САНХҮҮ ЭДИЙН ЗАСГИЙН ИХ СУРГУУЛЬ
ӨДРИЙН ХӨТӨЛБӨР
ЭКОНОМИКСИЙН ТЭНХИМ

Тугчин
Отгонбаярын ЛХАГВАСҮРЭН

ЭДИЙН ЗАСГИЙН ТААМАГЛАЛД МАШИН СУРГАЛТЫН АРГЫГ ХЭРЭГЛЭХ НЬ

Мэргэжлийн индекс
031101

Эдийн засгийн ухааны бакалаврын
зэрэг горилсон
Дипломын төсөл

Удирдсан
Д.Хашбаатар/МА/

Улаанбаатар. 2020

САНХҮҮ ЭДИЙН ЗАСГИЙН ИХ СУРГУУЛЬ
ЭКОНОМИКСИЙН ТЭНХИМ

Тугчин
Отгонбаярын ЛХАГВАСҮРЭН

ЭДИЙН ЗАСГИЙН ТААМАГЛАЛД МАШИН СУРГАЛТЫН АРГЫГ ХЭРЭГЛЭХ НЬ

Мэргэжлийн индекс
031101

Эдийн засгийн ухааны бакалаврын зэрэг
Горилсон Дипломын Төсөл

Удирдагч: . . . . Д.Хашбаатар /MA/

Шүүмжлэгч: . . . . Д.Гансүлд /MBA/

Улаанбаатар. 2020

УДИРТГАЛ

Сэдвийн нэр: *Машин сургалтыг эдийн засгийн таамаглалд ашиглах нь*

Товч танилцуулга: Эдийн засгийн таамаглалд загвар сонголтын асуудал хүнд сорилтуудын нэг байсаар ирсэн юм. Үүнийг дагаад маш олон төрлийн судалгаа хийгдсэн байдаг ба энэ судалгааны ажлаар симуляцийн аргаар загваруудыг үнэлж, гүйцэтгэлийг харьцуулах болно. Ингэхдээ түгээмэл ашиглагддаг AR, GARCH, TAR загварууд дээр машин сургалтын баггинг аргазүйг ашигласнаар нарийвчлал болон таамаглах чадвар хэрхэн сайжирч байгааг авч үзэх болно. Гүйцэтгэлийн хэмжихдээ түүврийн гаднах (out of sample) таамаглалын алдааны квадратуудын дундажаар (MSE) хэмжинэ. Уламжлалт таамаглалын аргууд нь тухайн тохиолдолд сайн ажиллаж болох ч ерөнхий тохиолдолд сайн гүйцэтгэлтэй байдаггүй. Судалгааны үр дүнд баггинг аргазүйн ачаар тухайн нэг хугацааны цуваанд тохирохгүй загвараар таамагласан тохиолдлууд буюу таамаглалын загвараа буруу сонгосон тохиолдолд эерэг үр дүнг өгч байна. Баггинг буюу бүүтстрап агригэйшн хэмээх машин сургалтын алгоритм нь бусад сургалтын суурь болж өгдөг, мөн сүүлийн жилүүдэд эдийн засгийн таамаглалд ашиглах талаар судалгааны ажлууд гарч байгаа нь энэхүү судалгааны ач холбогдлыг харуулж байгаа юм.

Эдийн засгийн бүтээлийн сангийн индекс: *C01, C15*

Түлхүүр үг: *Хугацаан цуваан таамаглал, Загвар сонголт , Баггинг аргазүй, Монте Карло симуляци*

ГАРЧГИЙН ТОВЬЁОГ

ХҮСНЭГТЭН МЭДЭЭЛЛИЙН ЖАГСААЛТ

ЗУРГАН МЭДЭЭЛЛИЙН ЖАГСААЛТ

ХАВСРАЛТЫН ЖАГСААЛТ

ТОВЧИЛСОН ҮГИЙН ЖАГСААЛТ

ЭГ	Экспоненциал Гөлийлгөлт
ӨҮП	Өгөгдөл Үүсгэх Процесс
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MSE	Mean Squared Error
TAR	Threshold Autoregression
VECM	Vector Error Correction Model

ОРШИЛ

Хугацаан цувааны таамаглалд загварын тодорхой бус байдал голлох бэрхшээлүүдийн нэг байдаг. Тухайн түүврийн хувьд сайн тохирч байна гэдэг нь энэ загвар оновчтой таамаглал хийнэ гэсэн үг биш юм. Түүврийн хувьд сайн тохирч болох ч таамаглалын загварын гүйцэтгэлийг төлөөлөхгүй гэдэг нь илэрхий. Таамаглалын загварыг түүврийн бус аргад үндэслэн рекурсив эсвэл роллинг таамаглалуудын тохирсон байдалд түшиглэн сонгох нь түгээмэл байдаг. Үүний дараа түүврийг бүхэлд нь ашиглан параметрүүдийг тооцно. Энэхүү түүврийн бус аргазүйг Кларк (2004) болон Вэст (2006) нар судалсан бөгөөд өмнө дурдсан түүвэрт суурилсан загвар сонголтыг AIC мэдээллийн шинжүүрийн тусламжтай гүйцэтгэдэг билээ. Мөн түүврийг 2 хэсэгт хувааж загвар сонголт болон параметр үнэлэх үүргийг бие биеэсээ үл хамаарах байдлаар тооцох арга ч байдаг бөгөөд дээрх 2 аргазүйтэй харьцуулахад өргөн хэрэглэгддэггүй юм.

Бидний сургалтад өргөнөөр ашиглагддаг хугацааны цувааны загварууд нь урт хугацааны туршид хувьсан өөрчлөгдөж сайжирсаар ирсэн билээ. Тухайлбал, авторегрессив , хөдөлгөөнт дундаж, Вектор авторегрессив, Ерөнхийлсөн авторегрессив нөхцөлт хетероскедастик гэх мэт загваруудыг дурдаж болно. Эдгээр загвар нь тухайн нэг даалгаврыг сайн гүйцэтгэх боловч ерөнхий тохиолдолд хангалттай сайн үр дүнг өгдөггүй.

Шийдвэрийн модны аргазүй, бүүтстрап, санамсаргүй ойгүүжүүлэлт гэх мэт алгоритмуудыг өгөгдлийн шинжилгээ, загвар хөгжүүлэлтэд өргөнөөр ашиглаж байгаа билээ. Ялангуяа Брэймэний “Баггинг аргазүй” (Breiman, 1996) хэмээх бүтээлд дурдсан түүний техникийн гажуудлыг нэмэгдүүлэхгүйгээр таамаглалын хэлбэлзлийг бууруулах шинж чанар нь машин сургалтын давуу талыг харуулж байгаа юм. Иймдээ ч энэхүү алгоритм нь макро эдийн засгийн таамаглалд ашиглах талаарх судалгааны ажлууд бичигдэх болсон байна.

Монгол Улсын хэмжээнд энэ төрлийн машин сургалтын аргазүй нь хөгжиж байгаа боловч хараахан эдийн засагт бидний уламжлалт загвар шиг олонд танигдаагүй байгаа юм. Монгол банк гэх мэт улсын том хэмжээний институцүүд таамаглалдаа нэгтгэх олон төрлийн аргыг ашиглаж байгаа ч машин сургалтын аль нэг аргазүйг нэвтрүүлээгүй байгаа нь энэ чиглэлийн судалгаа шинжилгээний хоосон орон зай байгааг илэрхийлж байна.

Судалгааны зорилго, зорилтууд

Энэхүү судалгааны гол зорилго нь таамаглалын загваруудыг симуляцийн аргаар шинжлэн, харьцуулах билээ. Ингэхдээ гурван төрлийн өгөгдөл үүсгэх процессийг (ӨҮП) ашиглан симуляцийн цувааг үүсгэж Авторегрессив (AR) загвараар таамаглал хийсэн үр дүнг баггинг ашиглан таамаглалтай харьцуулж тодорхой үр дүнд хүрэх болно. Энэ зорилгыг ойгомжтой, дэс дараалалтай болгохын тулд дараах зорилгуудад хуваах болох юм. Үүнд:

- Сүүлийн 30 жилийн хугацааны цуваан таамаглалын загваруудын хөгжлийн түүхийг хураангуйлах, мөн баггинг аргазүйтэй холбоотой судалгааны ажлуудын үр дүнг дурдах
- Баггинг аргазүйг тоймлон аргазүйн талаас судлах
- Симуляцийн судалгааг төлөвлөх
- Монте-Карло симулцийн үр дүнг харуулах
- Шинжилгээний үр дүнд үндэслэн дүгнэлт санал боловсруулах зэрэг багтана.

