VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Bakalauro darbas

Agregavimo alternatyvos skirtingo dažnio duomenų regresiniuose modeliuose

Alternatives of Aggregation in Regression Models with Variables Sampled at Different Frequencies

Benediktas Bilinskas

VILNIUS 2009

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

EKONOMETRINĖS ANALIZĖS KATEDRA

Darbo vadovas doc. Virmantas Kvedaras ____________
(parašas)

Darbas apgintas 2009 m. birželio mėn. ___ d.
Gynimo posėdžio protokolo Nr. ____________________
Darbas įvertintas ________________________________

Registravimo Nr. ________________________________
2009-05-28 ______________________________________

Turinys

Anotacija

Darbe tiriami skirtingo dažnio laiko eilučių ekonometriniai modeliai. Turėdami skirtingo dažnio ekonominius duomenis ir siekdami aprašyti žemo dažnio duomenis, tokiais modeliais galime geriau nei įprastais modeliais panaudoti informaciją, slypinčią aukšto dažnio duomenų dinamikoje. Darbe aprašomi žinomi bei pristatomi tys iš jų išvesti nauji skirtingo dažnio duomenų regresiniai modeliai, kurių teorinės savybės bei su imituotais ir realiais duomenimis gauti rezultatai palyginami tarpusavyje. Imitacinės analizės dalyje palyginamas modelių prognozių bei parametrų vertinimo tikslumas. Gauti rezultatai rodo, kad bent vienas naujas modelis yra geresnis nei likusieji. Empirinėje dalyje tradiciniais bei skirtingo dažnio duomenų modeliais prognozuojamas akcijų indeksų grąžų kintamumas. Pastarųjų modelių, tarp kurių dominuoja darbe pristatomi, prognozės yra tikslesnės nei tradicinių.

Summary

A typical time series regression model involves data sampled at the same frequency. In this paper I consider various approaches how to build a regression model with variables sampled at different frequencies (hence, VDF model), where dependent and independent variables are of low and high frequencies, respectively. I describe different aggregation forms, some of which were recently presented in scientific issues, but three, to my knowledge, are new ones.

The main purpose is to get and analyse the results of various aggregation functions. There is a need to find out, if there exist the best VDF model; if the results of VDF models are better in comparison with ones of typical regression models, and if they are, then in what way. To achieve the purpose, both a Monte Carlo experiment and analysis with real financial data were carried out.

The Monte Carlo experiment was done modelling eigth different data generating aggregation functions, when both parameter and forecasting accuracy of VDF models were estimated. The results show that traditional aggregation approaches often are inappropriate for usage. None of VDF models could be named being the most accurate, on the other hand, the results of new aggregation functions are among the best.

As an empirical example, multi-period out-of-sample forecasts of return volatilities of three stock indexes are compared. The forecasts are produced by several traditional econometrical forecasting approaches, i.e. direct, iterated and scaling, as well as using VDF models. The results show that the forecasts of VDF models are generally more accurate than those of traditional models. In conclusion, the forecasts of VDF models with new aggregation functions dominate.

1 Įvadas

Tipiniai laiko eilučių regresiniai modeliai naudoja vienodo dažnio duomenis. Šiuolaikiniai ekonominiai duomenys tampa vis išsamesni ir dažnai prieinami kaip aukštesnio dažnio nei anksčiau buvo įprasta. Norėdami sudaryti regresinį modelį, turėdami žemo dažnio (metinį, ketvirtinį, mėnesinį, dieninį, minutinį) paaiškinamąjį ir aukštesnio dažnio paaiškinančiuosius (ketvirtinius, mėnesinius, dieninius, minutinius ar sekundinius) kintamuosius, įprastai pastaruosius tradiciniais būdais agreguojame laike, tai yra paverčiame tokio pačio dažnio kaip ir turimas rūpimas kintamasis. Įprastas agregavimas dažniausiai reiškia aukštesnio dažnio duomenų vidurkio ar paskutinės (pirmosios, vidurinės) reikšmės ėmimą periode. Taip gali būti prarandama dalis informacijos, slypinčios duomenų dinamikoje. Skirtingo dažnio duomenų regresinių modelių sudarymas yra būdas šią informaciją panaudoti, nustatant ryšius tarp ekonominių kintamųjų. Dėl šios priežasties reikalinga skirtingo dažnio duomenų regresinių modelių palyginamoji analizė.

Darbe nagrinėjami keletas skirtingo dažnio duomenų modelių, kurie aprašyti mano skaitytoje literatūroje ([2-5] straipsniuose). Taip pat pristatomi trys nauji modeliai, kurių kiekvienas yra tam tikras aprašytų modelių praplėtimas. Pagrindinis tikslas yra atlikti skirtingo dažnio duomenų modelių empiriškai gautų rezultatų analizę. Reikalinga išsiaiškinti, ar galima išskirti vieną ar kelis geriausius modelius; kokioms sąlygoms esant konkretus modelis yra tinkamas arba netinkamas naudoti; kiek ir ar nagrinėjamų skirtingo dažnio duomenų modelių konkretaus tyrimo rezultatai yra geresni nei ekonometrijoje tradiciniais tapusių modelių.

Minėtiems uždaviniams spręsti, naudojausi dviem metodais. Pirmuoju atveju, modeliavimui naudojami generuoti skirtingo dažnio duomenys, o modeliai palyginami pagal duomenis generuojančių procesų parametrų ir prognozių tikslumą. Tradiciniais šioje dalyje laikomi įprasto agregavimo modeliai. Gauti rezultatai rodo, kad nagrinėjami modeliai stipriai pranoksta tradicinius, tačiau vieno geriausiojo išskirti negalima. Minėtų naujų modelių rezultatai yra tarp geriausiųjų. Antruoju atveju modeliuojami konkretūs finansinių indeksų grąžų duomenys, o modeliai palyginami pagal grąžų kintamumo prognozių tikslumą. Gauta, kad geresnieji skirtingo dažnio duomenų modeliai prognozuoja tiksliau nei tradiciniai prognozavimo metodai. Dviejų naujų modelių rezultatai, galima sakyti, pranoksta jau literatūroje nagrinėtų.

Toliau darbas išdėstytas tokiu būdu. Antrame skyriuje aprašomi naudojami žymėjimai bei skirtingo dažnio duomenų regresiniai modeliai. Trečiame skyriuje atliekama imitacinė analizė, kurioje palyginamas nagrinėjamų modelių gebėjimas aprašyti generuotus skirtingo dažnio duomenis. Ketvirtame skyriuje palyginamas skirtingo dažnio modelių finansinių duomenų kintamumo prognozių tikslumas tarpusavyje ir su tradiciniais ekonometriniais metodais. Paskutiniame skyriuje apibendrinami pasiekti rezultatai ir padaromos išvados.

2 Skirtingo dažnio duomenų regresiniai modeliai

Šiame skyriuje aprašomi konkretūs skirtingo dažnio duomenų regresiniai modeliai (toliau – SDD modeliai). Pirmojoje skyriaus dalyje, panašiai kaip [5] straipsnyje, pateikiami literatūroje jau nagrinėti bei pristatomi iš jų išvedami modeliai. Antrojoje skyriaus dalyje užrašoma, kaip gali būti praplėsti ir sudarinėjami ateities prognozavimui skirti SDD modeliai.

2.1 Bendras duomenų agregavimas

Nagrinėkime atsitiktinį diskretaus laiko procesą $\tilde{Y} = \big(Y_t,\ t = 0, \pm 1, \dots\big)$, kurį atitinka stebėjimai $Y = \{Y_t,\ t = 1, \dots, n\}$. $\tilde{Y}$ priklauso nuo atsitiktinio proceso $\tilde{X} = \big(X_j,\ j = 0, \pm 1, \dots\big)$, kuris stebimas aukštesnio dažnio duomenimis $X = \{X_t,\ t = 1, \dots, nm\}$. Abu kintamieji pradedami stebėti tuo pačiu laiku, o laiko momentu t Y_t atitinka aukšto dažnio duomenys $X_{(t-1)m+1}, \dots, X_{tm}$. Fiksuotas skaičius m parodo, kiek aukšto dažnio stebėjimų atitinka vieną žemo dažnio stebėjimą, ir vadinamas periodo ilgiu. Mažąja raide x žymimas vektorius $x = (x_1, \dots, x_m)$.

Funkciją $g\colon \mathbb{R}^m \to \mathbb{R}^d$, kur $d \le m,\ m, d \in \mathbb{N}$, vadinsiu (m, d) agregavimo funkcija, kuri m-vektorių transformuoja į galimai žemesnės eilės d-vektorių:

$$g(x_1, \dots, x_m) = (v_1, \dots, v_d). \tag{2.1}$$

(2.1) yra bendra agregavimo taisyklė, kuri galimai sumažina vektoriaus dimensiją. Bendresniu atveju sąlyga $d \le m$ nebūtinai turi galioti, tačiau dėl paties agregavimo tikslingumo ir galimo agregavimo funkcijos tiesiškumo kitas variantas darbe nėra nagrinėjamas. Agregavimo funkcija g gali taip pat priklausyti nuo papildomų ir, galbūt, vertintinų parametrų.

Atsitiktinio proceso $\tilde{X}$ realizaciją $X = \{X_j,\ j = 1, \dots, nm\}$ galima užrašyti kaip periodų po m seką $X = \big(X_t^{(m)},\ t = 1, \dots, n\big)$, kur $X_t^{(m)} = \big(X_{(t-1)m+1}, \dots, X_{tm}\big)$. Agreguoti aukšto dažnio duomenys gaunami, kai (2.1) funkcija pritaikoma kiekvienam duomenų periodui:

$$V_t = \big(V_{t,1}, \dots, V_{t,d}\big) = g\big(X_t^{(m)}\big), \tag{2.2}$$

kur $V_t = \big(V_{t,1}, \dots, V_{t,d}\big),\ t = 1, \dots, n$ yra d po agregavimo gautų laiko eilučių. Šiame darbe nagrinėjamas tik toks agregavimas, kada duomenys kiekviename periode agreguojami vienodai – ta pačia funkcija g.

Žemo dažnio ir agreguotiems duomenims taikomas regresijos modelis:

$$Y_t = \alpha + \sum_{j=1}^{d} \beta_j V_{t,d} + \varepsilon_t. \tag{2.3}$$

Parametrų vertinimo metodas priklauso nuo paklaidų ε_t savybių. Paklaidoms esant vienodai ir nepriklausomai pasiskirsčiusioms bei tenkinančioms bazinę prielaidą $\mathrm{E}(\varepsilon_t | V_t) = 0,\ \forall t =$

$1, \ldots, n$, (**??**) regresijos parametrams $\alpha, \beta_t, \ldots, \beta_d$ gali būti taikomas Mažiausių Kvadratų Metodas (toliau – MKM).

Dažnai patogu (m, d) agregavimo funkciją suvesti į $(m, 1)$ pavidalą, įtraukiant (**??**) regresijos koeficientus $\beta_1, \ldots, \beta_d$:

$$g(x_1, \ldots, x_m, \beta_1, \ldots, \beta_d) = \sum_{j=1}^{d} \beta_j v_j. \tag{2.4}$$

Galima išskirti du $(m, 1)$ agregavimo tipus. Vienas yra periodo efektų agregavimas, kitas – skalės efektų. Pirmojo tipo agregavimo funkcija bendruoju atveju gali būti užrašoma formule:

$$g(x) = \sum_{j=1}^{m} w(j) x_j. \tag{2.5}$$

Čia kiekvieno periodo elemento įtaką, arba svorį, nusako funkcijos w reikšmė, priklausanti nuo elemento vietos periode. Tradicinis ikimodelinis agregavimas, kai aukšto dažnio duomenys agreguojami iki periodų aritmetinių vidurkių ar paskutinės reikšmės, yra periodo efektų agregavimas. Pirmuoju atveju $w(j) = 1/m$, antruoju – $w(j) = \mathbf{1}_{j=m},\ j = 1, \ldots, m$.

Agreguojant pagal skalės efektus, periodo elemento svoris priklauso nuo jo paties reikšmės. Tai galima užrašyti formulėmis:

$$g(x) = \sum_{j=1}^{m} w(x, j) x_j, \text{ arba} \tag{2.6}$$

$$g(x) = \sum_{j=1}^{m} w(j) h(x, j), \quad x = (x_1, \ldots, x_m). \tag{2.7}$$

Periodo ir skalės efektų agregavimo funkcijos gali būti netiesinės, priklausyti nuo papildomų argumentų ar net būti šiek tiek kitokios struktūros nei čia pateikta.

2.2 Agregavimo alternatyvos

Toliau skyriuje aprašomi literatūroje nagrinėti bei nauji SDD modeliai, arba kitaip, konkrečios (2.5)-(2.7) agregavimo alternatyvos. Modeliai, kurie gali būti sudaromi agreguojant duomenis tiek pagal periodo, tiek pagal skalės efektus, iš pradžių yra aprašomi pirmuoju atveju, o tada trumpai – antruoju. Didesnė dalis agregavimo funkcijų yra literatūroje publikuotos, o trys, mano žiniomis, yra naujos.

2.2.1 ***Tiesioginis*** agregavimas

Tiesioginiu (angl. *direct*) periodo efektų agregavimu vadinsiu aukšto dažnio duomenų suskaidymą į m žemo dažnio laiko eilučių, kai kiekviena yra sudaryta iš toje pačioje periodų vietoje esančių elementų. Tai aprašanti (m, m) agregavimo funkcija yra:

$$g^{Tsg}(x_1, \ldots, x_m) = (x_1, \ldots, x_m). \tag{2.8}$$

Taigi agreguotos laiko eilutės $V_{t,1}, \ldots, V_{t,m}$ gaunamos pagal taisyklę $V_{t,i} = X_{(t-1)m+i}$, $i = 1, \ldots, m$, bei taikomas tiesinės regresijos modelis:

$$Y_t = \alpha + \sum_{j=1}^{m} \beta_j V_{t,j} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}\big(0, \sigma_\varepsilon^2\big), \tag{2.9}$$

kur $t = 1, \ldots, n,\ n > m + 1$. Modelio parametrai vertinami MKM:

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta_1, \ldots, \beta_m}{\arg\min} \left[\sum_{t=1}^{n} \left(Y_t - \alpha - \sum_{j=1}^{m} \beta_j V_{t,j} \right)^2 \right]. \tag{2.10}$$

Tiesioginis skalės efektų agregavimas gaunamas didėjimo ar mažėjimo tvarka surūšiavus aukšto dažnio duomenis kiekviename periode, tai yra (2.8) formulėje vietoje argumento x naudojant $h(x)$, kur

$$h(x) = \big(\max(x), \ldots, \min(x)\big). \tag{2.11}$$

Tiesioginis agregavimas, arba *Tiesioginė* regresija, gali būti taikoma nedažnai. Kadangi regresijos funkcijos parametrų skaičius lygus $m + 1$, tai, jei m nėra mažas ar turima ne itin daug žemo dažnio duomenų, *Tiesioginės* regresijos taikymas nėra efektyvus parametrų bei ateities prognozės tikslumo prasme. Tačiau *Tiesioginį* agregavimą verta nagrinėti vien dėl to, kad tai yra gana bendras atvejis, apimantis kitas toliau aprašomas periodo efektų agregavimo funkcijas.

