rolul acesteia din urmă fiind acela de a focaliza atenția asupra unei anumite zone a distribuției (teoretice) a proiecțiilor. În acest context, totuși, evaluarea performanței de prognoză, necesară pentru a ierarhiza diferitele metodologii, este mai dificil de realizat. O metodă des utilizată în practică în cazul densităților de repartiție a proiecțiilor presupune transformarea acestora într-o distribuție uniformă (engl. *integral probability transform*) pentru a testa dacă noile valori sunt independente și identic uniform distribuite cu medie 0 și varianță 1. În acest fel se apreciază dacă proiecțiile observate aparțin unei distribuții similare celei teoretice, generată de modelul corect specificat în raport cu procesul de generare a datelor (mai multe detalii sunt disponibile în Diebold *et al.*, 1998; Berkowitz, 2001). Totuși, în practică, există o posibilitate destul de ridicată ca modelul cu ajutorul căruia a fost realizată proiecția să prezinte un anumit grad de eroare la nivelul specificației (engl. *misspecification*). În acest context, soluția oferită de literatura de specialitate este aceea de a utiliza funcții de *scoring*, care contabilizează într-o singură valoare pierderea asociată unei proiecții eronate. Mai mult, pe baza funcțiilor de *scoring* se pot construi teste statistice care evaluează capacitatea modelelor de a prognoza evenimente în anumite zone ale distribuției. În acest sens, poate prezenta interes măsura în care un anumit model are capacitatea de a anticipa valori pentru creșterea economică plasate în jurul mediei istorice sau în alte segmente ale distribuției (Amisano și Giacomini, 2007).

Dintre metodologiile utilizate în această lucrare doar MF-BVAR permite evaluarea concomitentă a parametrilor modelului și a densităților de repartiție a prognozelor

$f(Y_{n,h}|\Omega_\tau)$, întrucât, pentru fiecare extragere din distribuția *a posteriori*, structura autoregresivă a modelului asigură obținerea unei prognoze pentru orizontul *h*, condiționat de setul de informații din Ω_τ, unde $\tau = 1:T:M$; *T* reprezintă ultima perioadă de timp pentru care sunt disponibile informații trimestriale, iar *M* este ultima lună din perioada *nowcasting* ($T+1$) pentru care sunt publicați indicatori statistici. Trebuie menționat faptul că proiecțiile $f(Y_{n,h}|\Omega_\tau)$ sunt calculate prin aplicarea operatorului de medie aritmetică simplă la nivelul prognozelor lunare $f\left(Y_{n,hm}\middle|\Omega_\tau\right)$ obținute prin iterarea în avans a modelului VAR lunar pentru orizontul *hm*, reprezentat de ecuația (15) – modelul MF-BVAR utilizează indicatorii în nivel.

Trecând la metodele cu factori comuni, unde se apelează la tehnici clasice de estimare, distribuția prognozelor $f(Y_{n,h}|\Omega_\tau)$ este evaluată ulterior estimării modelului. Procedura, care este o simulare de tip *Monte Carlo*, presupune recuperarea mediei și a variației coeficienților regresiei pentru a fi folosite pentru a iniția extrageri din distribuția normală multivariată, pe baza cărora sunt calculate multiple traiectorii pentru variabila de interes, ținându-se totodată seama de abilitatea modelului de a surprinde evoluțiile din date, precum și de eventuale corelații între parametrii acestuia. În cazul metodologiei MIDAS, întrucât abordarea prezintă un grad mai ridicat de complexitate (ecuația de regresie fiind una neliniară), simularea traiectoriei prognozelor se realizează prin aplicarea unor tehnici de *bootstrapping* la nivelul erorilor (Aastveit *et al.*, 2017).

O abordare des întâlnită în ceea ce privește combinarea mai multor densități de repartiție se numește *linear opinion pool* (Timmermann, 2006 prezintă și alte metodologii disponibile) și presupune o agregare convexă a distribuției prognozelor ($n = 1 \dots N$):

$$p\left(Y_{\tau,h}\right) = \sum_{n=1}^{N} \widetilde{\omega}_{n,\tau,h}\, f\left(Y_{n,hm}\middle|\Omega_\tau\right) \tag{25}$$

unde $\widetilde{\omega}_{n,\tau,h}$ sunt ponderile de agregare care trebuie să fie cuprinse între $0 \le \widetilde{\omega}_{n,\tau,h} \le 1$ și suma lor $\sum_{n=1}^{N} \omega_{n,\tau,h} = 1$. Această din urmă condiție garantează faptul că rezultatul procesului de agregare este o distribuție; totodată, procedura nu impune restricții cu privire la forma teoretică a densităților considerate.

Un avantaj important al combinării densităților de repartiție este reprezentat de faptul că distribuția rezultată poate prezenta asimetrii, acomodând astfel multiple caracteristici observate în practică (Aastveit *et al.*, 2019), chiar dacă cele considerate inițial erau simetrice. De pildă, dacă densitățile de repartiție ale prognozelor modelelor individuale sunt normale, cea rezultată de pe urma procesului de combinare poate prezenta multiple cocoașe, cozi leptokurtotice sau alte tipuri de asimetrii, noua distribuție purtând denumirea de „mixtură" (Boero *et al.*, 2011). Media acesteia este dată de combinarea mediilor densităților individuale utilizând ponderile de agregare $\omega_{n,\tau,h}$, în timp ce în cazul deviației standard noua valoare include, pe lângă componenta agregată, și una asociată discrepanțelor dintre momentele distribuțiilor (a se vedea *inter alia* Hall și Mitchell, 2007). Prin urmare, varianța distribuției rezultate este superioară celei care ar fi prevalat doar prin agregarea simplă a momentelor de ordin II

ale distribuțiilor, fapt care contrastează cu rezultatul teoretic obținut în cazul în care se combină doar tendințe centrale, pentru care varianța rezultată este cel puțin egală cu varianțele prognozelor individuale (mai multe detalii cu privire la acest din urmă aspect se găsesc în Granger și Ramanathan, 1984; Timmermann, 2006). O altă metodă des utilizată pentru a genera distribuții asimetrice (potrivită mai cu seamă în situația folosirii unui singur model) presupune specificarea unei distribuții normale formate din două componente care prezintă o medie comună și două deviații standard pentru cele două părți ale distribuției – Banca Angliei utilizează o astfel de ipoteză pentru construirea distribuției prognozelor, i.e. *fan chart*, (Wallis, 2004).

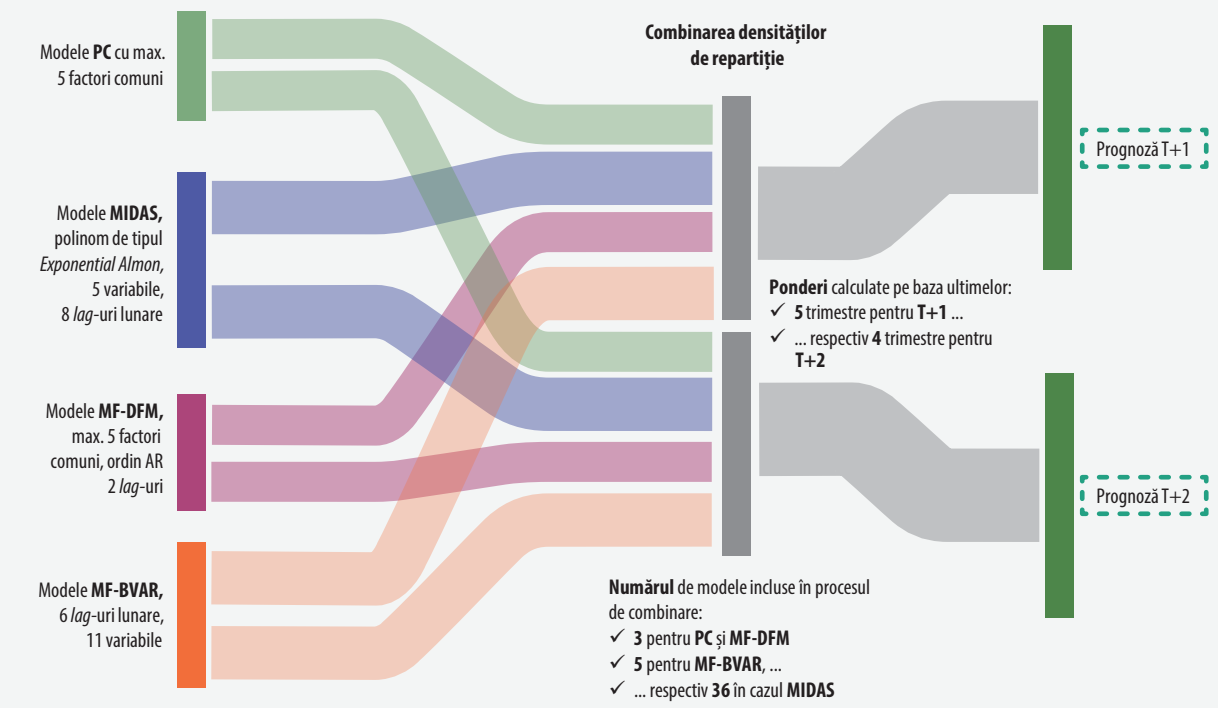
Principala problemă practică ce trebuie depășită în situația combinării densităților de repartiție a prognozelor se referă la modul în care sunt calculate ponderile de agregare $\omega_{n,\tau,h}$. Similar rezultatelor obținute de literatura de specialitate care tratează această problemă din perspectiva combinării mediei proiecțiilor, o opțiune fezabilă este aceea de a considera ponderi egale pentru fiecare model $\omega_{n,\tau,h} = 1/n$ (Hendry și Clements, 2004). Alte propuneri se referă la utilizarea metodologiei *Bayesian Model Averaging* care evaluează ponderile din perspectiva capacității modelului de a surprinde evoluțiile (istorice) observate la nivelul setului de indicatori (Garratt *et al.*, 2003; Boero *et al.*, 2011) sau la estimarea acestora utilizând tehnici de optimizare care minimizează distanța dintre distribuția realizată a proiecțiilor și cea teoretică (Hall și Mitchell, 2007; Geweke și Amisano, 2011). Această din urmă abordare, deși prezintă fundamente teoretice solide, este dificil de implementat în cazul unui număr ridicat de modele, mai ales dacă este prezent și un anumit grad de corelare la nivelul prognozelor (Jore *et al.*, 2010), care apare, de pildă, în situația utilizării unui set similar de indicatori.