Судалгааны объект

Энэхүү судалгааны ажил нь машин сургалтын баггинг аргазүйн гүйцэтгэлийг хэмжих зорилготой симуляцийн судалгаа билээ. Аливаа симуляцийн судалгаа нь судлаачийн үүсгэсэн өөрөөр хэлбэл ямар нэгэн хязгаарлалт дор үүсгэсэн хиймэл өгөгдөл дээр суурилдаг. Практикт хамгийн түгээмэл ашиглагддаг Авторегрессив загвар (AR), Theroshold Авторегрессив загвар (TAR) болон Ерөнхийлсөн Авторегрессив Нөхцөлт Хетероскедастик (GARCH) өгөгдөл үүсгэх процессуудыг ашиглах болно. ӨҮП бүрийн хувьд 1000 урттай, 100 ширхэг хугацааны цуваанд үндэслэн загварыг үнэлэх ба дараагийн 1, 6 болон 12 дахь утгыг таамаглах үүргийг гүйцэтгэх болно.

Судалгааны арга зүй

Монте-Карло симуляцийн арга нь тодорхой ӨҮП ийн дагуу санамсаргүй хувьсагчдыг олон удаа үүсгэн оруулж загварыг үнэлдэг. Санамсаргүй хэмжигдэхүүнүүд хосолсноор хугацааны цувааны маш олон хувилбарыг үүсгэх бөгөөд таамаглалын харьцуулалт баттай болох юм. Нарийвчлан авч үзвэл зохиомлоор үүсгэсэн хугацааны цуваан таамаглалд бүүтстрап дахин түүвэрлэлтийн хэмжээ, тухайн процессын үлдэгдэл санамсаргүй хэмжигдэхүүний вариац болон цувааны нэгж язгууртай ойрхон эсвэл цагаан шуугиан зэрэг шинж чанарууд таамаглалын гүйцэтгэлд хэрхэн нөлөөлж байгааг шинжлэх болно. Мөн 1, 6 болон 12 үеийн дараах таамаглалын MSE ийг оролцуулна.

Судалгааны ажлын практик ач холбогдол

Машин сургалтын загваруудын хөгжил салбар бүрт хүчээ авсан энэ цаг үед эдийн засгийн таамаглалын асуудлыг авч үзэх нь ач холбогдолтой юм. Симуляцийн аргын тусламжтай машин сургалтын аргуудын гүйцэтгэлийг харьцуулсан гадаадын хэд хэдэн судалгааны ажлууд сайн үр дүнг өгсөн байгаа нь Монгол Улсын хэмжээнд гүйцэтгэх ач холбогдолыг бий болгож байна. Онолыг эмпирикээр батлах, шалгах болон төрөл бүрийн арга техникийг симуляцийн аргаар турших нь өнөөгийн техникийн хөгжлийн хувьд боломжтой юм. Цаг хугацааг хэмнэн, асар их хэмжээний тооцооллыг богино хугацаанд гүйцэтгэх программ хангамжууд, тэдгээрийн гарц шийдлийг эзэмших нь 21 дүгээр зууны эдийн засагчдийн наад захын шаардлага болж хувирах цаг энүүхэнд ирсэн байна.

Энэхүү эрдэм шинжилгээний ажлын хүрээнд машин сургалт болон хугацааны цувааны огтлол дээр орших судалгааны чиглэлийг таньж мэдэхийг зорьсон юм. Ялангуяа бид өмнө дурдсан асуултууддаа хариулт олох юм. Баггинг алгоримтын тусламжтай таамаглалын тодорхой бус байдлыг шийдэж чадах уу?

Судалгааны ажлын бүтэц

Нэгдүгээр бүлэгт бид хугацааны цувааны загваруудын сүүлийн 30 жилийн түүх, хөгжлийн талаар маш товчхон авч үзэх болно. Бүх загваруудыг нарийвчлалай авч үзэх нь энэхүү ажилын гол зорилго биш тул Экспоненциал гөлийлгөлт (ЭГ) болон ARIMA загварыг сонгон авсан байгаа. Мөн баггинг аргазүйг эдийн засгийн таамаглалд ашигласан цөөн судалгааны ажлаар судлагдсан байдлаа өргөтгөн. Машин сургалтын эдийн засаг дахь хэрэглээний талаар товч дурдах болно.

Хоёдугаар бүлэгт баггинг аргазүйн үүсэл болон шийдвэрийн модны аргазүйн талаар дэлгэрэнгүй авч үзэх болно. Гуравдугаар бүлэгт Монте Карло симуляци хийх төлөвлөгөөг боловсруулж үр дүнг дөвөрдүгээр бүлэгт тайлбарлаж, улмаар дүгнэлт санал боловсруулах болно.

I БҮЛЭГ. СУДЛАГДСАН БАЙДАЛ

1.1 Хугацааны цувааны таамаглалын загваруудын товч түүх

1.1.1Экспоненциал гөлийлгөлт

Одоогоос дөчөөд жилийн өмнө экспоненциал гөлийлгөлтийн аргууд нь төрөл бүрийн нэг хэмжээст хугацааны цувааг гөлийлгөх албан ёсны арга хэрэгсэл хэмээгддэг байв. Экспоненциал гөлийлгөх аргыг бизнес, үйлдвэрлэлд өргөнөөр ашигладаг байсан боловч статистикчдын анхаарал татаагүйн улмаас сайн хөгжсөн статистикийн үндэс суургүй байжээ. Эдгээр арга нь 1950-1960-аад онд Браун (1959,1963) Холт (1957, 2004 онд дахин хэвлэгдсэн), Уинтерс (1960) нарын бүтээлүүдээс гарсан билээ. Пегелс (Pegels, 1969) нь хугацааны цуваag хандлага, улирлын нөлөөг шугаман эсвэл шугаман бус эсэхээс нь хамааруулж задлах энгийн боловч үр ашигтай аргазүйг гаргасан юм.

Экспоненциал гөлийлгөх аргууд нь 1985 онд хэвлэгдсэн хоёр бүтээлээс их түлхэц авсны улмаас энэ чиглэлээр дараагийн ажлуудыг хийх үндэс суурийг тавьсан юм. Нэгдүгээрт, Гарднер (1985) тухайн үеийн экспоненциал гөлийлгөх ажлын нарийвчилсан тоймыг хураангуйлж, синтезийг гаргасан бөгөөд бүдгэрсэн (damped) хандлагыг оруулахын тулд Пегелсийн ангиллыг сунгасан юм. Үүний дараа Снайдер (1985) ЭГ-ийг инновацийн төлөвийн загвараас (жишээ нь, алдааны нэг эх үүсвэр бүхий загвар) үүссэн гэж үзэж болно. Энэ ойлголт тухайн үед бараг үл мэдэгдэх байсан ч сүүлийн жилүүдэд экспоненциаль гөлийлгөх аргын дагуу төлөвийн загварууд дээр их хэмжээний ажил хийх үндэс суурийг тавьсан юм. (Gardner Jr., 1985)