2.2.2 *MIDAS* agregavimas

MIDAS (angl. *MIxed DAta Sampling*) regresija, arba agregavimas, pristatytas [2], yra dar visa neseniai aprašytas, bet jau plačiai taikomas SDD modelis. *MIDAS* regresiją patogu aprašyti kaip atskirą *Tiesioginio* agregavimo atvejį, kuris gaunamas, kai (2.9) formulėje parametrai $\beta_1, \ldots, \beta_m$ apribojami taip, kad įgautų kokios nors funkcijos pavidalą.

Konkrečiau, *Tiesioginio* agregavimo funkcija po parametrų įvertinimo gali būti užrašoma $(m, 1)$ pavidalu taip:

$$g^{Tsg}(x) = \sum_{j=1}^{m} \beta_{m+1-j} x_{m+1-j}. \tag{2.12}$$

MIDAS regresija gaunama šiuos svorius apribojus toliau aprašytu būdu. *MIDAS* agregavimo funkcijos turi bendrą pavidalą:

$$g^{MIDAS}\big(x, \lambda\big) = \sum_{j=1}^{m} w(j, \lambda) x_{m+1-j},\ \lambda \in \mathbb{R}^k,\ k < m,\ x = (x_1, \ldots, x_m) \tag{2.13}$$

(2.12) esantys svoriai apribojami iki funkcijos $w(j, \lambda)$ su vertinamu parametru λ. Straipsnyje

[3] pasiūlytos keletas funkcijų, iš kurių šiame darbe naudojamos šios:

$$w^{exp}(j, \lambda_1, \lambda_2) = \frac{\exp\left(\lambda_1 j + \lambda_2 j^2\right)}{\sum_{i=1}^{m} \exp\left(\lambda_1 i + \lambda_2 i^2\right)}, \tag{2.14}$$

$$w^{hyp}(j, \lambda_3) = \frac{b\left(\frac{j}{m}, \lambda_3\right)}{\sum_{i=1}^{m} b\left(\frac{i}{m}, \lambda_3\right)}, \tag{2.15}$$

kur $b(j, \lambda) = \frac{\Gamma(j+\lambda)}{\Gamma(j+1)\Gamma(\lambda)}$, $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R},\ \lambda_3 > 0.$

Formulėse (2.14) ir (2.15) apibrėžtas funkcijas ar svorius vadinsiu atitinkamai *eksponentiniais* ir *hiperboliniais*. Pirmoji su skirtingomis parametrų λ_1, λ_2 reikšmėmis gali įgyti įvairių pavidalų ir vertintina kaip gana lanksti, tačiau jautri nedideliems parametrų pokyčiams. Antroji labiau tinkama aprašyti lėtai kintantiems svoriams ir negali įgyti tiek daug formų, tačiau yra stabilesnė. Pažymėtina, kad tiek eksponentinė, tiek hiperbolinė funkcija apima tradicinį aritmetinio vidurkio agregavimą.

Realiems duomenims agreguota laiko eilutė $V_t^{(\lambda)}$ gaunama pagal taisyklę:

$$V_t^{(\lambda)} = \sum_{j=1}^{m} w(j, \lambda) X_{tm+1-j}. \tag{2.16}$$

Sudaromas regresijos modelis:

$$Y_t = \alpha + \beta V_t^{(\lambda)} + \varepsilon_t, \tag{2.17}$$

kurio įverčiai $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ gali būti gaunami netiesiniu MKM, kuris, tačiau, ne visada parametrus gerai įvertina. Kitu atveju įverčiai $\hat{\alpha}, \hat{\beta}$ gaunami MKM, kai parenkamas $\hat{\lambda}$:

$$\hat{\lambda} = \underset{\lambda \in \mathbb{R}^k}{\arg\min} \left[\min_{\alpha,\beta} \left(\sum_{t=1}^{n} \left(Y_t - \alpha - \beta V_t^{(\lambda)} \right)^2 \right) \right], \tag{2.18}$$

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha,\beta}{\arg\min} \left[\sum_{t=1}^{n} \left(Y_t - \alpha - \beta V_t^{(\hat{\lambda})} \right)^2 \right]. \tag{2.19}$$

$\hat{\lambda}$ įvertinimą galima realizuoti, įsivedus galimų įgyti reikšmių aibę $\lambda = \{\lambda_1, \ldots, \lambda_l\}$.

Aprašyta *MIDAS* regresija atitinka agregavimą pagal periodo efektus, o skalės efektų agregavimas gaunamas (2.13) formulėje vietoj pirmojo argumento naudojant (2.11) apibrėžtą.

MIDAS agregavimas ypač efektyvus dideliems periodų ilgiams m. Parametrizuotos svorių funkcijos dažniausiai negali tiksliai aprašyti visų periodo elementų koeficientų, taigi galimas įverčių paslinktumas ir nesuderinamumas. Tačiau, nors *Tiesioginės* regresijos įverčiai yra nepaslinkti ir suderinti, jie yra didelės dispersijos, arba neefektyvūs. Didėjant m, *MIDAS* gaunamas paslinktumas turi nusverti *Tiesioginio* agregavimo neefektyvumą.

2.2.3 *MMM* agregavimas

MMM (angl. *Min-Mean-Max*) skalės efektų $(m, 1)$ agregavimo funkcija $g^{MMM} : \mathbb{R}_+^m \to \mathbb{R}$ apibrėžiama formule:

$$g_p(x) = \frac{\sum_{j=1}^m x_j w_p(x_j)}{\sum_{j=1}^m w_p(x_j)}, \tag{2.20}$$

kur $w_p(x_j) = x_j^p,\ x = (x_1, \dots, x_m) \in \mathbb{R}_+^m,\ p \in \mathbb{R}$. Taigi

$$g_p^{MMM}(x) = \frac{\sum_{j=1}^m x_j^{p+1}}{\sum_{j=1}^m x_j^p}. \tag{2.21}$$

MMM agregavimo funkcijos pavadinimą atspindi faktas, kad esant $p = -\infty$, $p = 0$, $p = \infty$, agregavimo funkcijos reikšmės atitinkamai lygios $g_{-\infty}^{MMM}(x) = \min(x_1, \dots, x_m)$, $g_0^{MMM}(x) = \sum_{j=1}^m \frac{1}{m} x_j$, $g_\infty^{MMM}(x) = \max(x_1, \dots, x_m)$, taigi mažiausiai, vidutinei arba didžiausiai aukšto dažnio duomenų reikšmei periode.

Turimiems duomenims taikoma regresija:

$$Y_t = \alpha + \beta V_t^{(p)} + \varepsilon_t, \tag{2.22}$$

kur agreguoti duomenys gaunami pagal taisyklę:

$$V_t^{(p)} = \frac{\sum_{j=1}^m X_{(t-1)m+j}^{p+1}}{\sum_{j=1}^m X_{(t-1)m+j}^p}. \tag{2.23}$$

(2.22) parametrų įverčiams gauti gali būti taikomas netiesinis mažiausių kvadratų metodas, kuris, deja, ne visada gerai veikia. Alternatyvus vertinimo metodas yra analogiškas kaip ir *MIDAS* regresijos atveju:

$$\hat{p} = \arg\min_{p \in \mathbb{R}} \left[\min_{\alpha,\beta} \left(\sum_{t=1}^n \left(Y_t - \alpha - \beta V_t^{(p)} \right)^2 \right) \right], \tag{2.24}$$

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \left[\sum_{t=1}^n \left(Y_t - \alpha - \beta V_t^{(\hat{p})} \right)^2 \right]. \tag{2.25}$$

Praktikoje įvertis $\hat{p}$ gaunamas įsivedus galimų įgyti reikšmių aibę $p = \{p_1, \dots, p_l\}$.

MMM agregavimas naudojamas tik teigiamiems duomenims. Kitu atveju reikia apsiriboti $p > 0$ reikšmėmis arba kokiu nors būdu transformuoti aukšto dažnio duomenis į teigiamus. *MMM* yra skalės efektų agregavimas, tačiau, skirtingai nei *Tiesioginės* ir *MIDAS* skalės efektų regresijos atveju, agreguojama ne pagal (2.7), o pagal (2.6) formulę, kas svarbu sekančiame modelyje.

2.2.4 *T-MMM* agregavimas

T-MMM (*Tiesioginis Min-Mean-Max*) (m, m) agregavimas yra 2.2.3 skyrelyje aprašyto *MMM* agregavimo bendresnis atvejis. *MMM* yra skalės efektų agregavimas, tačiau, skirtingai nei *Tiesioginio* ir *MIDAS* agregavimo atveju, periodo elementai nėra sumaišomi, o tik paryškinamos jų reikšmės. Dėl šios priežasties galima papildomai įvesti periodo efektus aprašančius svorius po *MMM* agregavimo. Formulę (2.21) galima perrašyti tokiu būdu:

$$g_p(z) = \sum_{j=1}^{m} \frac{1}{m} z_j, \text{ kur} \tag{2.26}$$

$$z_j = m \frac{x_j^{p+1}}{\sum_{i=1}^{m} x_i^p}, \tag{2.27}$$

$x = (x_1, \ldots, x_m)$, $z = (z_1, \ldots, z_m)$. Tai parodo, kad periodo efektų svoriai po *MMM* agregavimo yra apribojami ir lygūs $1/m$. Periodo efektų svorių galima neapriboti ir laikyti, kad

$$g_p(z) = \sum_{j=1}^{m} w_j z_j. \tag{2.28}$$

Tai tapatu *Tiesioginio* agregavimo (2.12) funkcijai su argumentu $z = (z_1, \ldots, z_m)$. Apibendrinus *T-MMM* agregavimo funkcija išreiškiama formule:

$$g^{T\text{-}MMM}\big(x_1, \ldots, x_m\big) = \big(w_1 h_1(x), \ldots, w_m h_m(x)\big) = \left(w_1 \frac{m x_1^{p+1}}{\sum_{i=1}^{m} x_i^p}, \ldots, w_m \frac{m x_m^{p+1}}{\sum_{i=1}^{m} x_i^p}\right) \tag{2.29}$$

T-MMM agregavimo funkcija apima *Tiesioginį* ir *MMM* agregavimą, kada pirmasis gaunamas, kai $p = 0$, o antrasis, kai $w_i = 1/m$, $\forall i = 1, \ldots, m$.

Realiems duomenims m agreguotų laiko eilučių $V_{t,i}^{(p)}$, $i = 1, \ldots, m$ gaunamos pagal taisyklę:

$$V_{t,i}^{(p)} = \frac{m X_{(t-1)m+i}^{p+1}}{\sum_{j=1}^{m} X_{(t-1)m+j}^{p}}, \quad p \neq -1, \quad t = 1, \ldots, n. \tag{2.30}$$

Sąlygos $p = -1$ atveju visos laiko eilutės gaunamos lygios viena kitai, taigi atsiranda multikolinearumo problema. Agreguotiems duomenims formuluojamas regresijos modelis:

$$Y_t = \alpha + \sum_{i=1}^{m} \beta_i V_{t,i}^{(p)} + \varepsilon_t, \tag{2.31}$$

kurio parametrai vertinami netiesiniu MKM arba MKM, kaip ir *MMM* atveju, įvedus galimas įgyti reikšmes $p = \{p_1, \ldots, p_l\}$:

$$\hat{p} = \arg\min_{p \in \mathbb{R}} \left[\min_{\alpha, \beta_1, \ldots, \beta_m} \left(\sum_{t=1}^{n} \Big(Y_t - \alpha - \sum_{j=1}^{m} \beta_j V_{t,j}^{(p)} \Big)^2 \right) \right], \tag{2.32}$$

$$\left(\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_m\right) = \underset{\alpha, \beta_1, \ldots, \beta_m}{\arg\min} \left[\sum_{t=1}^{n} \left(Y_t - \alpha - \sum_{j=1}^{m} \beta_j V_{t,j}^{(\hat{p})} \right)^2 \right]. \tag{2.33}$$

2.2.5 *M-MMM* agregavimas

T-MMM agregavimas, kaip ir *Tiesioginis*, esant didesniam periodo ilgiui m, nėra tinkamas naudoti dėl didelio vertinamų parametrų skaičiaus. Kaip ir pereinant nuo *Tiesioginio* prie *MIDAS* agregavimo (2.31) regresijos parametrus galima apriboti, pavyzdžiui, (2.14) ar (2.15) funkcijomis. tokiu būdu iš *T-MMM* gauname (m, m) *M-MMM* (*MIDAS-MMM*) agregavimo funkciją:

$$g^{M\text{-}MMM}(x, \lambda, p) = \sum_{j=1}^{m} w(j, \lambda) h_j(x, p) = \sum_{j=1}^{m} w(j, \lambda) z_{m+1-j}, \tag{2.34}$$

kur $\lambda \in \mathbb{R}^k$, $z_j = m \frac{x_j^{p+1}}{\sum_{i=1}^{m} x_i^p}$, $z = (z_1, \ldots, z_m)$. Kai $p = 0$, (2.34) virsta *MIDAS* periodo efektų agregavimu, o *MIDAS* svoriams esant lygiems, (2.34) tampa *MMM* skalės efektų agregavimu.

Turimiems duomenims agreguota laiko eilutė $V_t^{(\lambda,p)}$ gaunama pagal taisyklę:

$$V_t^{(\lambda,p)} = \sum_{j=1}^{m} w(j, \lambda) \frac{m X_{tm+1-j}^{p+1}}{\sum_{i=1}^{m} X_{tm+1-i}^{p}}. \tag{2.35}$$

Tradiciškai sudaromas regredijos modelis:

$$Y_t = \alpha + \beta_i V_t^{(\lambda,p)} + \varepsilon_t, \tag{2.36}$$

kurio parametrai α, β vertinami MKM, bet parinkus λ ir p galimas įgyti reikšmes $\lambda \in \{\lambda_1, \ldots, \lambda_l\}$, $p \in \{p_1, \ldots, p_k\}$:

$$\left(\hat{\lambda}, \hat{p}\right) = \underset{\substack{\lambda \in \mathbb{R} \\ p \in \mathbb{R}}}{\arg\min} \left[\min_{\alpha, \beta} \left(\sum_{t=1}^{n} \left(Y_t - \alpha - \beta V_t^{(\lambda,p)} \right)^2 \right) \right], \tag{2.37}$$

$$\left(\hat{\alpha}, \hat{\beta}\right) = \underset{\alpha, \beta}{\arg\min} \left[\sum_{t=1}^{n} \left(Y_t - \alpha - \beta V_t^{(\hat{\lambda},\hat{p})} \right)^2 \right]. \tag{2.38}$$

Mano žiniomis, *T-MMM* ir *M-MMM* agregavimas iki šiol literatūroje nebuvo publikuotas. Šie du metodai aukšto dažnio duomenis agreguoja kartu ir pagal skalės, ir pagal periodo efektus. Vienas iš pagrindinių šio darbo tikslų yra palyginti šių modelių teikiamus rezultatus su kitų aprašytų modelių rezultatais.