În această lucrare, metoda selectată pentru calcularea ponderilor ($\omega_{n,\tau,h}$) se ancorează la performanța de prognoză din trecut a modelelor, evaluată utilizând funcții de *scoring,* atenția fiind îndreptată către cea denumită *Continuous Ranked Probability Score* (CRPS). Această variantă este de preferat în detrimentul celei alese de cele mai multe ori în literatură, care presupune utilizarea funcțiilor de *scoring* logaritmice, întrucât cele din urmă prezintă unele neajunsuri în situația în care valorile realizate ale variabilei de interes sunt plasate în jurul cozilor distribuției (Gneiting și Raftery, 2007). În esență, această strategie de calculare a ponderilor este similară celei utilizate în situația combinării mediei prognozelor (unde sunt vizate erorile medii), întrucât funcția de *scoring* sintetizează într-un singur număr ecartul dintre probabilitatea asociată valorii realizate și cea a mediei distribuției – se presupune că o valoare mai mică a funcției de *scoring* reflectă o performanță superioară de prognoză. Așadar, pe baza acestor evaluări, ponderile de agregare pot fi calculate ținând seama de performanța modelelor de-a lungul mai multor intervale de timp. Focalizarea asupra unei perioade mai recente (engl. *recursive weights*), de exemplu, oferă multiple avantaje în situația în care performanța modelelor este afectată de schimbări structurale la nivelul economiei. Jore *et al.*, 2010 arată că, spre deosebire de ponderi constante în timp, cele care se actualizează mai frecvent oferă o acuratețe de prognoză mai ridicată, vizibilă la nivelul tuturor fazelor ciclului economic; această strategie surprinde schimbările la nivelul structurii economiei prin alocarea unor ponderi mai însemnate modelelor care sunt mai robuste în acest context.

5.2. PERFORMANȚA DE PROGNOZĂ

Evaluările performanței de prognoză *out-of-sample* prezentate în Secțiunea 4 oferă o imagine mai fidelă cu privire la capacitatea metodologiilor considerate în această lucrare de a surprinde evoluțiile viitoare ale creșterii economice. Aceste evaluări permit alegerea celor mai potrivite specificații pentru fiecare categorie de modele, care urmează a fi reținute în cadrul final de prognoză, unde sunt combinate toate densitățile de repartiție.

Grafic 3. Ilustrarea procesului de combinare a proiecțiilor



Prin urmare, în cazul modelelor punte, sunt considerate specificații care includ 2, 3, respectiv 5 factori comuni, toate acestea prezentând o performanță similară de prognoză. O deteriorare este remarcată în cazul modelelor cu 7 factori comuni, în timp ce modelele bivariate generează proiecții care nu deviază semnificativ de la media pe termen lung a creșterii economice (ceea ce sugerează, printre altele, includerea unui set de informații deficitar la nivelul modelelor). O strategie similară este utilizată și pentru modelele cu factori comuni dinamici, fiind incluse specificații cu 2, 3, respectiv 5 factori comuni și un ordin autoregresiv de 2 *lag*-uri pentru modelul de extragere și prelungire a factorilor; în acest caz, diferențele din punct de vedere al acurateței prognozei nu sunt atât de evidente. În cazul modelelor MIDAS sunt preferate specificațiile care utilizează funcții polinomiale de tip *Exponential Almon* cu doi parametri, cinci variabile și opt *lag*-uri lunare. Din perspectiva perioadei anterioare pandemiei, această specificație nu generează cea mai redusă valoare pentru RMSFE (Tabel 3), însă, spre deosebire de celelalte variante evaluate, oferă o performanță superioară dacă sunt incluse și observațiile de pe parcursul crizei COVID-19 – RMSFE mediu generat în această situație este cu 15 la sută mai redus comparativ cu modelele cu trei variabilele, ambele fiind superioare celorlalte specificații considerate.

Totuși, pentru a îmbunătăți capacitatea de prognoză în perioade de evoluție stabilă a economiei (similar celei prepandemice) sunt eliminate 24 de modele care performează nesatisfăcător (30 la sută din numărul total). În cazul metodologiei MF-BVAR s-au avut în vedere specificații care prezintă un ordin autoregresiv de 6 *lag*-uri lunare, un coeficient de *tightness* de 0,1, precum și un factor comun de scalare pentru matricea de varianță-covarianță a erorilor variabil în timp; această variantă a fost preferată utilizării unor erori constante întrucât oferă performanțe superioare de prognoză în cazurile în care observațiile lunare aferente trimestrului de *nowcasting* sunt incomplete. În total sunt estimate cinci specificații utilizând câte 11 variabile alese dintr-un set restrâns de indicatori, prezentați în Anexă.

Mai departe, densitățile de repartiție generate de modele sunt combinate în funcție de performanța anterioară de prognoză, ponderile fiind evaluate cu ajutorul funcției de *scoring* CRPS. Perioada avută în vedere pentru calcularea acestora este de cinci trimestre pentru orizontul T+1, respectiv patru pentru T+2. Totodată, ponderile au fost ajustate pentru a acomoda numărul diferit de modele specificat pentru fiecare metodologie (cu $1/n_M$, unde n_M este numărul de modele pentru fiecare *M* categorie) – în subsecțiunea 5.3 sunt prezentate multiple evaluări cu privire la robustețea performanței de prognoză în contextul unor modificări la nivelul schemei de ponderare.

Tabel 6. Performanța comună de prognoză

Orizont de prognoză	T+1				T+2
	M0	M1	M2	M3	
T1 2013 – T1 2020					
Performanța comună	0,468	**0,439**	0,417	**0,415**	0,475
PC	0,485	0,462	**0,415**	0,416	0,492
MIDAS	0,486	0,442	0,444	0,439	0,486
MF-DFM	**0,464**	0,471	0,426	0,421	**0,467**
MF-BVAR	0,485	0,445	0,423	0,429	0,497
T2 2020 – T3 2020					
Performanța comună	5,539	1,864	2,196	**2,454**	5,900
PC	5,605	**0,838**	**2,140**	2,654	5,653
MIDAS	5,779	3,191	3,888	3,971	7,357
MF-DFM	**5,473**	9,581	4,448	3,455	**5,458**
MF-BVAR	5,637	2,268	2,332	2,609	6,508

Notă: În tabel este prezentată valoarea medie a funcției CRPS (engl. *Continuous Rank Probability Score*) multiplicată cu 100 și care reprezintă o generalizare a unei matrici simple precum eroare medie de prognoză la cazul mai sofisticat al prognozelor de probabilități. O valoarea mai mică indică o performanță superioară de prognoză. În coloana notată cu „M0", setul de informații disponibile nu include nicio observație lunară asociată trimestrului de *nowcasting* (T+1). Coloana „M1", „M2", respectiv „M3" indică numărul de observații lunare disponibile în cadrul unui trimestru. „PC" reprezintă modele punte estimate cu factori comuni. Prognoza pentru T+2 este condiționată de setul de informații disponibil în M3, T+1.

Performanța de prognoză a fiecărei categorii de modelele, precum și a distribuției agregate, din perspectiva funcțiilor de *scoring*, este ilustrată în Tabelul 6; o valoare inferioară sugerează o acuratețe superioară. În perioada prepandemică (T1 2013 – T1 2020), distribuția combinată oferă cea mai bună performanță de prognoză pentru trimestrul de *nowcasting* în 2 din 4 situații; în celelalte două cazuri, se observă o diferență extrem de redusă față de metoda care prezintă cea mai ridicată acuratețe de prognoză – MF-DFM în situația în care nicio observație lunară nu este disponibilă, respectiv PC când sunt incluse două observații lunare. În ceea ce privește performanța

de prognoză pe parcursul primului val pandemic (T2-T3 2020), doar într-o singură instanță distribuția combinată generează cea mai mică valoare a funcției de *scoring*, în celelalte cazuri fiind aproape de cea mai performantă metodă. Pentru orizontul T+2, metodologia MF-DFM pare a produce cele mai mici valori pentru funcția de *scoring*.