1980 оноос хойш хийсэн ихэнх ажил нь энэхүү аргазүйн эмпирик шинж чанарыг судлах (жишээлбэл, Бартоломей ба Свийт, 1989; Макридакис ба Хибон, 1991), тооцоолох эсвэл шинэ аргачлалтай холих талаарх саналууд (Ледолтер & Абрахам, 1984), таамаглалын гүйцэтгэлийн үнэлгээ (McClain, 1988; Sweet & Wilson, 1988) гэх чиглэлээр хөгжсөн байдаг. Hyndman, Koehler, Snyder, and Grose (2002) таксономизм (Тейлор, 2003 онд сунгасан) аргуудыг тайлбарлахад тустай ангилал хийсэн байдаг. Арга тус бүр нь хандлагын 5 хэлбэртэй (ямар ч, нэмэлт, чийгтэй нэмэлт, үржүүлсэн болон чийгтэй үржүүлэгч) ба улирлын гурван хэлбэрээс (байхгүй, нэмэлт, үржүүлэгч) бүрдэж байв. Ингээд 15 төрлийн арга байдаг бөгөөд бидний хамгийн сайн мэддэгээр ЭЭГ (хандлагагүй, улирлын нөлөөгүй), Холтын шугаман аргазүй (шугаман хандлага, улирлын нөлөөгүй), Холт-Винтерийн шугаман арга (шугаман хандлага, шугаман улирлын нөлөө) болон Холт-Винтерийн шугаман бус аргазүй (шугаман хандлага, шугаман бус улирлын нөлөө) зэрэг болно. (Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D, 2005)

ЭГ-ийн аргазүйн таамаглалын үр дүнтэй аргазүйн талар цөөн хэдэн судалгаа байдаг. Сатчелл, Тиммерманн (1995) ба Чатфилд нар (2001) нь ЭЭГ нь өндөр далайцтай өгөгдөл үүсгэх процесст оновчтой болохыг харуулсан. Жижиг хэмжээний симуляцийн судалгаагаар Хиндман (2001) ЭЭГ нь нэгдүгээр эрэмбийн ARIMA загвараас илүү сайн гүйцэтгэлтэй болохыг олсон бөгөөд үүнийг өгөгдөл хэвийн бус үед загвар сонголтын асуудалд өртдөггүйтэй нь холбож тайлбарласан. (Satchell, S., & Timmermann, A, 1995)

ЭГ-ийн аргазүйн шүүмжлэлийн нэг нь таамаглалын интервал гарган авах боломжгүйтэй холбоотой. Энэхүү асуудлыг шийдэх аналитик арга нь тухайн хугацааны цуваа нь детерминистик бөгөөд цагаан шуугианыг агуулсан байна хэмээн таамаглах юм. (Браун, 1963; Гарднер, 1985; Маккензи, 1986; Свийт, 1985) Гэхдээ ийм байдалтай бол ЭГ-өөс илүү регрессийн загварыг хэрэглэх нь ашигтай юм. Иймээс Ньюболд ба Бос (1989) нар энэхүү таамаглал дээр үндэслэсэн бүх хандлагыг эрс шүүмжилж байв. Бусад судлаач ЭГ-тэй ижил төстэй статистикийн арга хэрэгслийг ашиглан таамаглалын интервалыг гаргаж авахыг илэрхийлж байв. Жонстон ба Харрисон (1986) хэд хэдэн эх сурвалж бүхий алдааны загваруудад энгийн болон Холтийн ЭГ-ийн аргуудын таамаглалын вариацыг олжээ. Яр ба Чатфилд (1990) Holt – Winters-ийн шугамн бус аргатай ижил ARIMA загварыг олж авах замаар таамаглалын интервалуудыг олж авсан байдаг. (Newbold, P., & Bos, T, 1989)

1.1.2 ARIMA загварууд

Хугацааны цувааг судлах эртний оролдлогууд, ялангуяа 19-р зууны үед детерминист ертөнцийн үзэл санаа ноёлж байв. Юлегийн (Yule, 1927) хугацааны цуваан эконометрикт оруулсан хамгийн том хувь нэмэр нь стохастик процесс гэсэн ойлголт бөгөөд аливаа бодит амьдрал дээрх хугацааны цуваа нь стохастик процессоор тайлбарлагддаг гэжээ. Энэхүү энгийн санааг үндэслэн хугацааны цуваан аргуудыг боловсруулсан юм. Слуцкий, Уокер, Яглом, Юле зэрэг судлаачид эхлээд авторегрессив (AR) дараа нь хөдөлгөөнт дундаж (MA) загваруудын тухай ойлголтыг боловсруулсан. Волдын задралын теорем нь Колмогоровын шугаман таамаглалын асуудлыг боловсруулж, шийдвэрлэхэд хүргэсэн юм (1941). Түүнээс хойш параметрийн үнэлгээ, ялгавар, загвар сонгох, таамаглалын асуудалтай холбоотой их хэмжээний судалгааны ажлууд гарч ирэв. (жишээ нь, Ньюболд (1983) -ийн судалгаанд). (Yule, 1927)

“Хугацааны цуваан шинжилгээ: Таамаглал ба удирдлага”: Бокс ба Женкинс (1970) нар энэхүү бүтээлээрээ тухайн үеийн мэдлэгийг нэгтгэсэн байдаг. Түүнээс гадна судлаачид хугацааны цувааг задлах, үнэлэх болон шалгах зорилготой мөчлөгийн уялдаатай, олон талт гурван үе шаттай аргазүйг хөгжүүлсэн юм. (Бидний хэлж заншсанаар Бокс-Женкинсийн арга) Энэхүү ном нь орчин үеийн хугацааны цуваан шинжилгээ, таамаглалын онолд үнэмлэхүй хувь нэмэр оруулсан юм. Компьютер техникийн ачаар шинжлэх ухааны олон салбарт авторегрессив нэгдсэн хөдөлгөөнт дундаж (ARIMA) загвар болон өргөтгөсөн хувилбаруудыг ашиглах болсон байна. (Box, G. E. P., & Jenkins, G. M., 1970)

Нэг хэмжээст. Бокс-Женкинсийн амжилтыг үндэслэн олон төрлийн зан авирыг дуурайж чадах , эцсийн сонгосон загварыг үнэлэхэд хэт их параметр шаардахааргүй байлаа. Гэсэн хэдий ч жараад оны дунд үед загвар сонголт нь судлаачдын хувьд бэрхшээлтэй асуудал байсаар ирсэн ба тухайн нэг загварыг онцлох тодорхой алгоритм байхгүй байв. Иймээс дараахан ARMA загварыг оновчлох олон төрлийн математик аргууд гарч ирсэн байдаг нь Акайкын мэдээллийн шалгуур (AIC), Акайкын эцсийн таамаглалын алдаа (АЭТА) болон Байес мэдээллийн шалгуур (BIC) зэрэг багтана. Эдгээр арга нь 1 алхамт таамаглалын алдааг хамгийн бага байлгахаар сонгогдох бөгөөд хэт үнэлгээ бий болсон тохиолдолд торгуулийн зарчимтай байдаг. Мөн мэдээллийн шинжүүрүүд хөндлөн баталгаажуулалт, түүврийн хуваалтын зарчим дээр тулгуурлах тохиолдол гардаг (Вэст, 1996), бас түүврийн гаднах таамаглалын алдааг авч үзсэн судалгааны ажил цөөнгүй. (West, 1996)

ARMA загварын параметрүүдийг тооцох олон төрлийн арга байдаг. Эдгээр арга нь асимптот утгаараа ижил, тооцоолол нь ижил хэвийн тархалттай байх хандлагатай ч төгсгөлөг түүврийн шинж хувьд том ялгаа бий. Программ хангамжийн багцын харьцуулсан судалгаагаар Ньюболд, Агиаклоглу, Миллер (1994) нар энэ ялгаа нь нэлээд их байх боломжтой бөгөөд үүний үр дүнд таамаглалын үр дүнд нөлөөлж болохыг олсон юм. Тэд хамгийн их үнэний хувь бүхий шалгуурын хэрэглэхийг санал болгож байсан юм. Мөн параметрийн үнэлгээний алдааны таамаглалын магадлалын хязгаарт нөлөөллийг Зеллнер (1971) ажигласан. Тэрээр Байесийн шинжилгээг ашиглаж, ARMA загварт орсон параметрүүдийг санамсаргүй хувьсагч болгон харьцуулж ирээдүйн ажиглалтын таамаглалын тархалтыг гаргаж авсан. Ким (2003) жижиг түүвэр дэх АР загварын параметрийн үнэлгээ, таамаглалыг авч үзсэн билээ. Үр дүнд нь (бүүтстрап) алдаа засварласан параметрийн тооцоо нь хамгийн бага квадрат тооцооллоос илүү нарийвчлалтай таамаглал дэвшүүлдэг хэмээжээ. Лэндмэн, Дамодаран (1989) нар Жеймс-Стейний ARIMA-ийн параметрийн тооцоог MSE-ийн алдагдлын шалгуурын дагуу бусад аргаас харьцангуй нарийвчлалтай байгааг нотолсон болно. (Newbold, P., Agiakloglou, C., & Miller, J., 1994)