2.2.6 *Funkcinis* agregavimas

Aukšto dažnio duomenis galima laikyti funkciniais, t.y. nediskrečios funkcijos stebėjimais kai kurioms argumento reikšmėms[1]. Konkrečiau, laiko eilutę $X = \{X_j,\ j = 1, \dots, nm\}$ suskaidžius į n periodų po m stebėjimų, kiekvieno periodo duomenis galima laikyti kaip nežinomos funkcijos rekšmes nebūtinai vienodai nutolusioms argumento reikšmėms. Tardami, kad žemo dažnio duomenys Y priklauso nuo kokios nors įvertintų funkcijų charakteristikos, atitinkamu būdu galima agreguoti aukšto dažnio duomenis.

Aprašysiu paprastą funkcinio agregavimo variantą, patektą [5] straipsnyje. Bet kurio k-ojo periodo aukšto dažnio duomenis $X_{(k-1)m+1}, \dots, X_{km}$ interpretuojame kaip funkcijos $x_k(t)$ reikšmes argumento taškuose $t_1, \dots, t_m$ ir darome prielaidą, kad $x_k(t)$ turi tokį pavidalą:

$$x_k(t) = g_{k,1}\psi_1(t) + \cdots + g_{k,d}\psi_d(t) + \varepsilon_t. \tag{2.39}$$

Bazinės funkcijos $\psi_i(t)$ yra tiesiškai nepriklausomos, o šiuo atveju įgyja laiptinės funkcijos pavidalą:

$$\psi_1(t,a) = \mathbf{1}_{\{t\in(0,a_1]\}}(t), \dots, \psi_d(t,a) = \mathbf{1}_{\{t\in(a_{d-1},m]\}}(t), \quad t \in [0,m] \tag{2.40}$$

kur $a = (a_1, \dots, a_{d-1}) \in \mathbb{R}^{d-1}$ yra toks lūžio taškų vektorius, kad $a_j \in [j, m-1]$, $a_j \in \mathbb{R}$, $j = 1, \dots, d-1$.

Konkrečiam a ir $\forall k = 1, \dots, n$ formuluojamas regresijos modelis:

$$X_{(k-1)m+j} = g_{k,1}(a)\psi_1(t_j,a) + \cdots + g_{k,d}(a)\psi_d(t_j,a) + \varepsilon_{(k-1)m+j}, \quad j = 1, \dots, m, \tag{2.41}$$

kurio koeficientai dėl (2.41) regresijos paprastumo gali būti užrašyti:

$$\begin{aligned}
\hat{g}_{k,1}(a) &= \frac{1}{\lfloor a_1 \rfloor} \sum_{j=1}^{\lfloor a_1 \rfloor} X_{(k-1)m+j},\\
\hat{g}_{k,i}(a) &= \frac{1}{\lfloor a_i \rfloor - \lceil a_{i-1} \rceil + 1} \sum_{j=\lceil a_{i-1} \rceil}^{\lfloor a_i \rfloor} X_{(k-1)m+j},\\
\hat{g}_{k,d}(a) &= \frac{1}{m - \lfloor a_{d-1} \rfloor + 1} \sum_{j=\lceil a_{d-1} \rceil}^{m} X_{(k-1)m+j},
\end{aligned} \tag{2.42}$$

dėl kurio modelio (2.41) sudarinėti nebūtina. Parametro a įvertis gaunamas sudarius pagrindinę regresiją:

$$Y_t = \alpha + \beta_1 g_{t,1}(a) + \cdots + \beta_d g_{t,d}(a) + u_t, \tag{2.43}$$

$$\hat{a} = \underset{a\in\{a_1,\dots,a_l\}}{\arg\min} \left[\min_{\alpha,\beta_1,\dots,\beta_d} \left(\sum_{t=1}^{n} \left(Y_t - \alpha - \sum_{j=1}^{d} \beta_j g_{t,j}(a) \right)^2 \right) \right], \tag{2.44}$$

[1] Apie funkcinę duomenų analizę galima rasti knygoje [6].

$$\left(\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_d\right) = \underset{\alpha, \beta_1, \ldots, \beta_d}{\arg\min} \left[\sum_{t=1}^{n} \left(Y_t - \alpha - \sum_{j=1}^{d} \beta_j g_{t,j}(\hat{a}) \right)^2 \right]. \tag{2.45}$$

Įvertis $\hat{a}$ gaunamas pagal visas galimas $a \in \{a_1, \ldots, a_l\}$ reikšmes minimizavus (2.43) liekanų kvadratų sumą.

Skaičius d, parodantis, kuria eile agreguojami duomenys, čia gali būti laikomas parametru. Kadangi d parodo taip pat koeficientų $\beta_1, \ldots, \beta_d$ skaičių, tai įvertį $\hat{d}$ galima gauti pagal visas $d \in \{1, \ldots, m\}$ reikšmes minimizavus kurį nors informacinį kriterijų. Taip *Funkcinis* (m, d) agregavimas virsta į $\left(m, \hat{d}\right)$ agregavimą. Pažymėtina, kad tokiu atveju lūžio taškų a parinkimo variantų yra 2^{m-1}, taigi parametrų vertinimo skaičiavimams atlikti gali prireikti daug laiko.

Skalės efektų *Funkcinis* agregavimas gaunamas, kaip ir anksčiau, kiekviename periode surūšiavus didėjimo tvarka aukšto dažnio duomenis.

Kaip ir *MIDAS*, *Funkcinis* agregavimas reiškia tam tikrą *Tiesioginės* regresijos parametrų apribojimą. Šiuo atveju parametrai suskirstomi į grupes, kurių kiekvienoje jie lygūs tarpusavyje. Šios prielaidos atsisakoma toliau aprašytame *Laužtinės* agregavimo būde.

2.2.7 ***Laužtinės*** agregavimas

Laužtinės agregavimu vadinsiu metodą, kuris, nors išplaukė iš *Funkcinio* būdo, tačiau taip pat gali būti laikomas *MIDAS* regresijos variantu. Paprastumo dėlei, nagrinėkime *Funkcinį* $(m, 2)$ agregavimą. Šiuo atveju aukšto dažnio duomenų perioduose pagal (2.44) įvertinamas lūžio taškas, o agreguotas dvi laiko eilutes pagal (2.42) sudaro aukšto dažnio duomenų periodų vidurkiai iki lūžio taško ir už lūžio taško. Paprasta idėja yra imti ne vidurkį, kada periodo efektų svoriai yra lygūs, o aprašyuti juos kokia nors funkcija, sakykime, tiese. Taip pirmoje ir antroje periodo dalyja svoriai įgyja nesikertančių tiesių pavidalą, iš ko ir kilo metodo pavadinimas.

Konkrečiau, (2.42) formulę $(m, 2)$ atveju, pridėjus tiesiškai proporcingus svorius, galima užrašyti:

$$\begin{aligned} g_{k,1}(a) &= \sum_{j=0}^{\lfloor a \rfloor - 1} \big(\alpha_1 + \beta_1 j\big) X_{(k-1)m+1+j}, \\ g_{k,2}(a) &= \sum_{j=0}^{m - \lfloor a \rfloor - 1} \big(\alpha_2 + \beta_2 j\big) X_{(k-1)m+\lfloor a \rfloor+1+j}, \end{aligned} \tag{2.46}$$

kur $a, \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$, ir, įvertinus šiuos parametrus, gaunamos agreguotos laiko eilutės. Kad gautume parametrų MKM įverčius, reikia pertvarkyti modelį. (2.46) formulės pirmąją dalį galima perrašyti kitaip:

$$g_{k,1}(a) = \alpha_1 X_{(k-1)m+1} + \big(\alpha_1 + \beta_1\big) X_{(k-1)m+2} + \cdots + \big(\alpha_1 + (\lfloor a \rfloor - 1)\beta_1\big) X_{(k-1)m+\lfloor a \rfloor}. \tag{2.47}$$

Nors nebūtina, bet pravartu perrašyti (2.47) per pirmą ir paskutinį sumos svorius, reiškiančius tiesės pradžios ir pabaigos reikšmes. Pažymėjus $w_1 = \alpha_1$, $v_1 = \alpha_1 + (\lfloor a \rfloor - 1)\beta_1$, tada gauname:

$$\begin{aligned} g_{k,1}(a) = \ldots = w_1 &\left(\frac{\lfloor a \rfloor - 1}{\lfloor a \rfloor - 1} X_{(k-1)m+1} + \cdots + \frac{1}{\lfloor a \rfloor - 1} X_{(k-1)m+\lfloor a \rfloor - 1} \right) + \\ + v_1 &\left(\frac{1}{\lfloor a \rfloor - 1} X_{(k-1)m+2} + \cdots + \frac{\lfloor a \rfloor - 1}{\lfloor a \rfloor - 1} X_{(k-1)m+\lfloor a \rfloor} \right). \end{aligned} \tag{2.48}$$

Panašiai gaunama ir antroji laiko eilutė $g_{k,2}(a)$. Dideliuose (2.48) skliaustuose esančias sumas pažymėjus $V_{t,11}$ ir $V_{t,12}$, $t = 1, \ldots, n$, sudaromas pagrindinis regresijos modelis:

$$Y_t = \alpha + w_1 V_{t,11} + v_1 V_{t,12} + w_2 V_{t,21} + v_2 V_{t,22} + \varepsilon_t. \tag{2.49}$$

Lūžio taškas įvertinamas kaip ir *Funkcinio* būdo atveju, o (2.49) regresijos įverčiai gaunami MKM. Kartais galima norėti, kad svorių tiesės būtų tik teigiamoje koordinačių plokštumos dalyje. Tą užtikrina regresijos koeficientų, reiškianių tiesės pradžios ir pabaigos reikšmes, teigiamumas: $w_1, v_1, w_2, v_2 > 0$. Tuomet regresijos parametrai vertinami su šiuo apribojimu.

Laužtinės metodas yra *Funkcinio* būdo praplėtimas. Čia aprašytas $(m, 2)$ agregavimas tiesiogiai gali būti taikomas ir (m, d) atveju, tačiau didesniems d regresijos parametrų skaičius auga tiesiškai, o parametrų vertinimo laikas – eksponentiškai. Kita vertus, tik dvi vertinamo ilgio ir nuokrypių svorių tiesės gali aprašyti pakankamai įvarias formas.

2.3 Prognozavimas skirtingo dažnio duomenų modeliais

Skirtingo dažnio duomenų regresinius modelius galima naudoti gauti žemo dažnio duomenų ateities prognozėms. Tikėtina, kad aukšto dažnio duomenų naudojimas dėl galimo geresnio kintamųjų priklausomybės aprašymo padeda gauti tikslesnes prognozes nei įprastų modelių.

Padarius prielaidą, kad žemo dažnio duomenys priklauso nuo aukšto dažnio duomenų praeities, (2.3) modelis užrašomas:

$$Y_t = \alpha + \sum_{j=1}^{d} \beta_j V_{t-1,j} + \varepsilon_t, \tag{2.50}$$

kur $V_{t-1,j}$ yra pagal kokią nors (2.2) agregavimo funkciją gauti duomenys. Toks modelis naudingas tuo, kad vieno žingsio Y_t prognozėms gauti reikalingiems agreguotiems duomenims nereikia formuluoti jų prognozavimo modelio. Įvertinus modelio parametrus su duomenimis iki laiko momento n, vieno žingsnio prognozė $\hat{Y}_{n+1}$, kaip įprasta, gaunama pagal formulę:

$$\hat{Y}_{n+1} = \hat{\alpha} + \sum_{j=1}^{d} \hat{\beta}_j V_{n,j}. \tag{2.51}$$

Modelį (2.50) galima praplėsti, įtraukiant daugiau agreguotų duomenų vėlavimų:

$$Y_t = \alpha + \sum_{j=1}^{d} \beta_{j,1} V_{t-1,j} + \cdots + \sum_{j=1}^{d} \beta_{j,J} V_{t-J,j} + \varepsilon_t, \tag{2.52}$$

Modelis su daugiau vėlavimų gali stipriai pagerinti prognozių tikslumą. Vėlavimų skaičiaus galima ir nedidinti, o agreguoti daugiau nei vieno periodo duomenis kokia nors funkcija g:

$$\big(V_{t,1}, \ldots, V_{t,d}\big) = g\big(X_{tm-J+1}, \ldots, X_{tm}\big), \quad t = 1, \ldots, n, \tag{2.53}$$

ir tada sudaryti (2.50) regresijos modelį. Kiekvienas 2.2 skyrelyje aprašytas modelis su konkrečia agregavimo funkcija gali būti praplėstas panašiai, bet ne vieninteliu būdu. Be to, prognozavimo tipo modeliai naudoti 4 skyriuje aprašytiems rezultatams gauti, todėl reikalingas trumpas jų apibūdinimas.

2.3.1 Prognozavimas ***Tiesioginiu*** ir ***T-MMM*** modeliais

Tiesioginės regresijos modelio prognozavimui skirtas pavidalas yra

$$Y_t = \alpha + \sum_{j=1}^{m} \beta_j X_{(t-1)m+1-j} + \varepsilon_t. \tag{2.54}$$

Viršutinę (2.54) esančios sumos ribą galima pasirinkti ar net įvertinti kaip nelygią m:

$$Y_t = \alpha + \sum_{j=1}^{J} \beta_j X_{(t-1)m+1-j} + \varepsilon_t. \tag{2.55}$$

$J > m$ reiškia, kad įtraukiami daugiau nei vieno aukšto dažnio duomenų periodo nariai. Dažniausiai prasmingiau yra pasirinkti $J < m$, nes taip sumažinamas vertinamų parametrų skaičius. Pavyzdžiui, galima aprašyti mėnesinių duomenų priklausomybę ne nuo kiekvienos praeito mėnesio dienos, o nuo J paskutiniųjų. Jei m yra gana didelis, o žemo dažnio duomenų „atmintis" yra ypač trumpa, tokiu būdu *Tiesioginiu* agregavimu gautos prognozės gali būti tikslios.