În lipsa altor evaluări, ierarhizarea modelelor pe baza mediei valorilor funcției de *scoring* este o abordare potrivită, însă trebuie avut în vedere că, în cazul unui eșantion redus de proiecții *out-of-sample*, nivelul mediu poate fi semnificativ influențat de o singură instanță în care un anumit model a avut o performanță deosebită. În acest caz, prezintă interes construcția unor teste statistice care să evalueze riguros acuratețea de prognoză a modelelor. O astfel de metodă, propusă de Amisano și Giacomini, 2007, presupune testarea performanței de prognoză prin evaluarea mediei diferențelor funcției de *scoring* calculată pentru două densități de repartiție concurente, în raport cu deviație standard observată la nivelul eșantionului – această procedură este similară testului DM aplicat în cazul în care se evaluează doar media proiecțiilor, Diebold și Mariano, 2002. Din perspectiva acestei proceduri, nicio categorie de modele nu oferă o performanță superioară comparativ cu prognoza distribuției combinate, atât din punct de vedere al unor evoluții extreme ale creșterii economice (din perspectivă istorică), cât și a celor plasate în apropierea mediei pe termen lung (Tabel 7). Totodată, performanța mai ridicată a modelelor MF-DFM pentru orizontul T+2 poate fi atribuită capacității acestora de a surprinde mai bine evoluții neobișnuite ale creșterii economice. Totuși, trebuie menționat că în situația de față, dat fiind numărul relativ redus de prognoze *out-of-sample*, puterea testului este, într-o anumită măsură, limitată (Amisano și Giacomini, 2007 prezintă mai multe detalii în acest sens) – rezultatele asimptotice prezentate au fost în mare parte confirmate de versiunea testului care utilizează tehnici de *bootstrapping* pentru evaluarea distribuției asociate ipotezei nule (Giacomini, 2002).

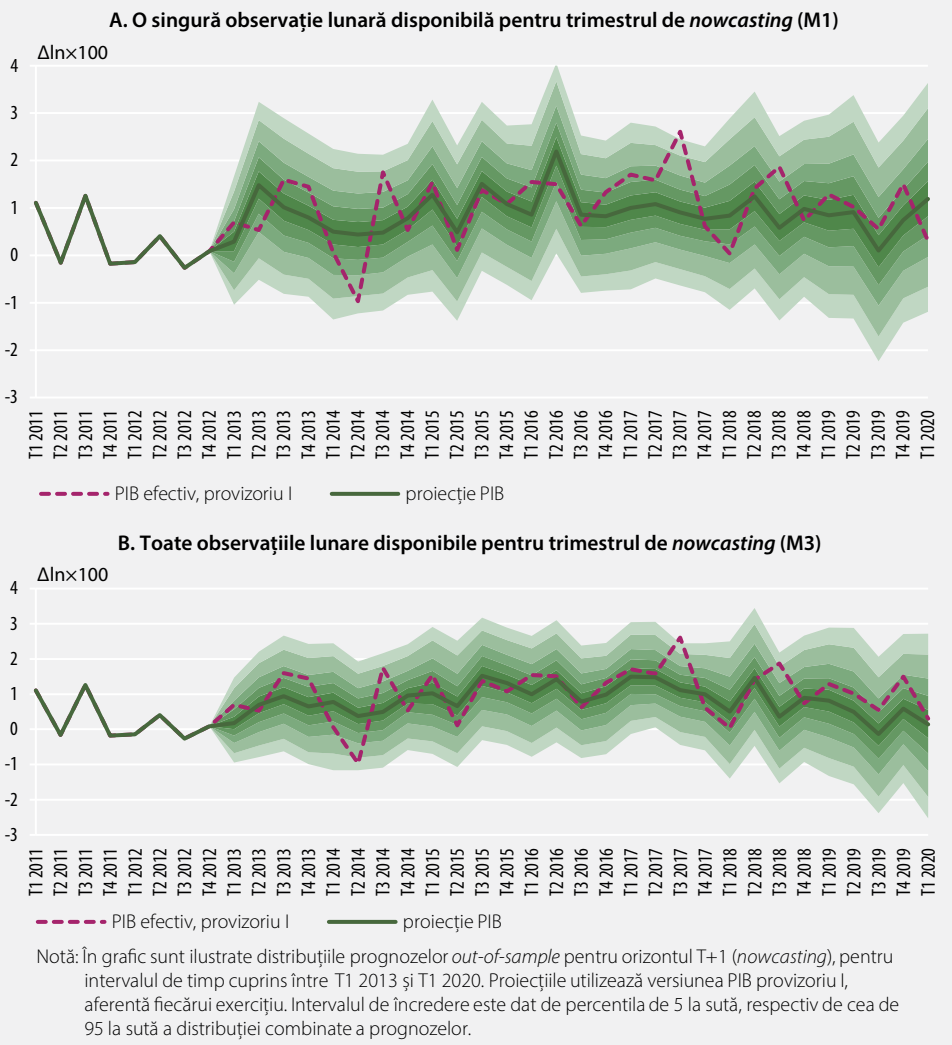
Tabel 7. Statistica testului WLR privind acuratețea de prognoză a densităților de repartiție

Orizont de prognoză	T+1				T+2
	M0	M1	M2	M3	
Statistica testului WLR pentru media distribuției					
PC	0,46	1,21	0,05	0,92	1,12
MIDAS	1,64	0,11	1,38	0,97	1,39
MF-DFM	-0,87	1,99	1,53	1,56	-0,30
MF-BVAR	1,18	0,86	0,18	0,46	-0,02
Statistica testului WLR pentru cozile distribuției					
PC	1,70	-0,94	-0,26	0,54	-0,35
MIDAS	1,15	1,02	0,99	0,99	1,79
MF-DFM	0,24	1,01	1,59	2,16	-1,98
MF-BVAR	-0,48	0,90	0,41	0,42	1,65

Notă: În tabel este prezentată statistica testului WLR (*Weighted Likelihood Ratio*) similar celui propus de Amisano și Giacomini, 2007. Ipoteza nulă presupune o performanță egală de prognoză a distribuției combinate comparativ cu cea generată de metodologiile individuale. O valoare pozitivă semnalează o performanță sporită de prognoză a celei dintâi, în timp ce una negativă, contrariul. Valoarea critică este ±1,64 pentru un interval de încredere de 90 la sută. Proiecțiile *out-of-sample* acoperă perioada cuprinsă între T1 2013 și T3 2020.

Performanța cadrului de prognoză în ceea ce privește anticiparea evoluției creșterii economice este ilustrată în Graficul 4, unde sunt prezentate prognozele pentru trimestrul de *nowcasting* în situația în care setul de informații include o singură observație (Panel A), respectiv toate observațiile lunare (Panel B). În grafic sunt vizibile două momente în care valoarea efectivă a creșterii economice s-a plasat la extremitățile distribuției prognozelor (T2 2014, respectiv T3 2017), în ambele situații dinamica trimestrială fiind exacerbată de evoluțiile consemnate în agricultură. Din acest punct de vedere, cadrul de prognoză prezintă o anumită sensibilitate întrucât setul de indicatori nu include informații suficiente care pot ajuta modelele să anticipeze într-o manieră adecvată traiectoria sectorului agricol – excluderea acestui sector din PIB conduce într-adevăr la o performanță superioară de prognoză a metodologiei; mai multe detalii sunt prezentate în subsecțiunea 5.3.

Grafic 4. Performanța de prognoză pentru trimestrul T+1 (*nowcasting*)