ARIMA-гийн автоматжуулсан загварчлал нь ч нэг алхмын таамаглалыг үнэн бодитой гаргаж өгдөг. (Hill, G., & Fildes, R, 1984) Программ хангамжийн хэд хэдэн борлуулагчид таамаглалыг тооцоолох автоматжуулсан (олон хэмжээстийг багтаасан) аргыг боловсруулсан байдаг. Ихэнхдээ эдгээр арга нь хар хайрцаг шиг ажилладаг тул таамаглалын алгоритмыг сонгох талаарх зарим удирдамжийг Чатфилд (1988) өгсөн байна.

Олон хэмжээст. Вектор ARIMA (BARIMA) загвар нь нэг хэмжээст ARIMA загварын олон хэмжээст ерөнхий хэлбэр юм. VARMA процессын эх олонлогийн шинж чанарын тухай анх Квинуил (1957) судалсан боловч тооцооллын хувьд 1980,1990 -ээд оноос л техникийн баталгаатай болсон юм. VARIMA загвар нь экзоген чанарын тухай таамаглал авч үздэг ба богино хугацааны хамаарлыг салгаснаар таамаглагчид болон бодлого боловсруулагчдад шинэ даалгавар бий болгосон байна. Riise and Tjøstheim (1984) нар VARMA-ийн таамаглалд параметрийн үнэлгээний тооцоолол хэрхэн нөлөөлж болохыг авч үзсэн бол Cholette and Lamy (1986) шүүлтүүрийг VARMA загварт хэрхэн ашиглаж болохыг харуулсан байдаг. (Riise, T., & Tjøstheim, D, 1984)

Ерөнхийдөө VAR ач холбогдолгүй маш олон хувьсагчдын улмаас хэт үнэлэх хандлагатай байдаг. Үр дүнд нь загвар түүврийн хувьд сайн тохирч болох ч түүврээс гаднах таамаглал нь муу гүйцэтгэлтэй болдог. Liu, Gerlow, and Irwin (1994) болон Simkins (1995) Зарим параметрүүдийг ердийн байдлаар хязгаарлахын оронд Литтерман (1986) болон бусад нь параметрүүд дээр тархалтын приор хийснээр эдийн засгийн олон хувьсагчид санамсаргүй алхаатай байдаг хэмээн илэрхийлж байв. BVAR загварыг макро эдийн засгийн таамаглалд (Artis & Zhang, 1990; Ashley, 1988; Holden & Broomhead, 1990; Kunst & Neusser, 1986), хөрөнгийн зах зээлийн таамаглахад (Рибейро Рамос, 2003), хөдөлмөрийн зах зээлийн таамаглалд (LeSage) ашигласан юм. & Магура, 1991), бизнесийг урьдчилан таамаглахад түлхүү ашигладаг. (Спенсер, 1993) Клинг ба Бесслер (1985) Litterman-ийн BVAR загварыг багтаасан олон төрлийн хугацаан цувааны олон хэмжээст аргуудыг түүврийн бус таамаглалаар харьцуулсан байдаг. (Liu, T. -R., Gerlow, M. E., & Irwin, S. H., 1994)

Энгле ба Гранжерийн (1987) коинтегрэшнйн үзэл баримтлал нь хязгаарлалтгүй VAR болон BVAR дээр алдаа засах загвар (VEC) -ийн таамаглах чадвартай холбоотой янз бүрийн сонирхолтой асуултуудыг үүсгэсэн юм. Shoesmith (1992,1995), Tegene and Kuchler (1994), Wang and Bessler (2004) нар VEC нь VAR-аас илүү гүйцэтгэлтэй байгааг эмпирикээр нотолсон бөгөөд ялангуяа урт хугацааны таамаглалд ялгаа их байгааг онцолсон байдаг. Shoesmith (1995), дараа нь Виллани (2001) нар мөн Литтерман (1986) Байесийн арга нь коинтегрэшнтай VAR-ын таамаглалыг хэрхэн сайжруулж болохыг харуулсан билээ. (Engle, R. F., & Granger, C. W. J, 1987)

1.1.3 Улирлын нөлөө

Улирлын нөлөөг загварчлан хамгийн хуучин арга нь хугацааны цуваг түүний улирлын бүрэлдэхүүн хэсгээр нь задлах процедур X-11 юм. Нэг хэсэг судалгаа нь улирлын нөлөөг загварлах аргуудын таамаглалд үзүүлэх нөлөөг судалж байв. Миллер ба Уильямс (2003, 2004) нь улирлын бүрэлдэхүүн хэсгийг тэг болгон засварлах замаар таамаглалын илүү нарийвчлалтай үр дүнд хүрч болохыг олж тогтоосон байна. X-11 арга, түүний хувилбарууд дээр хөгжүүлэлт хийсний эцэст тохируулгын хэд хэдэн шинэ аргыг боловсруулсан бөгөөд эдгээрийн хамгийн чухал нь TRAMO-SEATS (Go´mez & Maravall, 2001; Kaiser & Maravall , 2005) ба параметрийн бус арга STL (Кливленд, Кливленд, МакРэй, & Терпеннинг, 1990).

Зарим судлаачид стандарт нэгж язгуурт улирлын нөлөөг засварлах загваруудын өргөн хэрэглээг таатай үзэгдэл биш хэмээн үзэж байв. Осборн (1990) үзэхдээ эдийн засгийн хугацааны цуваа нь стохастик улирлын нөлөөтэй гэхээс илүү детерминистик улирын нөлөөтэй байдаг гэжээ. Франсес, Ромижн (1993) нар улирлын язгуур бүхий үечилсэн загвар таамаглалын гүйцэтгэлд сайнаар нөлөөлдөг тухай санал дэвшүүлсэн юм. Улирлын нөлөөний ялгаатай аргуудыг харьцуулсан эмпирик ажлуудын үр дүнд хамгийн сайн гүйцэтгэлтэй загвар нь өгөгдлийн мөн чанараас хамааран харилцан адилгүй байдаг хэмээн гарчээ. (Osborn, 1990)

1.2 Баггинг аргазүйг эдийн засгийн таамаглалд ашигласан судалгааны ажлууд

Баггинг нь таамаглалын загварын тодорхой бус байдалд тухайн таамаглалын нарийвчлалыг сайжруулахад зориулагдсан статистик аргазүй юм. (Breiman, 1996) Баггинг хэмээх үг нь бүүтстрап аггригэйшнт үгний товчлол билээ. Чухамдаа баггинг нь бүх боломжит таамаглагчдыг агуулсан олон тооны бүүтстрап түүвэр үүсгэж таамаглал хийж, үр дүнг дундажлах алгоритм юм. (Breiman, 1996)

Сүүлийн үед баггинг нь макро эдийн засгийн шинжилгээ, таамаглалд өргөнөөр ашиглагддаг болсон билээ. Панагиотелис, Афанасопулос, Хиндман, Цзян, Вахид (2019) нар Австралийн макро эдийн засгийн өгөгдөл дээр маш олон тооны хувьсагчийг хамруулан таамаглал хийж энэхүү аргазүйн гүйцэтгэлийг үнэлсэн байдаг. Нарийвчилж авч үзвэл тэд баггинг хийсэн LARS-ийг ДНБ-ий өсөлт, ХҮИ-ийн инфляц, IBR (АНУ дахь