Prognozavimas *T-MMM* modeliu panašus į prognozavimą *Tiesiogine* regresija. *T-MMM* atveju (2.31) regresija su pagal (2.30) agreguotais duomenimis prognozavimo regresijos pavidalą įgauna, kai užrašoma taip:

$$Y_t = \alpha + \sum_{j=1}^{m} \beta_j V^{(p)}_{t-1,m+1-j} + \varepsilon_t, \tag{2.56}$$

Viršutinę (2.31) regresijos sumos ribą dažniausiai geriausia pakeisti $J < m$, taip gaunant *MMM* agregavimą tik paskutiniams periodo nariams, bet sumažinant vertintinų regresijos parametrų skaičių.

2.3.2 Prognozavimas *MIDAS* ir *M-MMM* modeliais

MIDAS regresija gali būti praplėsta panašiai kaip ir *Tiesioginė*. Formulėje (2.16) pakeitus viršutinę sumos ribą skaičiumi J, gauname:

$$V_t^{(\lambda)} = \sum_{j=1}^{J} w\big(j, \lambda\big) X_{tm+1-j}, \tag{2.57}$$

ir sudaromas prognozavimo regresijos modelis su vienu vėlavimu:

$$Y_t = \alpha + \beta V_{t-1}^{(\lambda)} + \varepsilon_t. \tag{2.58}$$

M-MMM prognozavimo modelis gaunamas analogiškai.

Toks būdas, kuriuo tik vienu vėlavimu galima aprašyti ir tolimos praeities svorius, naudojamas pačių MIDAS regresijos autorių straipsnyje [2]. Tačiau galima tiesiogiai įtraukti viename periode agreguotų duomenų vėlavimus:

$$Y_t = \alpha + \beta_1 V_{t-1}^{(\lambda_1)} + \cdots + \beta_J V_{t-J}^{(\lambda_J)} + \varepsilon_t, \tag{2.59}$$

$$V_{t-j}^{(\lambda_j)} = \sum_{i=1}^{m} w\big(\lambda_j, i\big) X_{(t-j-1)m+i}.$$

Regresija (2.59) galimai geriau aprašo duomenų priklausomybę, tačiau, vertinant parametrus MKM, $\lambda_1, \ldots, \lambda_J$ parinkimas gali užtrukti. Vieno tokio modelio parametrų įvertinimas nesudėtingas, tačiau atlikti simuliacinę analizę ar skaičiuoti prognozių tikslumą paprastu kompiuteriu labai užtruktų.

2.3.3 Prognozavimas *MMM* modeliu

MMM prognozavimo regresijos atitikmuo su pagal (2.23) gautais agreguotais duomenimis yra:

$$Y_t = \alpha + \beta V_{t-1}^{(p)} + \varepsilon_t, \tag{2.60}$$

o jo praplėtimas su daugiau agreguotų vėlavimų:

$$Y_t = \alpha + \beta_1 V_{t-1}^{(p_1)} + \cdots + \beta_J V_{t-1}^{(p_J)} + \varepsilon_t. \tag{2.61}$$

Parametrų $p_1, \ldots, p_J$ įverčiai gali būti gaunami kaip formulėje (2.32) gaunamas $\hat{p}$, tačiau tai reikalauja didelies apimties kompiuterio skaičiavimų, kurių laikui sumažinti galima daryti parametrų apribojimą $p_1 = \cdots = p_J$.

Jeigu manoma, kad žemo dažnio duomenys priklauso nuo aukšto dažnio duomenų tik dalyje periodo, *MMM* būdu agreguoti duomenys gaunami atitinkamai: viršutinės (2.23) sumų ribos pakeičiamos J, $J < m$, ir sudaromas (2.60) regresijos modelis.

2.3.4 Prognozavimas *Funkciniu* ir *Laužtinės* modeliais

Jei negautume didelio parametrų skaičiaus, *Funkcinio* būdo prognozavimo modelis galėtų būti formuluojamas panašiai kaip ir *MMM* atveju (2.61) formulėje. Parametro skaičiaus problemą galimai išspręstų konservatyvus informacinis kriterijus, tačiau galimas ir kitas variantas. Kaip ir *MIDAS* prognozavimo modelio sudarymo atveju, galima duomenis agreguoti didesnėje ar mažesnėje erdvėje nei vienas periodas. Užtenka pakeisti m (2.40) formulėje bei atitinkamai perdaryti sekančias lygtis, ir formuluojamas *Funkcinio* būdo prognozavimo modelis:

$$Y_t = \alpha + \beta_1 g_{t-1,1}(a) + \cdots + \beta_d g_{t-1,1} + u_t. \tag{2.62}$$

Tokio modelio parametrų skaičius iš esmės nesikeičia, ir gali aprašyti tolimą ar artimą žemo dažnio duomenų priklausomybę nuo agreguotų duomenų. Per daug nesiplečiant, *Laužtinės* prognozavimo modelis formuluojamas pagal tokį patį principą.

3 Monte Carlo eksperimentas

Norint ne tik teoriškai palyginti 2.1 aprašytus skirtingo dažnio duomenų modelius, galima ir reikalinga imitacinė analizė. Monte Carlo eksperimentu tyriau, kaip gerai skirtingais modeliais aprašomi įvairūs duomenis generuojantys procesai. Imitacinė analizė yra [5] aprašyto eksperimento papildymas – įtrauktas prognozių prognozių tikslumo skaičiavimo atvejis bei periodo ilgis $m = 12$ pakeistas į $m = 22$. Pirmojoje skyriaus dalyje aprašomos eksperimento sąlygos, antrojoje pateikiami ir aptariami generuotų procesų parametrų vertinimo rezultatai, o trečiojoje tas pats atliekama, skaičiuojant prognozių tikslumą.

3.1 Duomenis generuojantys procesai

Duomenis generuojantys procesai yra analogiški [5] naudojamiems. Žemo dažnio procesas Y užrašomas lygtimi:

$$Y_t = \beta g_i(X_{(t-1)m+1}, \ldots, X_{tm}) + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathbb{N}\left(0, \sigma_\varepsilon^2\right), \tag{3.1}$$

o aukšto dažnio procesas X yra tiesiog teigiamų ir nepriklausomai normaliai pasiskirsčiusių atsitiktinių dydžių seka:

$$X_j = 10 + u_j, \quad u_j \overset{i.i.d}{\sim} \mathbb{N}\left(0, \sigma_u^2\right), \tag{3.2}$$

kur (3.1) ir (3.2) formulėse $t = 1, \ldots, n$; $j = 1, \ldots, nm$; $\sigma_\varepsilon^2 \geq 0$; $\sigma_u^2 > 0$; $\forall j, t,\ E(\varepsilon_t u_j) = 0$.

Naudojamos duomenis generuojančios funkcijos yra tokios:

1) $g_1(x) = \sum_{j=1}^m \frac{1}{m} x_j, \quad j = 1, \ldots, m;$

2) $g_2(x) = \sum_{j=1}^m w_j x_j, \quad w_j = \mathbf{1}_{\{j=m\}}, \quad j = 1, \ldots, m;$

3) $g_3(x) = \sum_{j=1}^m w_j x_j, \quad w_j = \frac{\exp(\lambda_1 j + \lambda_2 j^2)}{\sum_{i=1}^m \exp(\lambda_1 i + \lambda_2 i^2)}, \quad j = 1, \ldots, m;$

4) $g_4(x) = \sum_{j=1}^m w_j x_j, \quad w_j = \frac{1}{m}\left(1 + (-1)^j - \frac{\sum_{i=1}^m (-1)^i}{m}\right), \quad j = 1, \ldots, m;$

5) $g_5(x) = \sum_{j=1}^m w_j x_j, \quad w_j = \frac{x_j^p}{\sum_{i=1}^m x_i^p}, \quad p = 12, \quad j = 1, \ldots, m;$

6) $g_6(x) = \sum_{j=1}^m w_j x_j, \quad w_j = \mathbf{1}_{\max(x_1, \ldots, x_m) = x_j}, \quad j = 1, \ldots, m;$

7) $g_7(x) = \sum_{j=1}^m w_j s_j(x), \quad j = 1, \ldots, m.$
   Funkcija s surūšiuoja x didėjimo tvarka: $s(x) = s(x_1, \ldots, x_m) = \big(s_1(x), \ldots, s_m(x)\big) = \big(\min(x), \min(x/\min(x)), \ldots, \max(x)\big)$, w_j apibrėžti kaip g_3 atveju;

8) $g_8(x) = \sum_{j=1}^m w_j s_j(x), \quad j = 1, \ldots, m$, kur w_j apibrėžti kaip g_4 atveju.

Pirmosios dvi agregavimo funkcijos atstovauja tradiciniam priešmodeliniam agregavimui, kai pasirenkamas aukšto dažnio duomenų periodo vidurkis arba paskutinė reikšmė. Trečiojoje funkcijoje svoriai aprašomi *MIDAS* eksponentine funkcija, taigi natūraliai tokį procesą geriausiai turi aprašyti *MIDAS* regresija. Ketvirtoji funkcija yra periodinė, sudaryta iš dviejų svorių w_j reikšmių, kiekviena iš kurių kartojasi kas antrą žingsnį. Pirmosios keturios funkcijos atitinka periodo efektų agregavimą. Penktoji funkcija yra analogiška *MMM* agregavimui. Šeštoji agregavimo funkcija išrenka maksimalią reikšmę iš aukšto dažnio duomenų kiekvieno periodo. Septintoji funkcija didėjimo tvarka surūšiuoja kiekvieno periodo aukšto dažnio duomenų reikšmes, o svoriai parenkami kaip g_3 atveju, taigi šiam agregavimui tiesiogiai atstovauja *MIDAS* skalės efektų modelis. Aštuntoji funkcija tokia pati kaip ketvirtoji, tik agreguojama po minėto duomenų surūšiavimo. Paskutinės keturios funkcijos atitinka skalės efektų agregavimą.

Bendras sudaromų SDD modelių pavidalas ir iš jo išplaukiantys skirtingų agregavimo funkcijų modeliai yra analogiški 2 skyriuje aprašytiems.

3.2 Vertinamų parametrų tikslumo eksperimentas

Šio eksperimento tikslas yra įvertinti, kaip gerai SDD modeliai aprašo duomenis generuojančius procesus, matuojant pagal β tikslumą (3.1) formulėje. Analogiški skaičiavimai atlikti [5] straipsnyje. Taip pat papildomai įvertinamas *T-MMM* ir *M-MMM* modelių tikslumas, bei rezultatai palyginami su kitais modeliais. *Laužtinės* modelio rezultatai parametro tikslumo atveju skaičiuoti nebuvo dėl gana sudėtingos koeficientų interpretacijos. Koeficiento β įverčio tikslumas matuojamas vidutinės kvadratinės paklaidos MSE (angl. *Mean Squared Error*) kriterijumi, kuris užrašomas formule:

$$\mathrm{MSE}(\hat{\beta}, \beta) = \frac{1}{s}\sum_{i=1}^{s}(\hat{\beta}_i - \beta)^2, \quad \hat{\beta} = (\hat{\beta}_i, \ldots, \hat{\beta}_s), \tag{3.3}$$

o s lygus eksperimento simuliacijų skaičiui, kuris šiuo atveju lygus 10000.

Pirmojoje lentelėje pateikti parametro tikslumo įvertinimai, esant sąlygai $\sigma_\varepsilon^2 = 0$, kuri reiškia, kad žemo dažnio duomenys visiškai priklauso nuo aukšto dažnio duomenų po jų agregavimo su viena iš aprašytų agregavimo funkcijų. Atmetus atsitiktinių paklaidų sukeliamą parametrų netikslumą, tokia analizė parodo, kaip gerai modelio agregavimas atitinka duomenis generuojančio proceso funkciją. Jei MSE gaunamas lygus nuliui, tai reiškia, kad duomenis generuojančio proceso agregavimo funkcija yra tik atskiras modelio agregavimo atvejis. Rezultatai pateikti 1-oje lentelėje.

1-os lentelės stulpelius atitinka duomenis generuojančios agregavimo funkcijos, o eilutes – skirtingo dažnio duomenų modeliai. Paskutiniuose trijuose lentelės stulpeliuose pateiktas kiekvieno modelio parametro β vidutinis tikslumas, vertinant atitinkamai periodo efektus aprašančias agregavimo funkcijas g_1-g_4, skalės efektus – g_5-g_8 bei visas funkcijas g_1-g_8. Viršutinė lentelės dalis atitinka $n = 200$, $m = 3$ atvejį, apatinė atitinka $n = 100$, $m = 6$ atvejį. Taigi čia periodų ilgiai nedideli ir $m = 3$ atveju atitinka, pavyzdžiui, ketvirtinius ir mėnesinius, o $m = 6$ atveju – pusmetinius ir mėnesinius duomenis. Čia g_3 ir g_7 agregavimo funkcijose parinkta $\lambda_1 = -1{,}3, \lambda_2 = 0{,}1$. Paskutiniuose trijuose stulpeliuose trys mažiausi skaičiai (arba keturi, jei yra vienodų) yra pajuodinti.