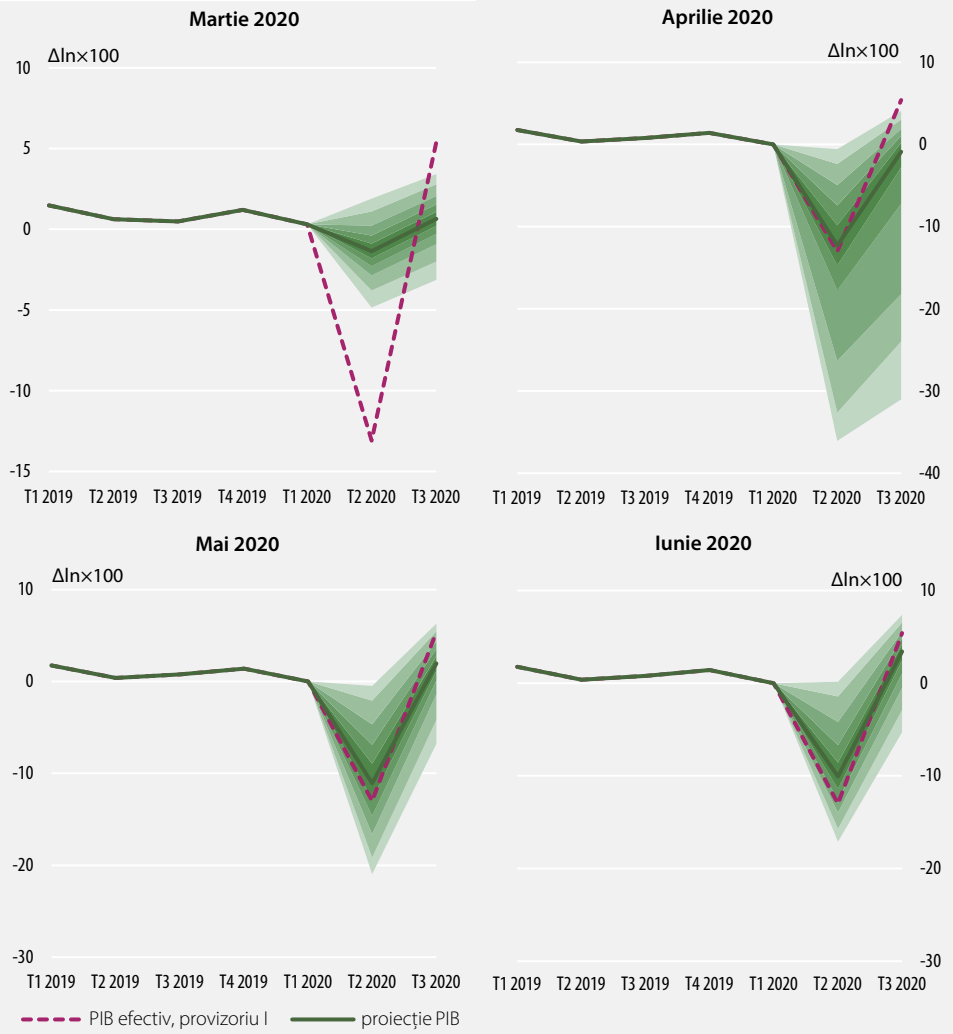


Notă: În grafic sunt ilustrate distribuțiile prognozelor *out-of-sample* pentru orizontul T+1 (*nowcasting*), pentru intervalul de timp cuprins între T1 2013 și T1 2020. Proiecțiile utilizează versiunea PIB provizoriu I, aferentă fiecărui exercițiu. Intervalul de încredere este dat de percentila de 5 la sută, respectiv de cea de 95 la sută a distribuției combinate a prognozelor.

Relevanța metodologiei de monitorizare în timp real crește în momentele imediat următoare materializării unor șocuri asupra activității economice. În cel mai recent episod de acest fel, cauzat de răspândirea alertă a virusului SARS-CoV-2, deteriorarea indicatorilor lunari s-a petrecut începând cu luna martie 2020. În aceste condiții,

în momentul în care au fost introduse observațiile lunare cu privire la evoluția principalelor ramuri economice din luna martie 2020, cadrul de prognoză în timp real a actualizat dinamica trimestrială a PIB pentru T1 2020 de la aproximativ 1,3 la sută (ultima perioadă din Graficul 4, Panel A) până la circa 0,1 la sută (Panel B). În același timp, noile informații au condus la ajustarea asimetriei distribuției prognozelor, de la una orientată către dreapta, la una care prezintă *skewness* negativ, fapt care sugerează unele riscuri de supraestimare a evoluției activității economice.

Grafic 5. Prognoze succesive pentru PIB în intervalul T2 2020 – T3 2020



Notă: În grafic este ilustrată distribuția prognozelor pentru intervalul T2 2020 – T3 2020, pe măsură ce setul de informații cu privire la evoluția activității economice este extins succesiv cu câte o observație lunară. Ultima observație trimestrială utilizată este T1 2020. Intervalul de încredere este dat de percentila de 5 la sută, respectiv de cea de 95 la sută a distribuției combinate a prognozelor.

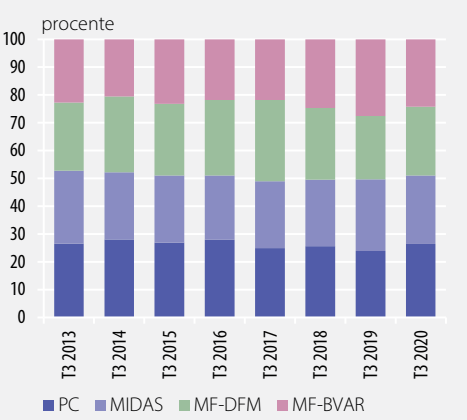
Actualizarea proiecției a generat revizuiri și mai ample pe parcursul T2 2020 (Grafic 5). Astfel, primele prognoze realizate pe baza datelor disponibile pentru luna martie 2020 sugerau o contracție economică de aproximativ -1,4 la sută. Extinderea setului informațional prin adăugarea observațiilor aferente lunii aprilie 2020 a condus la o actualizare a prognozei până la aproximativ -12,3 la sută, valoare apropiată de primele dinamici trimestriale ale PIB publicate de INS. În această etapă se remarcă, totuși, un grad ridicat de incertitudine, distribuția prognozelor conținând valori care se

plasau semnificativ sub pragul de -30 la sută (generate de modelele MF-DFM) alături de unele care sugerau mai degrabă o contracție ușoară. Bineînțeles, pe măsură ce noi observații lunare au devenit disponibile, încrederea asociată proiecțiilor (pentru ambele orizonturi) a început să se îmbunătățească. De menționat că punctul de minim al activității economice s-a consemnat în luna aprilie (BNR, 2020), datele publicate ulterior de INS indicând o redresare; tendința a fost preluată de modele (în principal de MF-DFM, care a cunoscut, de altfel, și cea mai consistentă actualizare a proiecției, până la -14,9 la sută). Proiecția agregată, care utilizează toți indicatorii lunari sugerează o contracție de circa -10 la sută pentru T2 2020, respectiv o revenire până la circa 3,4 la sută în T3 2020; primele valori oficiale pentru cele două trimestre au fost -13,1 la sută, respectiv 5,4 la sută (ritmuri aproximate prin diferență în logaritm).

5.3. ROBUSTEȚEA REZULTATELOR

Varianta preferată a cadrului de prognoză în timp real presupune combinarea prognozelor aparținând tuturor metodologiilor considerate pe baza unui sistem de ponderare care ține seama de numărul de modele asociat fiecărei categorii, dar și de eventuale modificări în timp a capacității de prognoză a acestora. Așadar, în această secțiune este investigată măsura în care modificări în oricare dintre dimensiunile menționate anterior pot conduce la o îmbunătățire a prognozelor – rezultatele acestor evaluări sunt prezentate în Tabelul 8. Astfel, neajustarea sistemului de ponderare pentru a ține seama de numărul de modele specificat pentru fiecare metodă determină o înrăutățire a acurateței prognozei. Această din urmă evoluție se datorează acordării unei ponderi mai însemnate metodologiei MIDAS, în condițiile în care, pentru explorarea întregului set de indicatori, fără a recurge la utilizarea altor tipuri de proceduri pentru sintetizarea informațiilor, a fost necesară includerea unui număr ridicat de modele – 36 în procedura finală de agregare, comparativ cu 3 pentru PC și MF-DFM, respectiv 5 pentru MF-BVAR. Modificarea sistemului de ponderare în sensul utilizării unei perioade mai scurte de evaluare a performanței modelelor (3 perioade, față de 5 perioade în varianta preferată pentru *nowcasting*) nu aduce schimbări importante din punct de vedere al acurateței prognozei. O majorare a valorii funcțiilor de *scoring* se remarcă în situația în care se consideră că fiecare metodologie prezintă o acuratețe istorică de prognoză similară (ceea ce se traduce în ponderi egale) sau când ponderile țin seama de performanța modelelor individuale pe termen lung, adăugând succesiv câte o perioadă (cea mai recentă dintre ele este evaluată pe baza performanței

Grafic 6. Evoluția ponderilor de agregare



Notă: În grafic sunt ilustrate ponderile pentru fiecare metodologie utilizată în cadrul procesului de agregare a densităților de repartiție.

medii a modelului în ultimele 30 de trimestre). Acest din urmă aspect sugerează că pe parcursul perioadei de evaluare puterea de predicție a modelelor a suferit modificări – fapt oarecum previzibil, dat fiind că economia autohtonă este supusă unor multiple procese transformaționale, necesare pentru a recupera decalajul față de economiile avansate. Evoluția ponderilor de agregare pe parcursul perioadei de evaluare este ilustrată în Graficul 6.