Холбооны сангийн ханштай дүйцэх банк хоорондын бэлэн мөнгөний ханш) -ын динамик хүчин зүйлийн загвар, Ridge регресс, LARS, Бэйсийн VAR зэрэгтэй харьцуулсан юм. Үр дүнд нь баггинг арга нь илүү нарийвчлалтай таамаглахад тус болно гэдгийг олж тогтоожээ. (Petropoulos, F., Hyndman, R. J., & Bergmeir, C., 2018)

Хирано ба Райт (2017) нар нь таамаглагч хувьсагчдын сонголтын талаар тодорхой бус үед Рао Блэквеллийн теорем ба Багингын дагуу таамаглалын загвараа сонгож гүйцэтгэлийг харьцуулах судалгаа хийсэн юм. Тэд загвар сонголт, параметрийн тооцоо хийхэд зориулж янз бүрийн схемүүдийн тархалтын шинж чанарыг судалж үзсэн: Akaike мэдээллийн шалгуурыг ашиглан түүвэрт тулгуурлан загвар сонгох, түүврийн бус өгөгдөлд тулгуурлан загвар сонгох, өгөгдлийг загвар сонголт болон параметрийн тооцоонд зориулж дэд хэсгүүд болгон хуваах аргууд. Тэд баггинг нь суурь асимптотик эрсдэлд хэрхэн нөлөөлж байгааг болон тэдгээрийн холбогдох таамаглалыг судалжээ. Эмпирик судалгаандаа тэд суурь параметрийн олон утгын хувьд уламжлалт аргаар хэрэгжүүлсэн тохиолдолд түүврийн бус болон хуваасан түүврийн схемүүд таарууxан ажилласан болохыг тогтоожээ. Гэхдээ Рао-Блэквелл теорем болон баггингын дагуу загвар сонгох аргуудтай хослуулсан үед үлэмж сайжирч байсныг онцолсон байдаг. (Hirano, K., & Wright, J. H, 2017)

Өмнө дурдсанчлан баггинг алгоритм нь таамаглалын тогтвортой байдалд үр ашигтайгаар нөлөөлдөг билээ. Ангиллын мод, регрессийн мод, шугаман ба шугаман бус хувьсагч сонголтыг ашигласан эмпирик болон онолын судалгаануудад энэ аргазүйн таамаглалд авчрах ерөнхий нөлөөг онцолсон байдаг. Гэвч баггингийн талаарх одоогийн ихэнх судалгаанууд нь хөндлөн өгөгдлийн эконометрикээр хязгаарлагдаж байна. Лий болон Янг (2006) нар баггинг алгоритмийн хэрэглээг хугацааны цуваа, бүр цаашилбал чанарын хувьсагч, квантилыг таамаглахад ашиглаж өргөтгөсөн байна. Чанарын таамаглал хийхийн тулд квантил тооцоолол, давамгай санал бүхий бинар таамаглалыг ашигласан бөгөөд ийм тохиолдолд баггинг аргазүй сайн гүйцэтгэлтэй байх боломжтойг харуулсан байна. Эмпирик хэрэглээний хувьд тэд сарын давтамжтай S&P500 болон NASDAQ хувьцааны индексийн өгөөжийг ашиглан үр дүнг танилцуулсан юм. (Lee, T.-H., & Yang, Y, 2006)

Иноуе, Килиан (2008) нар АНУ-ын ХҮИ-ийн хугацаан цуваан таамаглал хийхдээ багинг аргазүйг авч үзсэн байна. Тэд инфляцын таамаглахдаа динамик шугаман регрессд баггинг тохирч болох талаар судалж үзсэн юм. Ингэхдээ хэд хэдэн загваруудыг баггингтай болон баггинггүй харьцуулж үзсэн бөгөөд үүнд регрессийн загварууд, фактор загвар болон агшаасан регрессийн загварууд (ЛАССО-той) хамаарна. Тэдний эмпирик үр дүн нь баггинг нь инфляцыг урьдчилан таамаглах бэрхшээлтэй зориулалтад ч гэсэн таамаглалын алдааны квадратыг их хэмжээгээр бууруулж чаддаг болохыг харуулж байна. (Inoue, A., & Kilian, L, 2008)

Ли, Ту, Уллах (2014, 2015) ба Хиллебранд, Ли, Меейросоос (2014) үзүүлэлтүүдийн функцийг ашиглан янз бүрийн хатуу дүрэм хязгаарлалтад бүхий санхүүгийн өгөөжийг параметрийн, параметрийн бус болон хагас параметрийн регрессийн таамаглалын загваруудыг авч үзсэн болно. Судалгааны зорилго нь эдийн засгийн онол болон регрессийн функцээс гарч ирэх чухал эерэг эсвэл монотоник шинж чанартай эдийн засгийн олон төрлийн хязгаарлалт нэгтгэх байсан. Тэд хатуу хязгаарлалт бүхий үнэлгээний вариацыг бууруулахын тулд баггинг аргазүйг ашигласан юм. Тэд баггинг аргазүйг хэрэглэсэн хязгаарлалтын үнэлгээний асимптот чанарыг тооцож таамаглах ач холбогдлыг харуулсан байна. Баггинг хязгаарлалттай үнэлэгчид болон таамаглалын давуу талыг Монте-Карлогийн өргөн цар хүрээтэй симуляциар харуулсан болно. Санхүүгийн хөрөнгийн хураамжийг урьдчилан таамаглах эмпирик судалгаанд баггинг аргазүйг ашигласнаар урт хугацааны таамаглалын алдаа гаргах боломжийг багасгаж, хязгаарлалт бүхий таамаглалын үнэлгээг илүү тогтвортой болгож байсан юм. (Lee, T.-H., Tu, Y., & Ullah, A, 2015)

Жин, Су, Уллах (2014) түүврийн бус таамаглал хийхдээ баггинг аргазүйн хувиргасан хэлбэр болох загваруудыг нэгтгэх аргыг хэрэглэсэн юм. Хувиргасан хувилбар нь хугацааны цувааны хамаарлыг нарийвчлан үздэг бөгөөд таамаглалын алдааны квадратуудын дундаж утгыг баггинг хийгдээгүй таамаглалтай харьцуулан үзсэн юм. Тэдний Монте Карло симуляци нь энэхүү арга уламжлалт нэг алхамт шугаман таамаглал болон параметрийн бус таамаглалд ерөнхийдөө сайн гүйцэтгэлтэй, түүврийн хэмжээ бага байхад ч гайхалтай үр дүнтэй байгааг харуулж байна. Мөн тэд загвараа буруу тодорхойлсон ч баггинг аргазүйг ашигласан шугаман загвараар хийсэн таамаглал нь параметрийн бус таамаглалаас харьцангуй үр ашигтай байгааг ажигласан бөгөөд түүврийн бус таамаглал хийхдээ баггинг аргазүйг ашиглаж илүү тогтвортой үр дүнд хүрэх боломжтой санал болгосон юм. Тэд дараа нь судлагдсан байдалд санал болгосон хувьсагчдын хэтийн өгөөж буюу хөрөнгийн хураамжийг урьдчилан таамаглах хүчийг дахин судалж, Уэлч ба Гойлл (2008) -тай нийцсэн түүхэн илүүдэл хувьцааны өгөөжийн урьдчилсан тооцоо нь тухайн судалгаан дахь таамаглагч бусад хувьсагчийг гүйцэтгэлээрээ давж байгааг олж мэдсэн юм. Уламжлалт нэг шугаман урьдчилсан болон параметрийн бус таамаглалын аргыг хэрэглэн энэхүү үр дүнд хүрсэн байна. (Jin, S., Su, L., & Ullah, A, 2014)

Ардуино ба Медерос (2011) гөлгөр шилжилтийн мод хэмээх шинэ аргыг санал болгосон юм. Тэд инфляц ба бодит үйл ажиллагааны түрүүлэх үзүүлэлтүүд нь олон бүтцийн өөрчлөлт бүхий хамгийн чухал таамаглагчид болохыг тогтоожээ. Мөн тэд баггинг ашиглахтай холбогдуулан хоёр нөхцөлт моментийг урьдчилан таамаглах эмпирик нотолгоо өгсөн юм. (Audrino, F., & Medeiros, M. C, 2011)

1.3 Машин сургалтын алгоритм түүний хэрэглээ

1.3.1Шинэ өгөгдөл

"Том өгөгдөл" хэмээх хэллэг нь мэдээллийн цар хүрээ хурдацтай өөрчлөгдөж байгааг харуулж байна. Мөн өгөгдлийг шинж чанарт ч гэсэн өөрчлөлт гарсан байна. Машин сургалт нь стандарт тооцооллын аргуудад хэт их хэмжээст уламжлалт бус өгөгдлүүд, түүний дотор зураг, хэлний мэдээлэл зэрэг өгөгдлийг бид регресст оруулаад авч үзэх боломжтой болгодог.