1 lentelė. Parametro β tikslumas be paklaidų įtakos, $100 \times \mathrm{MSE}(\hat{\beta}, \beta)$

$m=3,\ n=200,\ \sigma_\varepsilon=0$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	0,00	1,03	0,24	0,45	0,38	0,35	0,09	0,08	0,43	0,23	0,33
Pask. reikšmė	44,5	0,00	79,8	79,5	45,9	44,8	44,6	44,6	51,0	45,0	48,0
Maksimali reikšmė	16,4	17,4	16,7	16,8	0,38	0,00	34,1	22,6	16,8	14,3	15,5
Tiesioginis per. ef.	0,00	0,00	0,00	0,00	0,38	0,34	0,09	0,08	**0,00**	0,22	**0,11**
MIDAS per. ef.	0,00	0,00	0,00	0,00	0,38	0,34	0,09	0,08	**0,00**	0,22	**0,11**
Funkcinis per. ef.	0,00	0,00	0,00	0,00	0,39	0,34	0,09	0,08	**0,00**	0,23	**0,11**
MMM	0,00	1,00	0,25	0,45	0,00	0,00	0,08	0,08	0,43	0,04	0,23
Tiesioginis sk. ef.	0,00	1,00	0,24	0,45	0,11	0,00	0,00	0,00	0,42	**0,03**	0,23
MIDAS sk. ef.	0,00	0,98	0,25	0,46	0,05	0,00	0,00	0,00	0,42	**0,01**	0,22
Funkcinis sk. ef.	0,00	1,02	0,24	0,45	0,11	0,00	0,00	0,00	0,43	**0,03**	0,23
T-MMM	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,08	**0,00**	0,04	**0,02**
M-MMM	0,00	0,09	0,02	1,11	0,00	0,21	0,12	0,08	0,31	0,10	0,20
$m=3,\ n=200,\ \sigma_\varepsilon=0$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	0,00	5,05	1,41	1,01	1,27	1,57	0,52	0,11	1,87	0,87	1,37
Pask. reikšmė	69,7	0,0	94,8	44,6	71,5	70,2	69,8	69,5	52,3	70,3	61,3
Maksimali reikšmė	36,1	38,5	36,7	36,6	2,19	0,00	63,6	24,5	37,0	22,6	29,8
Tiesioginis per. ef.	0,00	0,00	0,00	0,00	1,29	1,64	0,54	0,11	**0,00**	0,90	**0,45**
MIDAS per. ef.	0,00	0,00	0,00	1,06	1,37	1,60	0,54	0,11	0,27	0,91	0,59
Funkcinis per. ef.	0,00	0,00	0,00	0,00	1,33	1,61	0,53	0,11	**0,00**	0,90	**0,45**
MMM	0,00	5,32	1,37	1,06	0,00	0,00	0,29	0,08	1,94	0,09	1,02
Tiesioginis sk. ef.	0,00	5,36	1,44	1,06	0,26	0,00	0,00	0,00	1,97	**0,07**	1,02
MIDAS sk. ef.	0,00	5,16	1,42	1,04	0,11	0,00	0,00	0,06	1,91	**0,04**	0,97
Funkcinis sk. ef.	0,00	4,92	1,34	0,98	0,26	0,00	0,02	0,03	1,81	**0,08**	0,94
T-MMM	0,00	0,00	0,00	0,00	0,00	0,01	0,29	0,08	**0,00**	0,10	**0,05**
M-MMM	0,00	0,24	1,41	1,19	0,00	1,38	0,57	0,06	0,71	0,50	0,61

Tradiciniai grubūs agregavimo metodai duomenis parametro tikslumo prasme aprašo prasčiausiai, iš kurių blogiausius rezultatus rodo *Paskutinės reikšmės* agregavimas. Trijų bazinių modelių – *Tiesioginio*, *MIDAS* ir *Funkcinio* – rezultatai gana panašūs. Pastarieji periodo efektus aprašantys modeliai puikiai aprašo periodo efektų duomenis ir žymiai prasčiau skalės efektų duomenis, ir atvirkščiai. *MMM* modelio rezultatai panašūs į kitų skalės efektus aprašančių modelių rezutatus. Reikia pažymėti, kad *Tiesioginio* agregavimo rezultatai nėra prastesni nei kitų modelių dėl atsitiktinės paklaidos nebuvimo, kada parametrų dispersijos nedidina jų skaičius. Geriausi rezultatai gaunami *T-MMM* modeliu, kuris, „paveldėdamas" *Tiesioginio* ir *MMM* agregavimo savybes, tiek periodų, tiek skalės efektus aprašo labai gerai. *M-MMM* modelis tokių gerų rezultatų neduoda dėl periodo parametrų apribojimų. Periodų efektų agregavimo duomenis šis metodas aprašo prasčiau nei periodo efektų modeliai, tačiau geriau nei skalės efektų modeliai. Tas pats galioja ir skalės efektų generuotiems duomenims. Reikia pažymėti, kad esminių skirtumų tarp gautų rezultatų ir [5] straipsnyje pateiktų nėra.

Jau minėjau, kad dėl atsitiktinių paklaidų nebuvimo (3.1) lygtyje, modelių rezultatai ne visiškai gali būti palyginami. Skirtingus modelius paklaidos veikia skirtingai. 2-os lentelės struktūra yra analogiška 1-ai lentelei. Vienintelis skirtumas, kad (3.1) lygtyje nustatyta $\sigma_\varepsilon^2 = 1$, taigi rezultatams įtaką daro nemodeliuojama atstitiktinė paklaida, tačiau dabar galima detaliai palyginti visų modelių parametro tikslumą.

Rezultatai pateikiami 2-oje lentelėje. Kiekvieno stulpelio trys mažiausi skaičiai yra pajuodinti. Tradiciniai agregavimo metodai bendrai duoda prasčiausius rezultatus. *Paskutinės* ir *Maksimalios* reikšmės agregavimo funkcijomis gaunamas β tikslumas yra itin prastas, tuo

2 lentelė. Parametro β tikslumas su paklaidų įtaka, $100 \times \mathrm{MSE}(\hat{\beta}, \beta)$

$m = 3,\ n = 200,\ \sigma_\varepsilon = 1$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	**1,54**	2,57	1,79	2,03	1,90	1,83	1,57	1,60	1,98	1,73	1,85
Pask. reikšmė	45,0	**0,51**	80,3	79,7	46,7	45,5	45,2	45,2	51,4	45,7	48,5
Maksimali reikšmė	17,4	18,2	17,7	17,9	1,30	**0,86**	35,1	23,5	17,8	15,2	16,5
Tiesioginis per. ef.	**1,54**	1,53	1,54	**1,55**	1,89	1,88	1,65	1,62	**1,54**	1,76	1,65
MIDAS per. ef.	1,56	**1,16**	**1,46**	**1,35**	1,90	1,85	1,64	**1,58**	**1,38**	1,74	**1,56**
Funkcinis per. ef.	**1,54**	1,61	**1,52**	**1,57**	1,99	1,92	1,65	1,68	1,56	1,81	1,69
MMM	1,55	2,52	1,79	1,98	**1,33**	**1,21**	1,68	1,60	1,96	**1,46**	1,71
Tiesioginis sk. ef.	1,57	2,53	1,83	1,97	1,68	1,53	**1,54**	**1,53**	1,98	1,57	1,77
MIDAS sk. ef.	1,57	2,51	1,79	2,01	**1,21**	1,32	**1,45**	**1,41**	1,97	**1,35**	1,66
Funkcinis sk. ef.	1,55	2,50	1,72	1,96	1,67	1,52	**1,53**	**1,58**	1,93	1,58	1,75
T-MMM	1,57	1,55	**1,49**	1,58	1,34	**1,22**	1,77	1,64	**1,55**	1,49	**1,52**
M-MMM	**1,51**	**0,89**	2,35	2,25	**1,25**	1,27	1,64	1,59	1,75	**1,44**	**1,59**
$m = 3,\ n = 200,\ \sigma_\varepsilon = 1$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	**6,10**	11,6	7,47	7,43	7,29	7,82	6,57	6,35	8,15	7,01	7,58
Pask. reikšmė	70,7	**1,02**	95,2	45,9	72,1	71,3	70,9	70,6	53,2	71,2	62,2
Maksimali reikšmė	38,4	41,8	39,4	39,5	**4,65**	**2,50**	65,9	27,0	39,8	25,0	32,4
Tiesioginis per. ef.	6,55	6,74	**6,58**	**6,46**	7,90	8,32	6,90	6,61	**6,58**	7,43	7,01
MIDAS per. ef.	6,64	**3,23**	**5,14**	8,73	7,82	8,17	7,12	6,81	**5,94**	7,48	6,71
Funkcinis per. ef.	**6,13**	6,71	8,08	8,89	7,28	7,96	7,05	**6,26**	7,45	7,14	7,30
MMM	6,34	11,2	7,50	7,48	4,98	**4,19**	**5,59**	6,33	8,12	**5,27**	**6,70**
Tiesioginis sk. ef.	6,56	12,0	7,81	7,66	6,70	6,62	6,52	6,58	8,52	6,61	7,56
MIDAS sk. ef.	**6,13**	10,0	**7,13**	**6,86**	**4,48**	4,72	**5,41**	**6,13**	7,53	**5,19**	**6,36**
Funkcinis sk. ef.	6,38	12,0	7,75	**7,45**	6,88	6,23	6,49	**6,28**	8,40	6,47	7,44
T-MMM	8,00	6,91	7,17	7,52	5,31	**4,47**	6,91	8,52	7,40	6,30	6,85
M-MMM	6,89	**4,67**	7,84	8,23	**4,73**	4,74	**5,68**	6,71	**6,91**	**5,47**	**6,19**

tarpu paprasto *vidurkio* rezultatai tik šiek tiek nusileidžia sudėtingesniems modeliams.

Iš trijų bazinių periodo efektų modelių – *Tiesioginio*, *Funkcinio* ir *MIDAS* – geriausi rezultatai yra pastarojo. *Tiesioginės* regresijos rezultatai tik šiek tiek prastesni, daugiausiai dėl to, kad periodo ilgiai m, lemiantys vertintinų parametrų skaičių, yra nedideli. Lentelėje pateikti abiejų *Funkcinių* metodų rezultatai ne visai atitinka straipsnyje pateiktus ir $m = 6$ atveju yra prastesni. Pagrindinė skirtumo priežastis yra didelė skaičiavimų trukmė, dėl ko lentelėje pateikti skaičiai gauti pagal $d \in \{1, 2, 3\}$ reikšmes.

Skalės efektų agregavimo modelius pagal rezultatus galima skirstyti į dvi grupes. *Tiesioginio* ir *Funkcinio* modelių rezultatai tarpusavyje panašūs, tačiau prastesni nei *MMM* ir *MIDAS* modelių. Iš pastarųjų dviejų geriau atrodo *MIDAS* modelis. Skaičiavimai patvirtina [5] gautą keistą faktą, kad g_5 generuojančio proceso parametrą β geriau vertina *MIDAS* nei *MMM* modelis, nors g_5 yra generuotas pagal *MMM* agregavimo funkciją.

T-MMM ir *M-MMM* modelių rezultatai yra geresni nei tikėtasi. $m = 3$ atveju *T-MMM* vidutinis rezultatas geriausias tarp visų modelių, o *M-MMM* modelio tik šiek tiek prastesnis. $m = 6$ atveju žymiai geriau nei visi kiti modeliai parametrą β vertina *M-MMM* modelis, o dėl parametrų skaičiaus *T-MMM* rezultatai nėra itin geri. Galima daryti išvadą, kad, nevertindami geriausiai nei periodo efektų agregavimo procesų g_1-g_4, nei skalės efektų agregavimo procesų g_5-g_8, bendrai šie du modeliai duoda geriausius rezultatus. Tokie rezultatai sutampa su modelių teorinėmis savybėmis, kada *T-MMM* ir *M-MMM* yra tiek periodo, tiek skalės agregavimo modeliai. Reikia pažymėti, kad *M-MMM* modelio periodo efektų svoriai čia aprašomi (2.15) formulėje pateikta hiperboline funkcija, kuri negali įgyti tokių įvairių formų

3 lentelė. Parametro β tikslumas su paklaidų įtaka, $100 \times \mathrm{MSE}(\hat{\beta}, \beta)$, $m = 22$ atvejis

$m = 22,\ n = 100,\ \sigma_\varepsilon = 0,5$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	**5,84**	24,6	6,61	**6,80**	7,95	10,2	**5,72**	**5,90**	10,9	7,44	9,20
Pask. reikšmė	91,6	**0,26**	100,3	82,8	92,4	92,1	91,6	92,1	68,7	92,1	80,4
Maksimali reikšmė	70,3	74,6	69,2	69,2	16,5	**0,98**	85,5	65,2	70,8	42,0	56,4
Tiesioginis per. ef.	7,43	7,24	7,05	7,43	10,0	14,3	7,39	7,43	**7,29**	9,77	8,53
MIDAS per. ef.	6,48	**6,73**	**5,45**	7,60	9,24	12,5	6,24	**5,90**	**6,57**	8,47	**7,52**
Funkcinis per. ef.	6,62	**6,68**	**6,05**	7,97	8,52	13,2	7,02	6,82	**6,83**	8,89	**7,86**
MMM	6,13	29,0	**6,18**	7,37	**4,78**	**1,48**	**5,36**	6,27	12,2	**4,47**	8,32
Tiesioginis sk. ef.	7,59	34,70	8,31	8,48	8,06	7,18	7,28	6,97	14,8	7,37	11,1
MIDAS sk. ef.	**5,89**	29,2	6,90	**7,11**	**3,89**	3,26	6,26	6,36	12,3	**4,94**	8,61
Funkcinis sk. ef.	**5,81**	30,2	6,55	**7,24**	6,83	6,14	**6,10**	**6,15**	12,45	**6,31**	9,38
T-MMM	15,6	7,63	12,1	9,80	7,28	**2,27**	15,6	16,6	11,3	10,4	10,9
M-MMM	6,18	8,87	7,03	7,90	**3,71**	8,38	6,25	6,90	7,50	**6,31**	**6,90**

kaip eksponentinė. Reikia tikėtis, jei svoriai būtų vertinami pastarąja funkcija (didesnis ją aprašančių parametrų skaičius, vertinant MKM, neblogina pagrindinių regresijos parametrų tikslumo), rezultatai būtų dar geresni.

1-oje ir 2-oje lentelėse pateikiami rezultatai, kai periodo ilgiai m nėra dideli. Verta modelius palyginti ir dideliems periodo ilgiams – tada, kai aukšto dažnio duomenų yra žymiai daugiau nei žemo dažnio. 3-oje lentelėje pateikti rezultatai, kai $m = 22$, $n = 100$. Toks periodo ilgis būdingas finansinėms laiko eilutėms, kada nagrinėjami duomenys yra mėnesiniai ir dieniniai, turint omenyje, kad per mėnesį būna maždaug 22 darbo, arba prekybos, dienos. Eksperimento sąlygos nesikeičia nuo ankstesniųjų, išskyrus tai, kad g_3 ir g_7 agregavimo funkcijose parinkta $\lambda_1 = 0,5$ bei $\lambda_2 = -0,03$, nes $m = 22$ atveju buvusios šių parametrų reikšmės beveik atitinka paskutinės reikšmės agregavimą. Siekiant sumažinti skaičiavimų laiką, replikacijų skaičius sumažintas iki 1 000.

3-oje lentelėje pateikti rezultatai keičiasi nedaug nuo gautų su mažais periodų ilgiais. *Tiesioginio* ir *T-MMM* modelių rezultatai dėl parametrų gausos žymiai suprastėja kitų modelių atžvilgiu, išskyrus tai, kad *Tiesioginis* periodų efektų modelis g_1-g_4 duomenis vertina vis dar pakankamai gerai. Funkcijos g_5 duomenis *MIDAS* skalės efektų modelis vėl vertina geriau nei *MMM*, o, kas keista, pagal g_7 generuotus duomenis *MIDAS* skalės efektų modelis nepatenka net tarp geriausių. Svarbiausia, *M-MMM* modelis išlieka kaip geriausiai parametrus vertinantis modelis.