Tabel 8. Robustețea rezultatelor

Orizont de prognoză	T+1				T+2
	M0	M1	M2	M3	
Cadrul de referință	0,468	0,439	0,417	0,415	0,475
Model AR(1) cu *drift*	0,507	0,507	0,507	0,507	0,571
A. Scheme de ponderare alternative					
Ponderi neajustate	0,478	0,438	0,433	0,429	0,481
Ponderi medie trei perioade	0,468	0,441	0,417	0,415	0,473
Ponderi medii pe TL	0,469	0,442	0,417	0,417	0,474
Ponderi egale	0,470	0,442	0,421	0,419	0,476
B. Metodologii excluse din combinare					
PC	0,468	0,437	0,419	0,417	0,474
MIDAS	0,468	0,443	0,413	0,413	0,475
MF-DFM	0,474	0,435	0,418	0,417	0,481
MF-BVAR	0,469	0,449	0,422	0,419	0,474
C. Măsuri alternative pentru PIB					
PIB excluzând VAB Agricultură	0,344	0,367	0,342	0,334	0,338
Vintage PIB, prov II, T3 2020	0,484	0,465	0,448	0,431	0,481

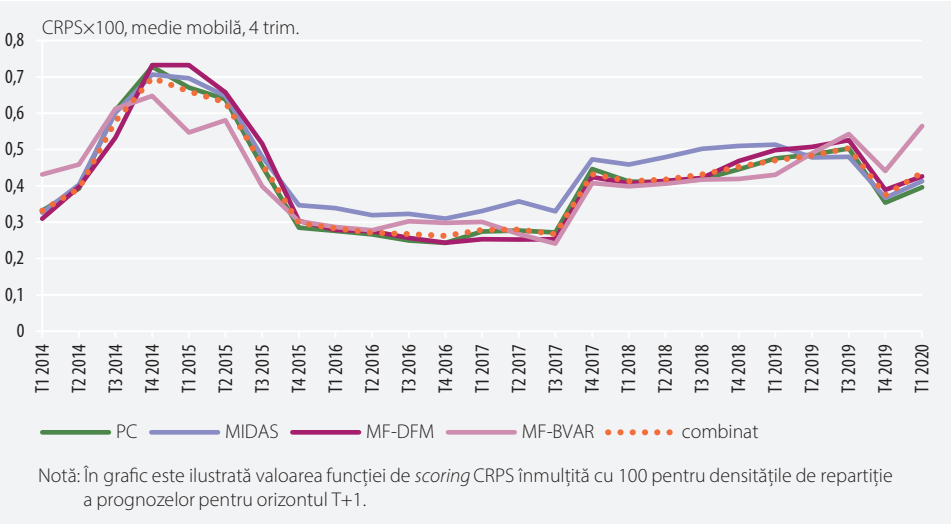
Notă: În tabel este prezentată valoarea funcției de *scoring* CRPS multiplicată cu 100 și care reprezintă o generalizare a unei matrici simple, precum eroare medie de prognoză la cazul mai sofisticat al prognozelor de probabilități. „Cadrul de referință" presupune combinarea tuturor modelelor utilizând ponderi variabile în timp, calculate pe baza a cinci perioade consecutive pentru T+1, respectiv patru trimestre pentru T+2, ajustate cu numărul de modele pentru fiecare metodologie. Prognozele se bazează pe versiunea PIB provizoriu I disponibilă în momentul realizării proiecției. „Ponderile neajustate" implică utilizarea unui sistem de ponderare care nu ține cont de numărul modelelor utilizate la nivelul fiecărei metodologii. În cazul poziției „Ponderi medii pe TL" se ține seama de întreaga perioadă de timp pe baza căreia sunt evaluate modelele, adăugând succesiv câte o observație. "Ponderi egale" presupun că fiecare metodologie contribuie cu 25 la sută în procesul de combinare a prognozelor.

Comparând metodologiile considerate în cadrul de prognoză din punct de vedere al funcției de *scoring*, o ușoară îmbunătățire se remarcă doar în cazul în care este exclusă metodologia MIDAS și numai pentru situația în care sunt utilizate două, respectiv trei observații lunare. Modelele de acest tip prezintă o sensibilitate mai ridicată la dimensiunea eșantionului de date (ele fiind estimate, de altfel, pe baza unor tehnici neliniare), capacitatea de prognoză a acestora ameliorându-se vizibil în a doua parte a intervalului de evaluare, odată cu majorarea numărului de observații (Grafic 7). Similar, excluderea modelelor care utilizează factori comuni îmbunătățește capacitatea de prognoză doar în situația în care se utilizează o singură observație lunară.

După cum a s-a menționat anterior, în absența unor serii de date care să ajute modelele să îi aproximeze evoluția într-o manieră robustă, agricultura reprezintă o sursă importantă a erorii de specificare. Chiar dacă sectorul are o contribuție limitată la formarea PIB (circa 4 la sută în 2019), variațiile ample ale producției, induse, în mare parte, de fenomene meteorologice extreme, tind să imprime o volatilitate ridicată creșterii trimestriale a PIB. Prin urmare, excluderea sectorului permite modelelor să

anticipeze mult mai precis creşterea economică, valoarea medie a funcţiei de *scoring* reducându-se cu aproximativ 20 la sută. Totodată, utilizarea celei mai recente versiuni a PIB conduce la o înrăutăţire cu aproximativ 5 la sută a funcţiei de *scoring* – această versiune a datelor de conturi naţionale prezintă o volatilitate mai înaltă a ratelor trimestriale de creştere pe parcursul perioadei de evaluare, fapt care afectează mai ales capacitatea de prognoză a modelelor MF-BVAR.

Grafic 7. Valoarea funcţiei de *scoring* (CRPS) pentru orizontul T+1



Notă: În grafic este ilustrată valoarea funcţiei de *scoring* CRPS înmulţită cu 100 pentru densităţile de repartiţie a prognozelor pentru orizontul T+1.

Dincolo de acest aspect, literatura de specialitate furnizează suficiente indicii cu privire la faptul că modelele econometrice tind să performeze mai bine atunci când sunt considerate primele versiuni de date publicate (Faust *et al.*, 2003; Koenig *et al.*, 2003; Clements şi Galvão, 2009), includerea de către institutele de statistică a unor noi surse de date contribuind la deteriorarea legăturilor dintre indicatorii lunari şi creşterea economică. Prin urmare, pentru a asigura o performanţă similară în exerciţiile reale de prognoză, este de preferat utilizarea primelor versiuni de date PIB pentru selecţia modelelor de prognoză. În contrast cu unele studii realizate pentru economii central şi est-europene (Rünstler *et al.*, 2009; Feldkircher *et al.*, 2015), în care metodologiile mai complexe aveau dificultăţi notabile de a depăşi performanţa celor autoregresive, cadrul de prognoză prezentat în această lucrare prezintă (pentru ambele orizonturi) o acurateţe a prognozelor simţitor mai ridicată (în medie cu 15 la sută).

Concluzii şi direcţii viitoare de cercetare

Această lucrare propune un cadru de monitorizare în timp real a activităţii economice care îmbină multiple metodologii capabile să utilizeze eficient fluxul de indicatori diseminaţi de institutele de statistică (sau de alte organisme), ulterior publicării

datelor de conturi naționale. Modelele utilizate acoperă o paletă largă: (i) modele punte, implementate cu ajutorul factorilor comuni, (ii) modele MIDAS restricționate, (iii) modele cu factori comuni dinamici, care includ variabile cu frecvențe mixte și (iv) modele VAR Bayesiene, care utilizează, de asemenea, serii de timp cu frecvențe mixte, dar și erori variabile în timp. Proiecțiile generate de fiecare dintre acestea sunt apoi combinate, având ca punct de reper acuratețea istorică. Strategia de agregare a proiecțiilor prezintă multiple avantaje, printre cele mai importante fiind reziliența la unele modificări structurale produse la nivelul economiei, precum și performanța satisfăcătoare în condițiile unui număr redus de observații. Totodată, în contrast cu abordarea uzuală, în lucrarea de față este preferată utilizarea întregii distribuții a prognozelor, care permite o evaluare adecvată a incertitudinii asociate traiectoriilor variabilelor de interes.

Procesul de combinare a densităților de repartiție oferă rezultate solide din punctul de vedere al acurateței proiecției, fiind consemnată doar o singură situație în care modelele individuale au avut o performanță superioară, semnificativă din punct de vedere statistic – este vorba de orizontul T+2, pentru care metodologia MF-DFM a prezentat o capacitate mai bună de a surprinde evoluții neașteptate ale creșterii economice, din punct de vedere al mediei istorice. În această zonă, o posibilă soluție pentru îmbunătățirea performanței poate fi reprezentată de migrarea către modele care folosesc exclusiv indicatori cu frecvență trimestrială, traiectoria creșterii economice putând fi astfel condiționată de cea prognozată pentru primul orizont (a se vedea Bańbura *et al.*, 2015).

În cazul *nowcasting* (trimestrul T+1), evaluarea istorică a performanței de prognoză (începând cu T1 2013) relevă doar două instanțe în care valoarea efectivă a creșterii economice se plasează în cozile distribuției prognozelor (T2 2014, respectiv T3 2017), în ambele situații dinamica trimestrială a PIB fiind exacerbată de evoluția agriculturii; din cauza disponibilității reduse a informațiilor, acest sector are o reprezentare deficitară la nivelul setului de indicatori incluși în baza de date, excluderea sa conducând la o îmbunătățire considerabilă a acurateței prognozei (în medie, cu 20 la sută). Unul dintre momentele în care metodologia de monitorizare în timp real a prezentat o utilitate crescută este dat de episodul de contracție economică cauzat de criza sanitară, publicarea primelor date cu frecvență lunară, care includeau efectele crizei pandemice asupra sectoarelor economice, conducând la o actualizare a proiecției pentru T2 2020 până la valori apropiate de cele publicate ulterior de INS.