Хиймэл дагуулууд дэлхийн зургийг хэдэн арван жилийн турш авч ирсэн бөгөөд одоо бид үүнийг пикселжсэн вектор төдийгүй эдийн засгийн хувьд ч ач холбогдолтой орц болгон ашиглаж болно. Дональдсон ба Стэйтегард (2016) хиймэл дагуулын мэдээллийг ашиглаж буй эдийн засгийн судлагдсан байдлын тоймд хувь нэмэр оруулсан байдаг. Тухайлбал, шөнийн гэрэлтүүлэг болон эдийн засгийн гарцын хамаарал (Хендерсон, Storeygard, Weil 2012) эсвэл ирээдүйн ургацын хэмжээг тооцоолох (Lobell 2013) судалгаануудыг дурдаж болно. Хиймэл дагуулын зураг нь бидэнд шууд ургацын хэмжээ гэх мэт мэдээллийг өгөхгүй ч зурганд суурилсан x вектор хэлбэртэй өгөгдлийг гаргана. Энэ өгөгдлийг $\hat{y}$ -ээр илэрхийлэгдсэн ургацын хэмжээний утгатай тааруулах болно. Хиймэл дагуулын зургийг ургацын хэмжигдэхүүн рүү шилжүүлэх нь таамаглалын асуудал юм. Чухамдаа машин сургалт нь энэхүү өгөгдлөөс эдийн засгийн ач холбогдолтой дохиолол гаргаж авах, задлах энгийн хэрэгсэл юм. (Lobell, David B, 2013)

Эдгээр шинэ мэдээллийн эх үүсвэр нь ялангуяа хөгжиж буй орнуудын ядуурлыг хянах, зорилтот чиглэлээр тогтоох гэх мэт эдийн засгийн мэдээлэл орхигдсон үед тохиолдолд онцгой ач холбогдолтой юм. (Blumenstock, Joshua Evan, 2016) Жан (2016) Африкийн таван орны хиймэл дагуулын мэдээллээс орон нутгийн эдийн засгийн үр дүнг урьдчилан таамаглах неирол сүлжээг сургасан байдаг. Машин сургах нь томоохон хэмжээний сүлжээний өгөгдлөөс эдийн засгийн таамаглал дэвшүүлэхэд ашиглагддаг. Жишээлбэл, Блуменсток, Кадамуро, Он (2015) нар гар утасны өгөгдлийг ашиглан Руанда дахь ядуурлыг хувь хүн бүрийн түвшинд үнэлсэн байдаг. Зурган мэдээллийг таних нь хиймэл дагуулын өгөгдлөөс гадна хөгжиж буй орнуудын эдийн засгийн гарцын таамаглалыг хийхэд ашиглагддаг. Нэг жишээ дурдахад Глаесер, Коминерс, Люка, Наик нар (2016) Google Street View-ээс авсан зургуудыг ашиглан Нью-Йорк болон Бостон дахь түвшний орлогыг хэмжсэн байна. (Blumenstock, Joshua E., Gabriel Cadamuro, 2015)

Хэл нь мэдээллийн өөр нэг шинэ хүчирхэг эх сурважийн нэг билээ. Хиймэл дагуулын зургийн нэгэн адил цахим постуудыг машин сургалтын тусламжтай шошгожуулж хэрэгцээтэй өгөгдөл гарган авах боломжтой юм. Кан, Кузнецова, Лука, Чой нар (2013) эрүүл ахуйн хяналт шалгалтын үр дүнг урьдчилан таамаглахын тулд Yelp.com дээрх ресторан бүрийн танилцуулгыг ашигласан юм. (Kang, Jun Seok, Polina Kuznetsova, Michael, 2013)

Санхүүчид болон эдийн засагчид Компюстат (дэлхийн хэмжээнд идэвхтэй, идэвхгүй байгаа компаниудын санхүүгийн, статистик, зах зээлийн мэдээллийн мэдээллийн сан) дээрх байгууллагын санхүүгийн мэдээлэлд ихээхэн найддаг. Гэхдээ байгууллагууд санхүүгийн байдлын дэлгэрэнгүй тайланг гаргадаг. АНУ-д олон нийтэд зарагддаг компаниуд жил бүр 10-K маягтыг гаргаж өгөх ёстой. Коган, Левин, Роутледж, Саги, Смит нар (2009) эдгээр хэлбэрийн зах зээлийн эрсдэлийн тодруулгын текстээс ойролцоогоор 10,000 ийм фирмүүдийн хэлбэлзлийг урьдчилан таамагласан бөгөөд энэ нь өнгөрсөн үеийн хэлбэлзэлд мэдэгдэхүйц ач холбогдолтой байгааг харуулсан байна.

Машины сургалт нь уламжлалт мэдээллийн санд боловсруулалт хийхэд ихээхэн ашиг тустай. Фейгенбаум (2015a, b) нь хувь хүний түүхэн мэдээлэлд тулгуурлан зорилтод хүмүүсийн уулзуулдаг машин сургалтын ангилагч загвар боловсруулсан байна. Эцэг хөвгүүдийг бүртгэлийн мэдээлэл болон бусад эх үүсвэр бүхий мэдээлэлд тулгуурлан холбох энэхүү загварын ачаар их хямралын үеийн хүмүүсийн нийгмийн үйл хөдлөлийг хэмжих боломжтой юм. Бернхайм, Бьоркгреген, Наеккер, Рангел (2013) судалгааны хариултыг ажиглагдах зан төлөвтэй холбосон: Судалгаанд хамрагдсан судалгааны хэсэг нь лабораторийн туршилтад оролцдог; Энэхүү өгөгдөл дээр сургагдсан машин сургалтын алгоритм нь судалгааны хариултуудын бодит сонголтыг урьдчилан таамаглаж, эдийн засагчдад тайлангийн төлөв байдлаас бодит байдлыг дүгнэх хэрэгслийг өгдөг. (Bernheim, Douglas, Daniel Bjorkegren, Jeffrey Naecker, and Anatonio Rangel, 2013)

1.3.2 Бодлогын шинжилгээ

Дараах бодлогын асуудлыг авч үзье: сэжигтнийг баривчилгааны дараа шүүх хурлыг гэртээ хүлээх эсвэл шоронд хүлээхийг шүүгч нар шийддэг. Энэ шийдвэр хуулийн дагуу шүүгчийн таамаглалд үндэслэн хийгдэнэ: суллагдсан шүүгдэгч шүүх хуралд буцаж ирэх үү эсвэл шүүх хурлаас зугтаах уу, цаашлаад гэмт хэрэгт холбогдож болох уу? Статистикийн арга хэрэгслүүд нь бодлогын таамаглалын асуудалд цөөн хэдэн замаар тусалдаг. (жишээ нь санамсаргүй байдлаар хийсэн хяналтын туршилт “бодлого нь ажилладаг уу?” гэсэн асуултад хариулахад тусалдаг) Энэ тохиолдолд, урьдчилан таамаглах алгоритм нь шүүгчийн шийдвэрийг сайжруулахад мөн адил тусалж чадах болов уу гэж гайхаж магадгүй юм. (Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, Ziad Obermeyer, 2015)