3.3 Prognozių tikslumo eksperimentas

Koeficientų tikslumas galimai nepilnai apibūdina paties modelio tikslumą. Jeigu labiau yra rūpimas ne modelio koeficientų reikšmės, o gebėjimas kuo geriau aprašyti duomenis, naudinga tirti kitas modelio charakteristikas. Taip pat neteisingas koeficientų įvertinimas gali netgi klaidingai apibūdinti modelį. Pavyzdžiui, *MMM* yra duomenų vidurkį keičiantis agregavimas. Neteisingas (2.24) formule gautas p įvertis keičia duomenų vidurkį, t.y. didesnė p reikšmė suteiks daugiau svorio didesnėms aukšto dažnio reikšmėms periode. Dėl to automatiškai mažės (2.22) regresijos koeficiento β reikšmė, bet tai nebūtinai reikš mažesnę gautą regresijos liekanų dispersiją. Tas pats galioja ir atvirkčiai. Pavyzdžiui, duomenų vidurkio nekeičiantys periodo efektus aprašantys modeliai tik dėl vidurkio (ne dėl didesnės paklaidų dispersijos) parametro prasme blogiau aprašys pagal g_5 agreguotus duomenis, nes, esant $p = 12$, agreguotų duomenų vidurkis tampa didesnis. Dėl šių priežasčių reikalinga tokia analizė, kada modeliai būtų lyginami ne pagal įvertintų parametrų tikslumą. Paprasčiausia yra lyginti modelių R^2 koeficientus, tačiau tai nėra visai tinkamas matas dėl skirtingo modelių parametrizavimo. Tinkamas matas yra vieno žingsnio modelio prognozės tikslumas, kada tiek pernelyg maža, tiek pernelyg didelė gauta liekanų dispersija prognozes blogina.

Konkrečiau, gavus agreguotus duomenis, sudaroma regresija:

$$Y_t = \alpha + \sum_{j=1}^{d} \beta_j V_{t,d} + \varepsilon_t, \quad t = 1, \dots, n-1. \tag{3.4}$$

Įvertinus regresijos parametrus, gaunama Y_n prognozė:

$$\hat{Y}_n = \hat{\alpha} + \sum_{j=1}^{d} \hat{\beta}_j V_{n,d}, \tag{3.5}$$

ir suskaičiuojama prognozės kvadratinė paklaida $e_n = (\hat{Y}_n - Y_n)^2$. Tokia procedūra atliekama daug kartų su naujai generuotais duomenimis ir suskaičiuojamas prognozės kvadratinių paklaidų vidurkis. 4-oje lentelėje pateikiami gauti rezultatai su 1 000 Monte Carlo replikacijų. Sudaryti modeliai bei duomenis generuojantieji procesai išlieka tokie patys, išskyrus tai, kad (3.1) liekanų dispersija lygi 0,25. Iš lentelės gerai matyti rezultatų paklaida. Kadangi vieno žingsnio prognozės paklaidos dispersija susideda iš modelio liekanų dispersijos bei parametrų netikslumų dalies (ji visada teigiama), tai prognozės paklaidų dispersija negali būti mažesnė nei modelio paklaidų dispersija. Lentelės mažiausias skaičius yra 0,22. Tokia paklaida gana didelė, tačiau, jei yra rūpima bendra aštuonių duomenis generuojančių funkcijų gautoji suma, paklaida stipriai mažėja iki priimtino lygio.

Tradiciniai agregavimo metodai išlieka kaip gana prastai prognozuojantys modeliai, tačiau skirtumas nuo sudėtingesnių modelių nėra toks ryškus. Paskutinės reikšmės agregavimas $m = 22$ atveju lenkia paprasto vidurkio agregavimą, daugiausia dėl pastarojo blogų rezultatų g_2 funkcijai. $m = 22$ atveju paskutinės reikšmės agregavimo prognozės yra tikslesnės už kai kurių SDD modelių. Be šių išimčių, visgi pastarieji modeliai pranoksta tradicinius agregavimo metodus.

Lyginant 4-os lentelės rezultatus su 1-3 lentelių, didžiausias skirtumas yra panašių modelių prognozių suvienodėjimas. Periodo efektų ar skalės efektų modelių prognozės tarpusavyje labai panašios, ko nebuvo galima pasakyti vertinant parametro tikslumą. Pavyzdžiui, *Funkcinio* modelio parametro tikslumas buvo visada prastesnis nei *MIDAS* gautas, kas nepasakytina prognozavimo atveju. Baziniai *MIDAS*, *Funkcinis* ir *MMM* modeliai prognozuoja beveik taip pat tiksliai. *Laužtinės* modelio rezultatai taip pat neišsiskiria ir yra labai panašūs į *Funkcinio* agregavimo. $m = 3$ atveju *Laužtinės* modelis nepateiktas, nes tada jis yra analogiškas *Tiesioginei* regresijai, kada laisvai vertinami keturi regresijos parametrai.

4 lentelė. Modelių prognozių tikslumas, MSFE

$m=3,\ n=100,\ \sigma_\varepsilon=0,5$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	0,25	0,93	0,42	0,53	0,46	0,46	0,32	0,30	0,53	0,39	0,46
Pask. reikšmė	0,44	0,26	0,74	0,88	0,63	0,72	0,54	0,52	0,58	0,60	0,59
Maksimali reikšmė	0,39	1,04	0,57	0,70	0,28	0,25	0,54	0,47	0,68	0,39	0,53
Tiesioginis per. ef.	0,26	0,26	0,25	0,26	0,44	0,48	0,33	0,30	**0,26**	0,38	**0,32**
MIDAS per. ef.	0,27	0,25	0,28	0,55	0,44	0,46	0,29	0,32	0,34	0,38	0,36
Funkcinis per. ef.	0,25	0,25	0,27	0,26	0,45	0,47	0,33	0,29	**0,26**	0,39	**0,32**
MMM	0,22	0,98	0,41	0,56	0,29	0,26	0,26	0,29	0,54	0,27	0,41
Tiesioginis sk. ef.	0,26	0,99	0,44	0,55	0,26	0,26	0,26	0,25	0,56	**0,26**	0,41
MIDAS sk. ef.	0,26	0,99	0,42	0,54	0,24	0,24	0,25	0,30	0,55	**0,26**	0,40
Funkcinis sk. ef.	0,27	0,95	0,40	0,53	0,26	0,26	0,26	0,24	0,54	**0,25**	0,40
T-MMM	0,26	0,22	0,25	0,24	0,24	0,24	0,27	0,29	**0,24**	**0,26**	**0,25**
M-MMM	0,23	0,29	0,32	0,54	0,23	0,26	0,25	0,29	0,35	**0,26**	**0,30**
$m=6,\ n=100,\ \sigma_\varepsilon=0,5$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	0,26	1,08	0,49	0,41	0,43	0,48	0,36	0,29	0,56	0,39	0,48
Pask. reikšmė	0,40	0,26	0,66	0,48	0,56	0,63	0,49	0,42	0,45	0,53	0,49
Maksimali reikšmė	0,35	1,22	0,59	0,53	0,27	0,26	0,52	0,34	0,67	0,35	0,51
Tiesioginis per. ef.	0,26	0,28	0,27	0,27	0,46	0,57	0,35	0,29	**0,27**	0,42	0,34
MIDAS per. ef.	0,26	0,24	0,23	0,42	0,42	0,53	0,34	0,27	**0,29**	0,39	0,34
Funkcinis per. ef.	0,26	0,27	0,27	0,33	0,42	0,50	0,34	0,26	**0,28**	0,38	**0,33**
MMM	0,25	1,19	0,48	0,42	0,24	0,25	0,26	0,25	0,58	**0,25**	0,42
Tiesioginis sk. ef.	0,27	2,00	0,53	0,47	0,28	0,26	0,27	0,27	0,82	0,27	0,54
MIDAS sk. ef.	0,27	1,10	0,50	0,40	0,28	0,26	0,25	0,27	0,57	**0,26**	0,42
Funkcinis sk. ef.	0,24	1,13	0,48	0,43	0,26	0,26	0,26	0,27	0,57	**0,26**	0,42
T-MMM	0,27	0,25	0,27	0,28	0,26	0,28	0,26	0,30	**0,27**	0,28	**0,27**
M-MMM	0,24	0,40	0,26	0,39	0,24	0,27	0,26	0,26	0,32	**0,26**	**0,29**
Laužtinės per. ef.	0,26	0,27	0,26	0,33	0,43	0,51	0,35	0,29	**0,28**	0,39	0,34
Laužtinės sk. ef.	0,26	1,12	0,51	0,43	0,26	0,27	0,26	0,26	0,58	**0,26**	0,42
$m=22,\ n=100,\ \sigma_\varepsilon=0,5$	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1-g_4	g_5-g_8	g_1-g_8
Vidurkis	0,26	1,22	0,28	0,27	0,34	0,48	0,26	0,24	0,51	0,33	0,42
Pask. reikšmė	0,29	0,25	0,32	0,32	0,38	0,51	0,31	0,31	0,30	0,38	0,34
Maksimali reikšmė	0,29	1,37	0,3	0,33	0,28	0,24	0,32	0,3	0,57	0,29	0,43
Tiesioginis per. ef.	0,31	0,32	0,32	0,33	0,46	0,60	0,37	0,35	0,32	0,45	0,38
MIDAS per. ef.	0,27	0,33	0,25	0,30	0,33	0,39	0,25	0,25	**0,29**	0,31	**0,30**
Funkcinis per. ef.	0,26	0,26	0,24	0,27	0,34	0,47	0,24	0,25	**0,26**	0,33	**0,29**
MMM	0,26	1,19	0,29	0,31	0,26	0,24	0,26	0,26	0,51	**0,26**	0,38
Tiesioginis sk. ef.	0,30	1,72	0,37	0,37	0,33	0,35	0,33	0,34	0,69	0,34	0,51
MIDAS sk. ef.	0,26	1,21	0,27	0,29	0,28	0,26	0,24	0,25	0,51	**0,26**	0,38
Funkcinis sk. ef.	0,27	1,33	0,27	0,29	0,26	0,24	0,27	0,25	0,54	**0,26**	0,40
T-MMM	0,31	0,33	0,34	0,30	0,30	0,37	0,36	0,35	0,32	0,35	0,33
M-MMM	0,25	0,57	0,27	0,30	0,26	0,29	0,26	0,24	0,35	**0,26**	0,31
Laužtinės per. ef.	0,25	0,27	0,26	0,29	0,33	0,49	0,26	0,26	**0,27**	0,33	**0,30**
Laužtinės sk. ef.	0,26	1,22	0,28	0,30	0,26	0,26	0,26	0,25	0,52	**0,26**	0,39

T-MMM ir *M-MMM* modeliai nebėra tokie dominuojantys. Nors mažiems periodų ilgiams jų prognozės yra tiksliausios, tačiau $m = 22$ atveju *M-MMM* modelio prognozės susilygina su *Funkcinio* ir *MIDAS* periodo efektų modeliais. Išlieka pagrindinė šių modelių savybė: geriausiai nevertindami nei pagal periodo, nei pagal skalės efektus generuotų duomenų, bendrai prognozes galima laikyti tiksliausiomis.

Iš pirmo ir antro eksperimento galima daryti išvadą, kad SDD modeliai dažniausiai, ir kartais labai stipriai, pranoksta tradicinius paprasto *vidurkio*, *paskutinės* ar *maksimalios* reikšmės agregavimo metodus. Pastarieji yra „rizikingi", kai kuriems generuotiems procesams stipriai netinkantys. Praktikoje dažniausiai naudojamas paprasto vidurkio agregavimas, o po to pereinama prie modeliavimo. Net ir subtiliausias, plačiai taikomas ekonometrinis modelis gali neduoti laukiamų rezultatų vien dėl to, kad buvo neišnaudota aukšto dažnio duomenų dinamikos informacija. Tuomet paprastas SDD modelis galimai pranoktų sudėtingą tipinį ekonometrinį modelį. Kokį būtent SDD modelį naudoti, pasakyti sunku, nes iš imitacinės analizės rezultatų negalima nei vieno išskirti kaip visada geriausiai tinkančio, tačiau visada galima išbandyti kelis.

4 Empirinių taikymų rezultatai – kintamumo prognozavimas

Šios darbo dalies tikslas yra įvertinti 2-ame skyriuje aprašytų skirtingo dažnio duomenų modelių tinkamumą realiems duomenims, lyginant su tradiciniais ekonometriniais metodais. Taip pat modeliai palyginami tarpusavyje. Tyrimo problema yra finansinių laiko eilučių trumpo ir ilgo laikotarpio kintamumo prognozavimas skirtingo dažnio duomenų modeliais. Ilgo laikotarpio kintamumo prognoze laikysiu 30 ar 60 dienų prognozę, nepaisant kokio dažnio duomenys yra modeliuojami. SDD modelių kintamumo prognozės palyginamos tarpusavyje ir su kai kurių tradicinių vienodo dažnio duomenų modelių prognozėmis.

4.1 Duomenys ir jų žymėjimai

Nagrinėjami duomenys yra Niujorko akcijų biržos akcijų paketo indeksas (angl. *New York Stock Exchange composite index*), *Dow Jones* akcijų paketo indeksas (angl. *Dow Jones composite average*) ir *NASDAQ*[2] panašus atitikmuo (angl. *NASDAQ composite*). Šiuos tris indeksus arba jų logaritmines grąžas žymėsiu atitinkamai *NYSE, Dow Jones* ir *NASDAQ*. Duomenys yra dieniniai iki 2009 metų balandžio mėn., o *NYSE, Dow Jones* ir *NASDAQ* prasideda atitinkamai nuo 1966, 1981, 1971 metų, ir stebėjimų nm yra atitinkamai 10 895, 7 151, 9 644.

Bet kurį pradinį indeksą žymėkime P_t. Iš jo gaunamos dieninės logaritminės grąžos (toliau – grąžos) $r_t = \log(P_{t+1}/P_t)$, $t = 1, \ldots, nm$. Nesikertančios m-dienų grąžos $R^{(m)}$ gaunamos pagal formulę $R_j^{(m)} = \log(P_{jm+1}/P_{jm-m+1})$. *Stebimąja* m-dienų dispersija (angl. *realized variance*) $RV^{(m)}$ vadinsiu grąžų kvadratų sumą periode, t.y.

$$RV_j^{(m)} = \sum_{i=1}^{m} r^2_{(j-1)m+i}, \quad j = 1, \ldots, n. \tag{4.1}$$

Stebimoji dispersija yra nestebimo kintamumo atitikmuo, kuris naudojamas modeliavime prognozių paklaidoms suskaičiuoti. Žemo dažnio duomenis atitinka m-dienų grąžos, *stebimoji* dispersija ar kintamumo prognozė, o aukšto dažnio – dieninės grąžos ar jų kvadratai.