Totuși, și la nivelul indicatorilor lunari există o anumită întârziere de publicare (de pildă, dintre indicatorii sectoriali, cifra de afaceri din comerțul cu amănuntul este publicată cel mai rapid, după aproximativ 35 de zile de la încheierea lunii de referință), interval de timp în care eventuale șocuri produc efecte asupra creșterii economice. În acest sens, avansul tehnologiei, concomitent cu intensificarea procesului de digitalizare a economiei, a condus la lărgirea setului de date care poate fi utilizat în exerciții de prognoză, (denumite generic *Big Data*). Acestea includ, de pildă, date care surprind vânzările operatorilor comerciali în timp real (*scanner data)*, căutări de produse pe principalele motoare de căutare *online* (*Google Trends*), rapoarte de mobilitate, date obținute din monitorizarea principalelor surse de știri *online*, etc. (v. Varian, 2014; Bok *et al.*, 2018; Buono *et al.*, 2018). Pe lângă obținerea unor evaluări

mai rapide referitoare la creșterea economică, aceste seturi de date (care adesea sunt fructificate prin tehnici de *machine learning,* Chakraborty și Joseph, 2017; Makridakis *et al.*, 2018; Doerr *et al.*, 2021) pot aduce îmbunătățiri notabile și în ceea ce privește evaluarea traiectoriei sectorului agricol, unde aprecierile operatorilor economici din sector pot fi preluate, de pildă, prin tehnici de monitorizare a fluxului de știri.

Anexă

În Anexă sunt prezenți indicatorii lunari folosiți în cadrul exercițiilor de prognoză. Majoritatea variabilelor sunt furnizate de INS (direct sau prin intermediul Eurostat) în termeni reali și ajustate sezonier, existând totuși unele situații în care aceste transformări nu sunt disponibile (marcate în Tabelul A1). Prin urmare, pentru ajustarea sezonieră a seriilor de timp s-a apelat la metoda *Tramo-Seats*, implementată în cadrul programului software *JDemetra+*[1], procesul ținând totodată seama și de efectele de calendar, acolo unde era necesar. În ceea ce privește procesul de deflatare a indicatorilor diseminați în termeni nominali, în situația în care a fost posibil s-a realizat o mapare între indicele de valoare și cel de preț în funcție de clasificarea CAEN Rev. 2; în cazul salariilor s-a optat pentru utilizarea indicelui prețurilor de consum (IPC). Indicii de preț aferenți serviciilor de piață prestate întreprinderilor au fost estimați utilizând tehnici econometrice pe baza celor disponibili la nivel trimestrial, precum și a remunerării salariaților pentru fiecare sector de activitate, componente ale IPC servicii și cursul de schimb nominal (procedura este descrisă în Secțiunea 3).

Tabel A1. Lista indicatorilor lunari incluși în baza de date

Nr.	Secțiune CAEN	Denumire indicator	Incluși în MF-BVAR	Prelucrări adiționale		Sursa
				ajustate sezonier	deflator	
1	A	Câștigul real net din agricultură		✓	IPC	INS
2		Numărul de salariați din agricultură		✓		INS
3	C-E	Cifra de afaceri din industria prelucrătoare	✓		IPPI	Eurostat
4		– bunuri de consum (MGI)			IPPI	Eurostat
5		– bunuri intermediare (MGI)			IPPI	Eurostat
6		– bunuri de capital (MGI)			IPPI	Eurostat
7		Producția industrială	✓			INS
8		– industria extractivă				INS
9		– industria prelucrătoare	✓			INS
10		- bunuri de capital (MGI)	✓			INS
11		- bunuri de folosință îndelungată (MGI)	✓			INS
12		- bunuri de uz curent (MGI)	✓			INS
13		- bunuri intermediare (MGI)	✓			INS
14		– industria energetică (MGI)				INS
15		Comenzi noi ind. prelucrătoare	✓	✓	IPPI	INS
16		– bunuri de consum (MGI)		✓	IPPI	INS
17		– bunuri intermediare (MGI)		✓	IPPI	INS
18		– bunuri de capital (MGI)		✓	IPPI	INS
19		Numărul de ore lucrate în ind. prelucrătoare	✓			INS
20		Câștigul real net, ind. prelucrătoare		✓	IPC	INS
21		Numărul de salariați, ind. prelucrătoare		✓		INS

[1] Mai multe detalii sunt disponibile accesând următorul link: https://ec.europa.eu/eurostat/cros/content/software-jdemetra_en

– continuare –

Nr.	Secțiune CAEN	Denumire indicator	Incluși în MF-BVAR	Prelucrări adiționale		Sursa
				ajustate sezonier	deflator	
22	F	Lucrări de construcții	✓			Eurostat
23		– clădiri rezidențiale				Eurostat
24		– lucrări de inginerie civilă				Eurostat
25		Autorizații de construire, clădiri	✓			Eurostat
26		Câștigul real net din construcții		✓		INS
27		Numărul de salariați din construcții		✓		INS
28	G, I, R	Comerț cu ridicata (excl. auto moto)	✓		IPC*	Eurostat
29		– bunuri de consum			IPC*	Eurostat
30		Comerț cu amănuntul (excl. auto, moto)	✓			Eurostat
31		Comerț excl. auto, moto, carburanți				Eurostat
32		Comerț cu carburanți				Eurostat
33		Comerț auto, moto și piese de schimb	✓			INS
34		– comerț auto				INS
35		Servicii prestate populației	✓			INS
36		– hoteluri și restaurante	✓			INS
37		– jocuri de noroc și alte activ. recreative				INS
38		– agenții de turism				INS
39		Câștigul real net din comerț		✓	IPC	INS
40		Numărul de salariați din comerț		✓		INS
41	H	Transport și depozitare	✓		IPSI	Eurostat
42		– terestru și prin conducte			IPSI	Eurostat
43		– depozitare și suport			IPSI	Eurostat
44		– activități poștale și de curierat			IPSI	Eurostat
45		Câștigul real net din transport		✓		INS
46		Numărul de salariați din transport		✓		INS
47	K	Credite acordate populației	✓	✓	IPC	BNR
48		Credite pentru achiz. de echip. și fin. stoc.	✓	✓	IPPI*	BNR, CRC
49		Credite pentru investiții imobiliare	✓	✓	ICC	BNR, CRC
50		Câștigul real net interm. financiare		✓	IPC	INS
51		Numărul de salariați interm. financiare		✓		INS
52	J	Informații și comunicații	✓		IPSI	Eurostat
53		– telecomunicații			IPSI	Eurostat
54		– servicii în tehnologia informației			IPSI	Eurostat
55		– servicii informatice			IPSI	Eurostat
56		Câștigul real net informații și comunicare		✓	IPC	INS
57		Numărul de salariați informații și comunicare		✓		INS

– continuare –

Nr.	Secțiune CAEN	Denumire indicator	Incluși în MF-BVAR	Prelucrări adiționale		Sursa
				ajustate sezonier	deflator	
58	L-N	Activități profesionale, științifice și tehnice	✓		IPSI	Eurostat
59		– consult. juridică, contab. și manag.			IPSI	Eurostat
60		– arhitectură și inginerie			IPSI	Eurostat
61		Activități administrative și servicii suport	✓		IPSI	Eurostat
62		– plasarea forței de muncă			IPSI	Eurostat
63		– investigație și protecție			IPSI	Eurostat
64		– activ. de curățenie și peisagistice			IPSI	Eurostat
65		– secretariat, servicii suport			IPSI	Eurostat
66		Câștigul real net tranzacții imob.		✓	IPC	INS
67		Câștigul real net activ. profesionale		✓	IPC	INS
68		Câștigul real net serv. administrative		✓	IPC	INS
69		Numărul de salariați tranzacții imob.		✓		INS
70		Numărul de salariați activ. profesionale		✓		INS
71		Numărul de salariați serv. administrative		✓		INS
72	O-T	Câștigul real net apărare		✓	IPC	INS
73		Câștigul real net învățământ		✓	IPC	INS
74		Câștigul real net sănătate și asist. soc.		✓	IPC	INS
75		Câștigul real net alte servicii		✓	IPC	INS
76		Numărul de salariați apărare		✓		INS
77		Numărul de salariați învățământ		✓		INS
78		Numărul de salariați sănătate și asist. soc.		✓		INS
79		Numărul de salariați alte servicii		✓		INS
80	Macro	Importul de bunuri				Eurostat
81		Exportul de bunuri				Eurostat
82		Câștigul real net, mediu economie	✓	✓	IPC	INS
83		Numărul de salariați, mediu economie	✓	✓		INS
84		Cursul de schimb real efectiv – IPC				BIS
85		Cursul de schimb real efectiv – IPPI				ECB
86		Cursul de schimb nominal efectiv				BIS
87		Indicele activității bursiere (BET)	✓			BVB
88		Servicii prestate întreprinderilor	✓		IPSI	Eurostat
89		Indicatorul de încredere în economie	✓			CE
90		ROBOR 3M	✓			BNR
91		Rata șomajului BIM	✓			INS
92		Indicatorul de încredere din industrie	✓			CE
93		– așteptări privind producția				CE
94		Indicatorul de încredere din comerț	✓			CE
95		Indicatorul de încredere din construcții				CE

IPC – indicele prețurilor de consum; IPPI – indice de preț producție industrială, IPSI – indice de preț servicii prestate întreprinderilor; MGI – Mari Grupe Industriale; IPC* – indicele prețurilor de consum exclusiv modificările cotei TVA; IPPI* – indicele prețurilor producției industriale pentru bunuri de capital, livrate pe plan local.