Батлан даалтын асуудал гэх мэт урьдчилсан таамаглалын бодлогын асуудал олон газарт гарч ирдэг (Клейнберг, Людвиг, Муллайнатан, Обермейер 2015). Жишээлбэл, томоохон хэмжээний судлагдахуун нь нэмэлт багш ажилд авах үр нөлөөг тооцдог бөгөөд энэ нь нэмж багш ажилд авах эсэх талаар шийдвэр гаргахад хэрэглэгдэнэ. Гэхдээ яг аль багшийг ажилд авах тухай шийдвэр нь мөн л таамаглахыг шаарддаг бөгөөд тухайн үед байгаа хувийн мэдээллийг ашиглан шийлвэр гаргадаг. (Кейн ба Стейгер 2008; Добби 2011; Джейкоб нар, 2016) Халфин (2016) нь машин сургалт нь эдгээр хувийн шийдвэр гаргалтын таамаглалын нарийвчлалыг хэрхэн сайжруулж болох талаар зарим нотолгоог өгсөн байдаг. Мөн Абелсон, Варшней, Саан (2014) (2014), Макбрайд ба Николс (2016), Энгстром, Херш, Ньюхаус (2016) нар одоогийн ядуурлын шалгуур үзүүлэлтүүдтэй харьцуулахад ядуурлын зорилтот түвшнийг сайжруулах зорилгоор машин сургалтыг ашигласан юм. Эдгээр урьдчилан таамаглах асуудал нь бидний хариулахыг хүссэн асуултуудтай нягт холбоотой: нэмэлт багшийн нөлөө нь тухайн багш хэрхэн сонгогдсоноос хамаарна; шилжүүлгийн хөтөлбөрийн үр нөлөө нь хэр зэрэг нарийвчлалтай зорилт болгосон байхаас хамаарна. (Sendhil Mullainathan, Jann Spiess, 2017)

Урьдчилан таамаглах бодлогын асуудлыг шийдвэрлэхэд эдийн засагчид чухал үүрэг гүйцэтгэж чадна. Нэгдүгээрт, урьдчилан таамаглах нь чухал боловч, дан ганц машин сурах нь хангалтгүй юм: эконометриктэй холбоотой хэд хэдэн бэрхшээл гарч ирнэ. Шүүгчийн шийдвэр гаргах тохиолдлын хувьд аль алгоритмыг ашигласнаар таамаглалыг сайжруулах боломжтой эсэхийг шийдэхийн тулд суурь, үндсэн асуудлыг шийдэх хэрэгтэй: бид зөвхөн батлагдсан тохиолдолд л гэмт хэргийг хүлээн зөвшөөрнө. Хоёрдугаарт зан төлөвийн асуудал үүсдэг. Алгоритм нь таамаглал хийхэд тусалж байгаа ч энэ хэрэгслийг сонгох болсон хүчин зүйлсийг ойлгох хэрэгтэй. Алгоритмд итгэх итгэлийг ямар хүчин зүйл тодорхойлдог вэ? Илүү энгийн алгоритмд их итгэх нь зөв үү? Хувийн мэдээллийг оновчтой ашиглахыг шүүгчдийг хэрхэн дэмжиж байгаа вэ? Эдгээр асуулт нь технологийн, мэдээллийн эдийн засаг, зан төлөвийн эдийн засгийн асуудлуудыг нэгтгэдэг. (Dietvorst, Berkeley J., Joseph P Simmons, and Cade Massey, 2015)

1.3.3Онолыг шалгах

Машин сургалтын эцсийн хэрэглээ нь таамаглалын чадвартай холбоотой онолуудыг шууд туршиж үзэх явдал юм. Санхүүгийн үр ашигтай зах зээлийн онолын хүрээнд, жишээлбэл, ирээдүйн талаар урьдчилан таамаглах чадваргүй болох нь гол таамаглал юм. Мориц ба Зиммерманн (2016) АНУ-ын компаниудын өнгөрсөн үеийн өгөөж нь хувьцааныхаа цаашдын үнэд нөлөөлөхүйц хүчтэй болохыг машин сургалтын аргуудыг ашиглан харуулсан. (Moritz, Benjamin, Tom Zimmermann, 2016)

Машины сургалтыг онол хэр сайн гүйцэтгэлтэй байгааг хэмжих жишиг стандарт бий болгоход ашиглаж болно. Хамгийн анхаарал татдаг асуудал бол онол хэдий зөв байсан ч тайлбарлахаар зорьж буй системийн хэлбэлзлийн багахан хэсгийг тодорхойлж магадгүй. R^2 нь дангаараа энэ асуудлыг авч үздэггүй бөгөөд нийт хэлбэлзлийг үүгээр хэмжих боломжгүй. Клейнберг, Лян, Муллайнатан (2015) онолын таамаглалын хүчийг оновчтой таамаглагчтай харьцуулахыг санал болгож байсан. Үүнтэй холбоотойгоор Пейсахович, Наеккер (2015) зан төлөвийн эдийн засгийн эрсдэлтэй, хоёрдмол утгатай загвар сонголтуудын загваруудыг машин сургалтын шалгуур үзүүлэлттэй харьцуулж үзсэн байдаг.

II БҮЛЭГ. ОНОЛЫН УХАГДАХУУН БА ЗАГВАР

2.1 Шийдвэрийн модны арга зүй

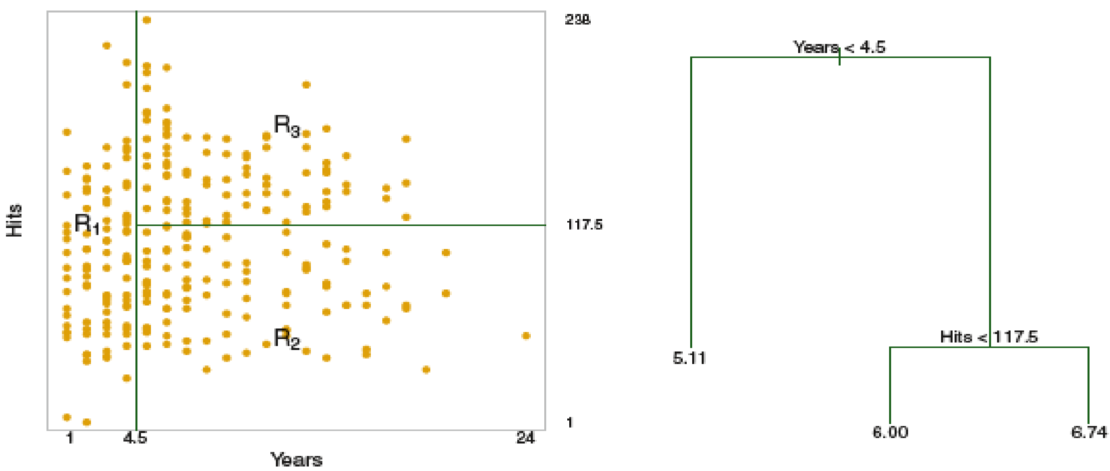
Өгөгдсөн ажиглалтын тусламжтай үзэгдлийг таамаглахдаа бид хуваасан бүлгүүдийн дундаж болон моодыг ашигладаг. Өгөгдлийг тодорхой хэсгүүдэд хуваах дүрмийг дээрээс доош салаалах модоор хураангуйлдаг бөгөөд үүнийг шийдвэрийн мод хэмээдэг. Шийдвэрийн модонд суурилсан аргазүй нь энгийн, тайлбарлахад хялбар юм. Энэхүү бүлгээр бид шийдвэрийн модонд суурилсан аргазүйн ерөнхий хэлбэрийг авч үзэх ба баггинг, санамсаргүй ойжуулалт болон бүүстинг гэх зэрэг хэд хэдэн модыг нэгтгэх замаар таамаглал хийдэг аргуудыг дурдана. Шийдвэрийн мод нь тоон болон чанарын шинжтэй асуудлыг шийддэг бөгөөд эхлээд регрессийн модыг дараа нь ангиллын номыг авч үзэх болно.