4.2 Tradiciniai prognozavimo metodai

Pirmiausia aprašysiu keletą paprastų vienodo dažnio duomenų finansinių grąžų dispersijos prognozavimui naudojamų modelių, kuriuos vadinsiu *tradiciniais*. Jų prognozių gavimas reikalingas norint įvertinti, kiek ir ar SDD modelių naudojimas padeda gauti tikslesnes prognozes.

[2]Visi duomenys gauti iš internetinio puslapio http://finance.yahoo.com

4.2.1 *Tiesioginis* prognozavimo metodas

Tiesioginiu (angl. *Direct*) prognozavimo metodu vadinsiu daugelio periodų dispersijos prognozės gavimo būdą, kai modeliuojami m-dienų grąžų duomenys, nesiremiant aukštesnio dažnio duomenų teikiama informacija. Sekant [4] straipsniu, *Tiesioginė* prognozė gali būti gauta *GARCH* šeimos modeliais, pasiūlytais [1], m-dienų nesikertančioms grąžoms. Prognozė gaunama kaip vieno žingsnio modelio prognozė.

$GARCH(1{,}1)$ m-dienų grąžoms turi tokį pavidalą:

$$\begin{aligned} R_t^{(m)} &= \Sigma_t^{(m)} \varepsilon_t^{(m)}, \quad \varepsilon_t^{(m)} \sim \mathbb{N}(0,1), \\ \left(\Sigma_t^{(m)}\right)^2 &= \alpha_0 + \alpha_1 R_{t-1}^{(m)2} + \beta_1 \left(\Sigma_{t-1}^{(m)}\right)^2, \end{aligned} \tag{4.2}$$

kur $\Sigma_t^{(m)2}$ yra vertinamas kintamumas, arba sąlyginė dispersija. Įvertinus parametrus Didžiausio tikėtinumo metodu naudojant informaciją iki laiko momento n, gaunama kintamumo (sąlyginės dispersijos) vieno žingsnio prognozė:

$$\left(\hat{\Sigma}_{n+1}^{(m)}\right)^2 = \hat{\alpha}_0 + \hat{\alpha}_1 R_n^{(m)2} + \hat{\beta}_1 \left(\hat{\Sigma}_n^{(m)}\right)^2. \tag{4.3}$$

Galimai tikslesnės prognozės gali būti gaunamos *ARMA-GARCH* šeimos modeliais. Čia pirmiausiai $ARMA(p,q)$ modeliu aprašomas sąlyginis grąžų vidurkis:

$$R_t^{(m)} = \alpha + \sum_{j=1}^{p} \beta_j R_{t-j}^{(m)} + \sum_{j=1}^{q} \gamma_j U_{t-j}^{(m)} + U_t^{(m)}, \tag{4.4}$$

o *GARCH* modelis taikomas įvertintoms paklaidoms $\hat{U}_t^{(m)}$.

4.2.2 *Iteracinis* prognozavimo metodas

Kintamumą *GARCH* modeliais galima prognozuoti ir naudojant aukščiausio dažnio grąžas. Sekant [4] straipsniu, *Iteraciniu* (angl. *Iterated*) prognozavimo metodu vadinsiu m-dienų ateities dispersijos prognozės gavimo būdą, kai *GARCH* modeliams naudojami vienodo ir aukščiausio turimo dažnio duomenys. Prognozė skaičiuojama m žingsnių į priekį, o galutinė m-dienų kintamumo prognozė, sudarius (4.2) modelį dieninėms grąžoms r_j ir kintamumui σ_j^2, gaunama pagal formulę:

$$\left(\hat{\Sigma}_{n+1}^{(m)}\right)^2 = \sum_{j=1}^{m} \hat{\sigma}_{nm+j}^2, \tag{4.5}$$

kur $\hat{\sigma}_{nm+j}^2$ yra j žingsnių į priekį (4.2) prognozė, parametrams įvertinti naudojant nm stebėjimų.

4.2.3 *Normavimo* prognozavimo metodas

Normavimo (angl. *Scaling*) prognozavimas yra pats paprasčiausias ir dažnai naudojamas ilgo laikotarpio kintamumo prognozėms gauti. Sudarius tokį patį modelį, kaip ir *Iteracinio* metodo atveju, gaunama vieno žingsnio kintamumo prognozė $\hat{\sigma}^2_{nm+1}$, o galutinė m-dienų kintamumo prognozė gaunama ją padauginus iš m:

$$\left(\hat{\Sigma}^{(m)}_{n+1}\right)^2 = m \times \hat{\sigma}^2_{nm+1}. \tag{4.6}$$

4.2.4 Prognozavimas *ARMA* modeliais

Naturalu grąžų kintamumą, kurio stebimas atitikmuo yra *stebimoji* dispersija, prognozuoti šio dydžio sudarytu *ARMA* šeimos modeliu (pagal [4]):

$$RV^{(m)}_t = \alpha + \sum_{j=1}^{p} \beta_j RV^{(m)}_{t-j} + \sum_{j=1}^{q} \gamma_j \varepsilon^{(m)}_{t-j} + \varepsilon^{(m)}_t, \tag{4.7}$$

kur p ir q gali būti parenkamas, pavyzdžiui, AIC informaciniu kriterijumi. Įvertinus modelio parametrus Didžiausio tikėtinumo metodu, gaunama m-dienų kintamumo vieno žingsnio prognozė:

$$\left(\hat{\Sigma}^{(m)}_{n+1}\right)^2 = \hat{\alpha} + \sum_{j=1}^{p} \hat{\beta}_j RV^{(m)}_{n+1-j} + \sum_{j=1}^{q} \hat{\gamma}_j \hat{\varepsilon}^{(m)}_{n+1-j}. \tag{4.8}$$

4.3 kintamumo prognozavimas skirtingo dažnio duomenų modeliais

SDD modelius taip pat galima naudoti ir kintamumo prognozavimui. Tariant, kad *stebimoji* m-dienų dispersija priklauso nuo praeities dieninių grąžų kvadratų, formuluojamas regresijos modelis:

$$RV_t = \alpha + \sum_{j=1}^{d} \beta_j V_{t-1,j} + \epsilon_t, \tag{4.9}$$

kur $V_t = g(r^2_{(t-1)m+1}, \dots, r^2_{tm})$, $t = 1, \dots, n$ yra kokia nors (m, d) agregavimo funkcija g gauti žemo dažnio duomenys. Agreguotų duomenų vėlavimų skaičių galima padidinti:

$$RV_t = \alpha + \sum_{j=1}^{d} \beta_{j,1} V_{t-1,j} + \ldots + \sum_{j=1}^{d} \beta_{j,J} V_{t-J,j} + \varepsilon_t. \tag{4.10}$$

Skyrelyje 2.3 aprašyta, kaip daugiau vėlavimų galima įtraukti kiekvienam nagrinėjamam modeliui. Įvertinus parametrus, skaičiuojama vieno žingsnio m-dienų kintamumo prognozė:

$$\left(\hat{\Sigma}^{(m)}_{n+1}\right)^2 = \hat{\alpha} + \sum_{j=1}^{d} \hat{\beta}_{j,1} V_{n,j} + \cdots + \sum_{j=1}^{d} \hat{\beta}_{j,J} V_{n-J+1,j}. \tag{4.11}$$

Prognozavimas SDD modeliais turi pranašumų. Pirma, kitaip nei *Tiesioginiu* prognozavimo metodo atveju, panaudojama informacija, esanti aukšto dažnio duomenyse. Antra, skaičiuojant m-dienų dispersijos prognozę skirtingiems m, modeliu visada prognozuojama tik vieną žingsnį į priekį, t.y. nepasikliaujama *Iteraciniu* prognozavimo metodu, kas, tikėtina, turėtų didinti prognozių tikslumą.

4.4 Prognozių tikslumo vertinimo kriterijus

Grąžų kintamumas nestebimas, todėl reikalingas jį atstojantis matas, kurį naudojant būtų galima lyginti prognozes. Pagal [4], toks matas gali būti *stebimoji* dispersija $RV^{(m)}$, kuri apibrėžta (4.1) formule. Turint kintamumo atitikmenį, galima gauti prognozės paklaidas:

$$e_t^{(m)} = \left(\hat{\Sigma}_{n+1}^{(m)}\right)^2 - RV_t^{(m)}. \tag{4.12}$$

Gavus prognozės paklaidas $e^{(m)} = (e_{t_1}^{(m)}, \ldots, e_{t_2}^{(m)})$, reikalingas tikslumo matas. Pagal [4], tikslumą galima matuoti kvadratinių prognozės paklaidų vidurkiu MSFE (angl. *mean squared forecast error*):

$$\text{MSFE}\left(e^{(m)}\right) = \frac{1}{t_1 - t_2 + 1} \sum_{t=t_1}^{t_2} \left(e_t^{(m)}\right)^2. \tag{4.13}$$

MSFE yra suderintas tikslumo matas ta prasme, kad prognozės paklaidoms skaičiuoti naudojamas tikro kintamumo atitikmuo asimptotiškai nekeičia modelių lyginimo rezultatų nuo tų, kuriuos gautume, jei stebėtume tikrą kintamumą.

4.5 Rezultatai

Aprašytais modeliais buvo suskaičiuotos kiekvieno indekso grąžų 5, 10, 20, 30 ir 60 dienų kintamumo prognozės. *ARMA* modelių eilė buvo parenkama pagal AIC informacinį kriterijų. *GARCH* modelių eilė pagal informacinius kriterijus parinkinėjama nebuvo, o rezultatai skaičiuoti su kelių tipų modeliais. Prognozės skaičiuotos pradedant maždaug nuo pusės kiekvieno indekso turimų duomenų, skirtų įvertinti parametrus, pridedant kas kartą po m dienų papildomų duomenų.

SDD modeliai su keletu agreguotų duomenų vėlavimais buvo sudaromi taip, kaip aprašyta 2.3 skyrelyje. *2.56* formulėje esantis skaičius J, parodantis, pagal kiek duomenų vykdomas agregavimas, buvo parenkamas toks, su kuriuo modelis tiksliausiai prognozuoja. J buvo parenkamas pagal iš anksto nustatytas galimas reikšmes. *Tiesioginiame* ir *T-MMM* modeliuose J galėjo įgyti reikšmes iš aibės $\{2, 3, 4, 5, 6, 7, 8\}$. Taigi m-dienų kintamumas čia aprašomas daugiausia pagal aštuonių paskutinių praeito periodo dienų duomenis. Visi kiti modeliai gali būti sudaryti įtraukiant tolimesnės praeities duomenis. Konkrečiai, priklausomai nuo periodo ilgio m, J galėjo įgyti tokias reikšmes:

- $m = 5: \quad J \in \{3, 5, 7, 10, 15, 20, 30\}$;

- $m = 10: \quad J \in \{3, 5, 7, 10, 15, 20, 30, 40\}$;
- $m = 20: \quad J \in \{5, 7, 10, 15, 20, 40, 60, 80\}$;
- $m = 30: \quad J \in \{5, 10, 20, 30, 45, 60, 90, 120\}$;
- $m = 60: \quad J \in \{5, 10, 20, 30, 40, 50, 60, 90, 120, 180, 240\}$.

Šios sąlygos buvo vienodos kiekvienam nagrinėjamam SDD modeliui.

Toks modelių sudarymo principas gali būti laikomas metodologiškai ne visiškai teisingu, nes J nėra vertinamas. Kitu būdu, skaičiuojant kiekvieną prognozę, J galėtų būti parenkamas pagal mažiausią informacinio kriterijaus reikšmę. Toks būdas yra teisingesnis, bet tada gaunami rezultatai yra žymiai prastesni. Natūralu, kad *stebimoji* dispersija labiau priklauso nuo artimesnės praeities kvadratinių grąžų. Daugumoje modelių vertinamų parametrų skaičius nepriklauso nuo J. Taigi modelių paklaidų dispersija dažniausiai yra mažesnė nedidelėms J reikšmėms, kurios tokiu atveju ir yra parenkamos. Tačiau tai nereiškia, kad tada yra gaunamos tikslesnės prognozės. Kelios praeito periodo dienų kvadratinių grąžų reikšmės, aprašydamos *stebimąją* dispersiją, mažiems J įgyja didelį svorį. Iš to išplaukia, kad kintamumo prognozė yra pernelyg jautri didesniems grąžų šokams, taigi prognozės dispersija gaunama labai didelė. Dideliems J prognozė yra „glotnesnė", ir kintamumo prognozavimo atveju tai lemia, kad ji yra tikslesnė. Dėl šios priežasties J parinkimas pagal informacinį kriterijų šiuo atveju nėra tinkamas, todėl reikia naudoti kitus būdus. Kita vertus, naudojamą metodologiją galima laikyti ir teisinga. Jei yra norima atlikti kintamumo ateities prognozę, praeities duomenis galima naudoti, kaip norima. Jei modelis parenkamas pagal geriausias praeities prognozes, tokiu atveju skaičių J galima tiesiog parinkti.

5-7 lentelėse pateiktas trijų indeksų kiekvieno modelio prognozių tikslumas. Kiekvienas skaičių stulpelis atitinka konkretaus m m-dienų kintamumo prognozės MSFE kriterijaus reikšmes. Paskutinis lentelių stulpelis yra tam tikras svertinis modelių visų prognozės ilgių vidurkis, pagal kurį galima lyginti bendrą modelių prognozės tikslumą. Kiekviename stulpelyje pajuodinti skaičiai reiškia tris (arba daugiau, jei trečioje vietoje yra keli modeliai su vienodai tiksliomis prognozėmis) geriausiai prognozuojančius modelius. *Tiesioginiam*, *Iteraciniam* ir *Normavimo* prognozavimo metodams atstovauja $AR(1)$- $GARCH(1,1)$ modeliai. Toks žymėjimas reiškia, kad $GARCH(1,1)$ taikomas po grąžų sąlyginio vidurkio aprašymo $AR(1)$ modeliu. Kitų bandytų šios šeimos modelių prognozės nebuvo tikslesnės.

Iš 5-os lentelės matyti, kad *Dow Jones* grąžų kintamumą geriausiai prognozuoja *Laužtinės* skalės efektų modelis, kurio visų laikotarpių prognozės yra itin tikslios. Tradicinio *Iteracinio* prognozavimo metodo rezultatai yra taip pat geri ir nenusileidžia daugumai SDD modelių. Kitų šiai klasei priklausančių ne tradicinio agregavimo modelių prognozės gana panašios, iš kurių prasčiausios yra gautos iš *Tiesioginės*, *T-MMM* ir *MMM* agregavimo funkcijų. Kita vertus, *M-MMM* yra vienas iš tiksliausias prognozes duodančių modelių. Reikia paminėti, kad modelio su *paprasto vidurkio* agregavimo funkcija prognozės yra nedaug blogesnės už geriausiąsias.