Notă: În unele situații seriile de tip au fost ajustate sezonier și exprimate în termeni reali utilizând indicii de preț aferenți fiecărei categorii de produse.

Bibliografie

Aastveit, K. A., Foroni, C., și Ravazzolo, F. „Density Forecasts With Midas Models", *Journal of Applied Econometrics*, Vol. 32(4), 2017, pp. 783-801

Aastveit, K. A., Mitchell, J., Ravazzolo, F., van Dijk, H. K. „The Evolution of Forecast Density Combinations in Economics", în *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press, 2019

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., Veronese, G. „New Eurocoin: Tracking Economic Growth in Real Time", *Review of Economics and Statistics*, Vol. 92(4), 2010, pp. 1024-1034

Amisano, G., Giacomini, R. „Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business & Economic Statistics*, Vol. 25(2), 2007, pp. 177-190

Andreou, E., Ghysels, E., Kourtellos, A. „Forecasting with Mixed-Frequency Data", *The Oxford Handbook of Economic Forecasting*, (April 2018), 2012, pp. 1-24

Angelini, E., Henry, J., Marcellino, M. „Interpolation and Backdating with a Large Information Set", *Journal of Economic Dynamics and Control*, Vol. 30(12), 2006, pp. 2693-2724

Ankargren, S., Jonéus, P. „Simulation Smoothing for Nowcasting with Large Mixed-frequency VARs", *Econometrics and Statistics*, 2020

Armeanu, D., Andrei, J. V., Lache, L., Panait, M. „A Multifactor Approach to Forecasting Romanian Gross Domestic Product (GDP) in the Short Run", *PLOS ONE*, Ediție de W.-X. Zhou, Vol. 12(7), 2017

Armesto, M. T., Engemann, K. M., Owyang, M. T. *Forecasting with Mixed Frequencies*, Review, Federal Reserve Bank of St. Louis, Vol. 92(Nov), 2010, pp. 521-536

Baffigi, A., Golinelli, R., Parigi, G. „Bridge Models to Forecast the Euro area GDP", *International Journal of Forecasting*, Vol. 20(3), 2004, pp. 447-460

Bai, J., Ghysels, E., Wright, J. H. „State Space Models and MIDAS Regressions", *Econometric Reviews*, Vol. 32(7), 2013, pp. 779-813

Baltagi, B. H., Kao, C., Wang, F. „Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes", *Journal of Econometrics*, Vol. 220(2), 2021, pp. 349-365

Bańbura, M., Giannone, D., Lenza, M. „Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-sections", *International Journal of Forecasting*, Vol. 31(3), 2015, pp. 739-756

Bańbura, M., Giannone, D., Modugno, M., Reichlin, L. „Now-Casting and the Real-Time Data Flow", în *Handbook of Economic Forecasting*, 2013, pp. 195-237

Bańbura, M., Giannone, D., Reichlin, L. „Large Bayesian Vector Auto Regressions", *Journal of Applied Econometrics*, Vol. 25(1), 2010, pp. 71-92

Bańbura, M., Rünstler, G. „A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting GDP", *International Journal of Forecasting*, Vol. 27(2), 2011, pp. 333-346

Bańbura, M., Rünstler, G., Angelini, E. „Estimating and Forecasting the Euro Area Monthly National Accounts from a Dynamic Factor Model", OECD Journal: *Journal of Business Cycle Measurement and Analysis*, Vol. 2010(1), 2010, pp. 1-22

Bates, B. J., Plagborg-Møller, M., Stock, J. H., Watson, M. W. „Consistent Factor Estimation in Dynamic Factor Models with Structural Instability", *Journal of Econometrics*, Vol. 177(2), 2013, pp. 289-304

Bencivelli, L., Marcellino, M. G., Moretti, G. „Selecting Predictors by Using Bayesian Model Averaging in Bridge Models", Bank of Italy, Temi di discussione, Nr. 872, 2012

Berkowitz, J. „Testing Density Forecasts, With Applications to Risk Management", Journal of Business & Economic Statistics, Vol. 19(4), 2001, pp. 465-474

BNR „Impactul economic al primelor măsuri adoptate în scopul aplatizării curbei epidemice", în *Raport asupra inflației*, noiembrie 2020

Boero, G., Smith, J., Wallis, K. F. „Scoring Rules and Survey Density Forecasts", *International Journal of Forecasting*, Vol. 27(2), 2011, pp. 379-393

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., Tambalotti, A. „Macroeconomic Nowcasting and Forecasting with Big Data", *Annual Review of Economics*, Vol. 10(1), 2018, pp. 615-643

Brave, S. A., Butters, R. A., Justiniano, A. „Forecasting Economic Activity with Mixed Frequency BVARs", *International Journal of Forecasting*, Vol. 35(4), 2019, pp. 1692-1707

Buono, D., Kapetanios, G., Marcellino, M., Mazzi, G. L., Papailias, F.	„Big Data Econometrics: Now Casting and Early Estimates", BAFFI CAREFIN Centre Research Paper, Nr. 2018-82, 2018
Camacho, M.	„Short-term Forecasting for Empirical Economists: A Survey of the Recently Proposed Algorithms", Foundations and Trends® in *Econometrics*, Vol. 6(2), 2013, pp. 101-161
Carriero, A., Clark, T. E., Marcellino, M.	„Common Drifting Volatility in Large Bayesian VARs", *Journal of Business & Economic Statistics*, Vol. 34(3), 2016, pp. 375-390
Chakraborty, C., Joseph, A.	„Machine Learning at Central Banks", Bank of England Working Papers, Nr. 674, 2017
Clark, T. E., McCracken, M. W.	„Averaging Forecasts from VARs with Uncertain Instabilities", *Journal of Applied Econometrics*, Vol. 25(1), 2010, pp. 5-29
Clements, M. P., Galvão, A. B.	„Forecasting US Output Growth using Leading Indicators: An Appraisal using MIDAS Models", *Journal of Applied Econometrics*, Vol. 24(7), 2009, pp. 1187-1206
Clements, M. P., Hendry, D. F.	„Intercept Corrections and Structural Change", *Journal of Applied Econometrics*, Vol. 11(5), 1996, pp. 475-494
Cogley, T., Sargent, T. J.	„Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US", *Review of Economic Dynamics*, Vol. 8(2), 2005, pp. 262-302
Diebold, F. X., Gunther, T. A., Tay, A. S.	„Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review*, Vol. 39(4), 1998, p. 863
Diebold, F. X., Mariano, R. S.	„Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, Vol. 20(1), 2002, pp. 134-144.
Diron, M.	„Short-term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-time Performance Based on Vintage Data", *Journal of Forecasting*, Vol. 27(5), 2008, pp. 371-390
Doerr, S., Gambacorta, L., Garralda, J. M. S.	„Big Data and Machine Learning in Central Banking", BIS Working Papers, No. 930, 2021
Doz, C., Giannone, D., Reichlin, L.	„A Two-step Estimator for Large Approximate Dynamic Factor Models based on Kalman Filtering", *Journal of Econometrics*, Vol. 164(1), 2011, pp. 188-205
Durbin, J., Koopman, S. J.	Time Series Analysis by State Space Methods. 2-lea ed, Oxford University Press, 2012

Elliott, G., Komunjer, I., Timmermann, A. — „Estimation and Testing of Forecast Rationality under Flexible Loss", *Review of Economic Studies*, Vol. 72(4), 2005, pp. 1107-1125

Elliott, G., Timmermann, A. — „Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions", *Journal of Econometrics*, Vol. 122(1), 2004, pp. 47-79

Economic Forecasting, Princeton University Press, 2016

Faust, J., Rogers, J. H., Wright, J. H. — „Exchange Rate Forecasting: The Errors We've Really Made", *Journal of International Economics*, Vol. 60(1), 2003, pp. 35-59

„News and Noise in G-7 GDP Announcements", *Journal of Money, Credit, and Banking*, Vol. 37(3), 2005, pp. 403-419

Feldkircher, M., Huber, F., Schreiner, J., Tirpák, M., Tóth, P., Wörz, J. — *Bridging the Information Gap: Small-scale Nowcasting Models of GDP Growth for Selected CESEE Countries*, *Focus on European Economic Integration*, Oesterreichische Nationalbank (Austrian Central Bank), (2), 2015, pp. 56-75

Fernández, C., Ley, E., Steel, M. F. J. — „Benchmark Priors for Bayesian Model Averaging", *Journal of Econometrics*, Vol. 100(2), 2001, pp. 381-427

Forni, M., Giannone, D., Lippi, M., Reichlin, L. — *Opening the Black Box: Structural Factor Models with Large Cross Sections*, *Econometric Theory*, Vol. 25(5), 2009, pp. 1319-1347

Foroni, C., Marcellino, M. G. — „A Survey of Econometric Methods for Mixed-Frequency Data", *SSRN Electronic Journal*, 2013

Gălățescu, A. A., Labhard, V. — „Predicting the 'Great' Recession and the 'Poor' Recovery – Results from a Suite of Non-linear Models for Romania and the Euro Area", National Bank of Romania, *Occasional Papers* Nr. 28, 2019

Garratt, A., Lee, K., Pesaran, M. H., Shin, Y. — „Forecast Uncertainties in Macroeconomic Modeling", *Journal of the American Statistical Association*, Vol. 98(464), 2003, pp. 829-838

Geweke, J. — *Contemporary Bayesian Econometrics and Statistics*, Hoboken, NJ, USA: John Wiley & Sons, Inc. (Wiley Series in Probability and Statistics), 2005

Geweke, J., Amisano, G. — „Optimal Prediction Pools", *Journal of Econometrics*, Vol. 164(1), 2011, pp. 130-141

Ghysels, E. — „Macroeconomics and the Reality of Mixed Frequency Data", *Journal of Econometrics*, Vol. 193(2), 2016, pp. 294-314

Ghysels, E., Santa-Clara, P., Valkanov, R. — „The MIDAS Touch: Mixed Data Sampling Regression Models", CIRANO Working Papers, Nr. 2004s-20, 2004

„There is a Risk-return Trade-off after All", *Journal of Financial Economics*, Vol. 76(3), 2005, pp. 509-548

Ghysels, E., Sinko, A., Valkanov, R. — „MIDAS Regressions: Further Results and New Directions", *Econometric Reviews*, Vol. 26(1), 2007, pp. 53-90

Ghysels, E., Valkanov, R. I., Rubia, A. — „Multi-Period Forecasts of Volatility: Direct, Iterated, and Mixed-Data Approaches", *SSRN Electronic Journal*, 2009

Giacomini, R. — „Comparing Density Forecasts via Weighted Likelihood Ratio Tests: Asymptotic and Bootstrap Methods", *SSRN Electronic Journal*, 2002

Giannone, D., Reichlin, L., Small, D. — „Nowcasting: The Real-time Informational Content of Macroeconomic Data", *Journal of Monetary Economics*, Vol. 55(4), 2008, pp. 665-676

Gneiting, T., Raftery, A. E. — „Strictly Proper Scoring Rules, Prediction, and Estimation", *Journal of the American Statistical Association*, Vol. 102(477), 2007, pp. 359-378

Golinelli, R., Parigi, G. — „The Use of Monthly Indicators to Forecast Quarterly GDP in the Short Run: An Application to the G7 Countries", *Journal of Forecasting*, Vol. 26(2), 2007, pp. 77-94

Granger, C. W. J. — „Outline of Forecast Theory using Generalized Cost Functions", *Spanish Economic Review*, Vol. 1(2), 1999, p. 161

Granger, C. W. J., Ramanathan, R. — „Improved Methods of Combining Forecasts", *Journal of Forecasting*, Vol. 3(2), 1984, pp. 197-204

Hall, S. G. Mitchell, J. — „Combining Density Forecasts", International Journal of Forecasting, Vol. 23(1), 2007, pp. 1-13

Hamilton, J. D. — *Time Series Analysis*, Princeton University Press, 1994

Havrlant, D., Tóth, P., Wörz, J. — *On the Optimal Number of Indicators – Nowcasting GDP Growth in CESEE, Focus on European Economic Integration*, Oesterreichische Nationalbank (Austrian Central Bank), (4), 2016, pp. 54-72

Hendry, D. F., Clements, M. P. — „Economic Forecasting: Some Lessons from Recent Research", *Economic Modelling*, Vol. 20(2), 2003, pp. 301-329

Hendry, D. F., Clements, M. P. — „Pooling of Forecasts", *The Econometrics Journal*, Vol. 7(1), 2004, pp. 1-31

Jore, A. S., Mitchell, J., Vahey, S. P. „Combining Forecast Densities from VARs with Uncertain Instabilities", *Journal of Applied Econometrics*, Vol. 25(4), 2010, pp. 621-634

Karlsson, S. „Forecasting with Bayesian Vector Autoregression", în Elliott, G. și Timmermann, A. (ed.) *Handbook of Economic Forecasting*, Vol. 2, Part B, 2013, pp. 791-897

Klein, L. R., Sojo, E. „Combinations of High and Low Frequency Data in Macroeconometric Models", în Klein, L. R. și Marquez, J. (ed.) *Economics in Theory and Practice: An Eclectic Approach. Advanced Studies in Theoretical and Applied Econometrics*, vol 17, 1989, pp. 3-16

Koenig, E. F., Dolmas, S., Piger, J. „The Use and Abuse of Real-time Data in Economic Forecasting", *Review of Economics and Statistics*, 2003

Koop, G. *Bayesian Econometrics*, John Wiley & Sons, 2003

Kuzin, V., Marcellino, M., Schumacher, C. „MIDAS vs. Mixed-frequency VAR: Nowcasting GDP in the Euro Area", *International Journal of Forecasting*, Vol. 27(2), 2011, pp. 529-542

Larsson, R., Clements, M. P., Hendry, D. F. „Forecasting with Difference-stationary and Trend-stationary Models", *The Econometrics Journal*, Vol. 4(1), 2001, pp. S1-S19

León, Á., Nave, J. M., Rubio, G. „The Relationship between Risk and Expected Return in Europe", *Journal of Banking & Finance*, Vol. 31(2), 2007, pp. 495-512

Litterman, R. B. „A Statistical Approach to Economic Forecasting", *Journal of Business & Economic Statistics*, Vol. 4(1), 1986, pp. 1-4

Lutkepohl, H. „Approximation of Arbitrary Distributed Lag Structures by a Modified Polynomial Lag: An Extension", *Journal of the American Statistical Association*, Vol. 75(370), 1980, p. 428

Makridakis, S., Spiliotis, E., Assimakopoulos, V. *Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward*, PLOS ONE, Ediție de A. R. Hernandez Montoya, Vol. 13(3), 2018, p. e0194889

Marcellino, M., Schumacher, C. „Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP", *Oxford Bulletin of Economics and Statistics*, Vol. 72(4), 2010, pp. 518-550

Mariano, R. S., Murasawa, Y. „A New Coincident Index of Business Cycles based on Monthly and Quarterly Series", *Journal of Applied Econometrics*, Vol. 18(4), 2003, pp. 427-443

Matheson, T. D. „New Indicators for Tracking Growth in Real Time", OECD Journal: *Journal of Business Cycle Measurement and Analysis*, Vol. 2013(2), 2014, pp. 51-71

Nalban, V.	„Modelarea incertitudinii asociate datelor statistice în contextul revizuirilor istorice", Banca Națională a României, *Caiete de studii* Nr. 33, 2014
Pesaran, M. H., Timmermann, A.	„Selection of Estimation Window in the Presence of Breaks", *Journal of Econometrics*, Vol. 137(1), 2007, pp. 134-161
Riedl, A., Wörz, J.	*A Simple Approach to Nowcasting GDP Growth in CESEE Economies*, Oesterreichische Nationalbank (Austrian Central Bank), (Q4/18), 2018, pp. 56-74
Rossi, B.	*Advances in Forecasting under Instability*, 2013, pp. 1203-1324
Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Van Nieuwenhuyze, C.	„Short-term Forecasting of GDP using Large Datasets: A Pseudo Real-time Forecast Evaluation exercise", *Journal of Forecasting*, Vol. 28(7), 2009, pp. 595-611
Rusnák, M.	„Nowcasting Czech GDP in Real Time", *Economic Modelling*, Vol. 54, 2016, pp. 26-39
Schorfheide, F., Song, D.	„Real-Time Forecasting With a Mixed-Frequency VAR", *Journal of Business & Economic Statistics*, Vol. 33(3), 2015, pp. 366-380
Schumacher, C.	„A Comparison of MIDAS and Bridge Equations", *International Journal of Forecasting*, Vol. 32(2), 2016, pp. 257-270
Stock, J. H., Watson, M. W.	„A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", în Engle, R. F. și White, H. (ed.) *Cointegration, Causality, and Forecasting*, Festschrift in Honour of Clive W. J. Granger, Oxford University Press, 1999, pp. 1-44
Stock, J. H., Watson, M. W.	„Forecasting Using Principal Components From a Large Number of Predictors", *Journal of the American Statistical Association*, Vol. 97(460), 2002, pp. 1167-1179
Tănase, A.	„Estimarea și prognoza dinamicii PIB și ale componentelor PIB prin modelul cu factori comuni dinamici. Aplicație pentru România", Banca Națională a României, *Caiete de studii* Nr. 34, 2014
Timmermann, A.	„Forecast Combinations", în Elliott, G., Granger, C. W. J. și Timmermann, A. (ed.) *Handbook of Economic Forecasting*, Elsevier B.V, 2006, pp. 135-196
Tóth, P.	„Nowcasting Slovak GDP by a Small Dynamic Factor Model", *EKONOMICKY CASOPIS*, Vol. 65(2), 2017, pp. 107-124

Varian, H. R. „Big Data: New Tricks for Econometrics", *Journal of Economic Perspectives*, Vol. 28(2), 2014, pp. 3-28

Wallis, K. F. „An Assessment of Bank of England and National Institute Inflation Forecast Uncertainties", *National Institute Economic Review*, Vol. 189, 2004, pp. 64-71