5 lentelė. *Dow Jones* grąžų kintamumo prognozių tikslumas, $10^6 \times$ MSFE

Pr.metodas	Modelis	5	10	20	30	60	Sv. vid.
Normavimo	AR(1)-GARCH(1,1)	0,427	1,439	5,024	11,320	25,37	5,489
Iteracinis	AR(1)-GARCH(1,1)	0,424	1,390	4,342	8,902	16,31	4,620
ARMA	ARMA(AIC)	0,446	1,473	4,313	8,500	23,30	4,985
Tiesioginis	AR(1)-GARCH(1,1)	0,516	1,505	4,483	8,444	28,15	5,412
Skirtingo duomenų dažnio	Paskutinė reikšmė	0,516	1,565	4,513	8,830	25,67	5,393
	Maksimali reikšmė	0,466	1,558	4,613	8,814	26,42	5,331
	Paprastas vidurkis	0,433	1,454	4,170	7,911	23,08	4,835
	Tiesioginis per. ef.	0,436	1,445	4,360	13,22	25,87	5,592
	Tiesioginis sk. ef.	0,447	1,475	4,070	7,920	22,01	4,811
	MIDAS sk. ef.	0,438	1,469	4,115	7,824	22,86	4,824
	Funkcinis per. ef.	0,444	1,488	4,41	8,089	21,82	4,903
	Funkcinis sk. ef.	0,433	1,414	4,042	7,483	23,29	4,740
	MMM	0,442	1,493	4,235	9,118	23,09	5,031
	T-MMM	0,436	1,445	4,158	13,471	21,52	5,380
	M-MMM	0,448	1,463	4,023	7,914	19,90	4,700
	Laužtinės per. ef.	0,440	1,441	4,359	7,483	17,45	4,588
	Laužtinės sk. ef.	0,425	1,342	3,905	7,182	15,73	4,272

6 lentelė. *NYSE* grąžų kintamumo prognozių tikslumas, $10^6 \times$ MSFE

Pr.metodas	Modelis	5	10	20	30	60	Sv. vid.
Normavimo	AR(1)-GARCH(1,1)	0,336	0,740	2,742	5,133	16,739	4,865
Iteracinis	AR(1)-GARCH(1,1)	0,299	0,735	2,571	4,630	16,350	4,559
ARMA	ARMA(AIC)	0,278	0,742	2,496	4,940	16,430	4,528
Tiesioginis	AR(1)-GARCH(1,1)	0,336	0,805	2,661	5,760	21,200	5,245
Skirtingo duomenų dažnio	Paskutinė reikšmė	0,336	0,805	2,661	5,760	21,200	5,245
	Maksimali reikšmė	0,300	0,867	2,794	5,703	17,660	5,074
	Paprastas vidurkis	0,281	0,744	2,495	4,867	15,440	4,480
	Tiesioginis per. ef.	0,294	0,778	2,890	7,421	33,070	6,021
	Tiesioginis sk. ef.	0,273	0,730	2,571	4,922	15,510	4,477
	MIDAS per. ef.	0,278	0,742	2,490	4,582	14,190	4,354
	MIDAS sk. ef.	0,280	0,734	2,396	4,808	15,310	4,409
	Funkcinis per. ef.	0,282	0,753	2,510	7,218	29,910	5,616
	Funkcinis sk. ef.	0,279	0,750	2,496	4,321	12,540	4,244
	MMM	0,281	0,744	2,496	4,985	19,240	4,684
	T-MMM	0,294	0,775	2,977	7,326	38,430	6,291
	M-MMM	0,278	0,742	2,446	4,582	14,110	4,334
	Laužtinės per. ef.	0,282	0,752	2,600	7,041	30,540	5,648
	Laužtinės sk. ef.	0,278	0,737	2,667	4,621	13,047	4,367

NYSE grąžų kintamumą SDD modeliai geriau prognozuoja nei tradiciniai. Iš pastarųjų geriausius rezultatus duoda *ARMA* ir *Iteracinis* metodai, gana tiksliai prognozuojantys visų laikotarpių kintamumą. Galima pastebėti, kad skalės efektų modelių rezultatai geresni nei periodo efektų, iš kurių visi, išskyrus *MIDAS*, 60-dienų kintamumą prognozuoja itin prastai. Labai gerai prognozuoja tiek bendrai, tiek ilgo laikotarpio kintamumą man labiausiai rūpimi *M-MMM* ir *Laužtinis* skalės efektų modeliai, tačiau jų išskirti kaip absoliučiai dominuojančių negalima.

NASDAQ atveju lyginamoji analizė pasikeičia nedaug. Keleto, tačiau ne visų, SDD modelių prognozės yra tikslesnės nei tradicinių modelių. Kaip geriausieji išskirtini *MIDAS*, *M-MMM* ir *Laužtinės* modeliai. *Paprasto vidurkio* agregavimas taip pat gana nedaug nusileidžia geriausiems SDD modeliams.

Iš trijų lentelių galima daryti išvadą, kad SDD geresniųjų modelių finansinių grąžų kinta-

7 lentelė. *NASDAQ* grąžų kintamumo prognozių tikslumas, $10^5 \times$ MSFE

Pr.metodas	Modelis	5	10	20	30	60	Sv. vid.
Normavimo	AR(1)-GARCH(1,1)	0,243	0,665	2,806	7,152	35,33	5,29
Iteracinis	AR(1)-GARCH(1,1)	0,258	0,702	2,982	7,127	24,65	5,01
ARMA	ARMA(AIC)	0,249	0,777	2,915	6,147	23,47	4,84
Tiesioginis	AR(1)-GARCH(1,1)	0,259	0,988	3,213	6,725	26,80	5,48
Skirtingo duomenų dažnio	Paskutinė reikšmė	0,302	0,766	3,584	6,364	32,33	5,65
	Maksimali reikšmė	0,277	0,929	3,357	7,108	24,25	5,47
	Paprastas vidurkis	0,253	0,785	2,910	6,650	21,90	4,88
	Tiesioginis per. ef.	0,314	0,808	3,220	6,59	28,75	5,51
	Tiesioginis sk. ef.	0,260	0,795	3,211	4,815	22,97	4,77
	MIDAS per. ef.	0,250	0,688	2,582	5,629	22,67	4,50
	MIDAS sk. ef.	0,261	0,793	2,930	6,358	20,52	4,83
	Funkcinis per. ef.	0,274	0,710	2,90	5,744	27,49	4,95
	Funkcinis sk. ef.	0,271	0,803	2,913	5,509	25,32	4,93
	MMM	0,257	0,787	2,939	6,837	23,13	4,99
	T-MMM	0,310	0,836	3,586	5,559	21,29	5,11
	M-MMM	0,258	0,689	2,581	5,632	22,72	4,54
	Laužtinės per. ef.	0,278	0,774	2,895	5,764	17,25	4,62
	Laužtinės sk. ef.	0,257	0,777	2,743	5,958	19,84	4,64

mumo prognozės yra tikslesnės už tradicinių prognozavimo metodų prognozes. Iš pastarųjų *Tiesioginio* ir *Normavimo* prognozavimo metodų rezultatai gerokai nusileidžia *Iteracinio* ir *ARMA* modelių prognozėms, kurias, tačiau, visada pranoksta bent keli SDD modeliai. Iš tradicinių agregavimo metodų visiems trims indeksams dominuoja *paprasto vidurkio* modelis, kuriuo gauti rezultatai nedaug prastesni nei geriausieji. Šis faktas gali sufleruoti, kad kintamumas gana vienodai priklauso nuo artimesnių ir tolimesnių praeities grąžų kvadratų. *Paskutinės* ir *Maksimalios* reikšmės agregavimo metodų prognozės visiems trims indeksams yra gana prastos.

Galima išskirti SDD modelius, kurie bendrai geriausiai prognozuoja visų indeksų grąžų kintamumą (tai pamatyti, patogu sudėti trijų lentelių paskutiniuosius stulpelius). Geriausius rezultatus rodo *Laužtinės* skalės efektų modelis, kuris ypač gerai prognozuoja ilgo laikotarpio kintamumą. Toliau iš eilės rikiuojasi *M-MMM*, *MIDAS* periodo efektų ir *Funkcinis* skalės efektų modeliai. Iš to išpaukia, kad *M-MMM* modelis, kaip *MMM* ir *MIDAS* agregavimo funkcijų praplėtimas, ir *Laužtinės* modelis, kaip *Funkcinio* agregavimo metodo praplėtimas, realiems duomenims „pasiteisino". Taip pat galima pastebėti, kad skalės efektų modeliai bendrai geriau prognozuoja nei periodo efektų atitikmenys (išskyrus *MIDAS* regresiją).

Pastebėtina, kad *Tiesioginio* periodo ir skalės efektų bei kartais *T-MMM* modelių prognozės yra gana tikslios. Dėl parametrų skaičiaus gausos šie modeliai aprašo m-dienų kintamumumo priklausomybę tik nuo kelių praeito periodo paskutinių dienų grąžų. Jeigu visada taip būtų gaunamos tikslios kintamumo prognozės, tai sufleruotų apie neilgą šio dydžio atmintį. Tačiau šie modeliai yra nepastovūs. Pirma, skirtingo laikotarpio prognozės gali visiškai skirtis – būti, pavyzdžiui, mažesniems m gana tikslios, didesniems m itin prastos. Antra, praeities grąžų dienų skaičiaus J nedidelis skirtumas gali visiškai pakeisti prognozes. Tarkime, modelis gerai prognozuoja, kai agreguojama pagal keturių dienų duomenis, tačiau prognozės kardinaliai pasikeičia, kai agreguojama pagal penkių dienų duomenis. Dėl šių priežasčių *Tiesioginės* ir *T-MMM* agregavimo funkcijų modeliai nėra patikimi.

Skirstyti modelius į gerai prognozuojančius tik trumpo ar tik ilgo laikotarpio kintamumą sunku. Tradicinių modelių ilgo laikotarpio kintamumo prognozės sąlyginai prastesnės nei trumpo. To paties negalima pasakyti apie visus SDD modelius. Čia tik *Funkcinio* periodo efektų, *Tiesioginio* ar *T-MMM* prognozės didesniems m prastesnės. Kitų modelių aiškios prognozių tikslumo priklausomybės nuo m nėra.

5 Išvados

Darbe atlikta iš įvairių agregavimo funkcijų formuojamų skirtingo dažnio duomenų (SDD) regresinių modelių palyginamoji analizė. Be literatūroje nagrinėtų, pristatomos ir trys iš jų išvestos naujos agregavimo funkcijos – *T-MMM*, *M-MMM* ir *Laužtinės*. Rezultatai gauti modeliuojant generuotus duomenis bei realius finansinių grąžų duomenis.

Imitacinės analizės dalyje modeliai pagal vertinamų parametrų ir prognozių tikslumą palyginami tarpusavyje, modeliuojant aštuonis duomenis generuojančius procesus. Iš gautų rezultatų galima daryti išvadą, kad tradicinės priešmodelinės paprasto vidurkio, paskutinės ar maksimalios reikšmės agregavimo funkcijos yra dažnai labai netikslios, todėl jų naudojimas yra gana rizikingas. Sudėtingesnių agregavimo funkcijų rezultatai gana panašūs – periodo efektų modeliai beveik vienodai gerai aprašo pagal periodo efektus generuotus duomenis ir prastai – pagal skalės efektus. Atvirkščia išvada galioja skalės efektų agregavimo funkcijoms. Nauja *M-MMM* agregavimo funkcija yra bene tinkamiausia naudoti, nes gana gerai aprašo tiek pagal periodo, tiek pagal skalės efektus generuotus procesus. Realūs duomenys gali būti generuoti itin sudėtingų procesų, dėl to ši *M-MMM* modelio savybė yra svarbi.

Empirinių taikymų dalyje lyginamas kelių tradicinių ir nagrinėjamų SDD modelių finansinių grąžų kintamumo prognozių tikslumas. Iš gautų rezultatų galima daryti išvadą, kad tradiciniais metodais gautos prognozės yra ne tokios tikslios kaip SDD modelių. Pastarieji taip pat pranoksta tradicinius paprasto vidurkio, paskutinės ar maksimalios reikšmės agregavimo būdus. Bendrai geriausius rezultatus duoda su naujomis *Laužtinės* skalės efektų ir *M-MMM* agregavimo funkcijomis formuojami modeliai, kuriems, tačiau, tik nežymiai nusileidžia *MIDAS* regresija ir *Funkcinis* būdas. Be to, galima daryti išvadą, kad skalės efektų modelių prognozės yra tikslesnės nei periodo efektų modelių. Minėti SDD modeliai apytiksliai viedonai tiksliai prognozuoja tiek artimos, tiek tolimos ateities kintamumą.

SDD modelių analizė toliau gali būti vystoma bent keliomis kryptimis. Pirma, galimas naujų agregavimo funkcijų kūrimas. Reikalingos *T-MMM* ir *M-MMM* funkcijų, kurios aukšto dažnio duomenis agreguoja tiek pagal periodo, tiek pagal skalės efektus, alternatyvos. Antra, darbe analizuojamos tik modelių kiekybinės charakteristikos – vertinamų koeficientų ir prognozių tikslumas. Galima ir platesnė modelių analizė, kada modeliai lyginami pagal kitus, nebūtinai kiekybinius kriterijus. Trečia, pageidautinas vartotojui „draugiško" programinio paketo sukūrimas, kuris leistų skirtingo dažnio duomenų modelius sudarinėti be specialaus pasiruošimo. Kita vertus, tam reikia konkrečių agregavimo metodų pripažinimo bei su parametrų vertinimu susijusių problemų efektyvių sprendimų.

Literatūra

[1] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 1986, t. 31, p. 307-327.

[2] E. Ghysels, P. Santa-Clara, A. Sinko, R. Valkanov. The MIDAS touch: mixed data sampling regression models, 2004, *working paper*, University of North Carolina.

[3] E. Ghysels, P. Santa-Clara, A. Sinko, R. Valkanov. MIDAS regressions: further results and new directions, *Econometric Reviews*, 2007, t. 26, p. 53-90.

[4] E. Ghysels, A. Rubia, R. Valkanov. Multi-period forecasts of volatility: direct, iterated, and mixed-data approaches, 2009, *working paper*, University of North Carolina.

[5] V. Kvedaras, A. Račkauskas. Regression models with variables of different frequencies: the case of a fixed frequency ratio, 2008, *working paper*, Vilniaus Universitetas.

[6] J.O. Ramsay, B.W. Silverman. *Functional data analysis*, 2nd ed., New York: Springer, 2005.