2.1.1 Регрессийн мод

Шийдвэрийн мод нь навч (эцсийн зангилаа), дотоод зангилаанаас бүрдэх бөгөөд дээрээс доош чиглэсэн буюу доошоо харсан модтой ижил тул навч (эцсийн зангилаа) нь хамгийн доор байрлана. Зангилаа бүр мөчрөөр холбогдоно. Эцсийн зангилаа нь эцсийн үр дүнг, дотоод зангилаа нь өгөгдлийг хуваах шалгуурыг илэрхийлдэг.

Хялбар байх үүднээс Бэйсболын тамирчиний цалинг регрессийн мод ашиглан дүрслэх жишээг авч үзье. Тамирчины цалинд лигт тоглож буй жил, өнгөрсөн жил гүйцэтгэсэн цохилтын тоо нөлөөлдөг гээд шийдвэрийн модыг дүрсэлбэл доорх мод үүснэ. Зурагт регрессийн мод өгөгдөлд хэрхэн тохирч байгааг болон модыг оройгоос салбарлуулан хуваах дүрмийг харж болно. R_1, R_2, R_3 нь өмнө дурдаж байсан эцсийн зангилаа буюу модны навчийг илэрхийлэх ба модны эцэст байрлана. Дотоод зангилаа болох “$Year < 4.5$” болон “$Hits < 117.5$” нь өгөгдлийг гурван бүлэгт хувааж байгаа юм. Ингээд таамаглалын загвар бэлэн боллоо. Тухайлбал, лигт 4.5 болон түүнээс дээш жил тоглосон, өнгөрсөн улирал 117.5 – оос бага цохилт хийсэн тамирчны цалин 3 дугаар навч буюу 6 хэмээн таамаглагдаж байна.

Зураг II-1. Шийдвэрийн мод



Энгийн регрессийн модыг байгуулах аргыг хураангуй авч үзвэл таамаглагч хувьсагчдыг үл давхцах бүлгүүдэд хуваах ба нэг бүлэгт хамаарах ажиглалтуудын хувьд нэг ижил таамаглал хийх буюу хамаарах хувьсагчийн дундаж утгыг ашиглана. Нэгдүгээр алхмыг хэрэгжүүлэхэд бид таамаглалыг алдааны квадратуудын нийлбэр (ҮКН) буюу дараах тэгшитгэлийг хамгийн бага байлгах шаардлагыг авч үздэг.

$$\sum_{j=1}^{J} \sum_{i \in R_j} \left(y_i - \hat{y}_{R_j}\right)^2 \tag{1}$$

Энд $\hat{y}_{R_j}$ нь j дүгээр бүлэгт буй хамааран хувьсагчдын дундаж утга юм. Бид дээрээс доош чиглэсэн рекурсив хоёртын хуваалтыг ашигладаг бөгөөд модны орой буюу бүх таамаглагчдыг бүлгүүдэд хуваах замаар мөчир үүсгэж модыг доош салбарлуулдаг юм. Рекурсив хоёртын хуваалтыг бид эхлээд таамаглагч X_i болон түүнийг $\{X|X < s\}$ болон $\{X|X \geq s\}$ гэсэн 2 хэсэгт хуваах s хуваалтын цэгийг сонгох шаардлагатай. Өөрөөр өгөгдлийг бүлэг бүрийн хувьд ҮКН буюу таамаглалын үлдэгдлийн квадратуудын нийлбэрийг хамгийн бага байхаар хуваах таамаглагчид болон хуваалтын цэгийг олох явдал юм.

2.1.2 Мод тайрах

Дээр дурдсан шийдвэрийн мод ургуулах үйл явц нь сургалтын өгөгдлийн хувьд сайн үр дүнг өгч болох ч өгөгдлийг хэт үнэлэх, тестийн өгөгдлийн хувьд муу гүйцэтгэлтэй байх эрсдэлтэй юм. Учир нь ургуулсан мод маань хэт цогц бүтэцтэй байх магадлалтай. Цөөн хуваалттай намхан мод багахан хэмжээний гажуудалтай, бага хэлбэлзэлтэй, тайлбарлахад хялбархан байх хандлагатай. ҮКН-ийг байж болох хамгийн ихээр бууруулах хуваалтыг агуулсан модыг ургуулах нь намхан мод үүсгэх ч хэт богиныг харсан хуваалт нь ирээдүйн үнэ цэнэ бүхий хуваалтыг үгүйсгэх эрсдэлтэй.

Маш том мод ургуулаад түүнийгээ тайрч засах замаар дэд модуудыг гаргаж авах нь илүү сайн арга юм. Мэдээж дэд модуудыг сонгож авахад ч гэсэн шалгуур байх бөгөөд энэ нь шалгуурын алдааны түвшин бага байх явдал юм. Үүнийг бид cross- validation- валидэйшн болон валидэйшн аргуудын тусламжтай гүйцэтгэнэ. Гэхдээ бүх боломжит дэд мод дээр үүнийг тооцох нь утгагүй бөгөөд цөөн хэдэн дэд модыг сонгож авах хэрэгтэй. Үүнийг зардалд суурилж тайралт эсвэл хамгийн сул холбоос тайралт хэмээн нэрлэдэг. Боломжит бүх модыг сонгохын оронд бид сөрөг биш утгатай α параметрээр индексжүүлсэн дэд моднуудыг авч үзнэ. $T \subset T_0$ дэд модонд харгалзах α нь дараах тэгшитгэлийг хамгийн бага утгатай байхаар сонгогддог.

$$\sum_{m=1}^{|T|} \sum_{i: x_i \in R_m} \left(y_i - \hat{y}_{R_m}\right)^2 + \alpha|T| \tag{2}$$

Энд $|T|$ нь T модны эцсийн зангилаа, R_m нь m дүгээр эцсийн зангилаанд харгалзах тэгш өнцөгт, $\hat{y}_{R_m}$ нь R_m тэгш өнцөгт буюу m дүгээр сургалтын өгөгдлийн дундаж буюу таамаглагдсан утга. Чухамдаа α параметр нь сургалтын өгөгдлийн тохирсон байдал болон дэд модны Complexity_хоёрын ацан шалааг зохицуулдаг. Хэрвээ $\alpha = 0$ бол $T = T_0$ болох буюу дээрх тэгшитгэл нь сургалтын өгөгдлийн алдааг хэмжих юм. Бид α -ийн утыг cross- validation болон валидэйшн аргуудын тусламжтай олох ба энэ утгад харгалзах дэд моднуудыг өгөгдлөөс сонгон авна. Алхмуудыг Алгоритм 1-д харуулав.

Алгоритм 1 Регрессийн мод ургуулах

1. Рекурсив хоёртын хуваалтын тусламжтай сургалтын өгөгдөл дээр том мод ургуулах. Энэ үйл явц нь эцсийн зангилаад хамгийн бага ажиглалтын тооны хязгаар хүртэл үргэлжилнэ.
2. Ургуулсан том модоо α функцийн тусламжтай тайрч засан хамгийн шилдэг дэд модны цуваг олж авна.
3. K-дахин cross validation – г ашиглан α – ийн утгыг олно. Энэ нь сургалтын өгөгдлийг K ширхэг бүлэгт хуваана гэсэн үг. $k = 1, \dots, K$ байх бүлгийн хувьд:
   a. 1 болон 2 дугаар алхмыг K дугаар бүлгээс бусад хэсэгт давтан гүйцэтгэнэ
   b. α функцийн хувьд өгөгдлөөс үлдсэн K дугаар бүлгийг ашиглан таамаглалын алдааны квадратыг үнэлнэ

   α функцийн тусламжтай үр дүнг дундажлах замаар дундаж алдааг хамгийн бага байлгах α – г сонгон авна.
4. Сонгогдсон α – д харгалзах дэд моднуудыг олж авна.

2.1.3 Ангиллын мод

Ангиллын мод нь регрессийн модтой төстэй боловч тоон биш чанарын хувьсагчийг таамаглах болно. Регрессийн модыг санавал бид нэг ижил эцсийн зангилаанд хамаарах ажиглалтын хувьд хамааран хувьсагчийн тооцож, дундажлаж замаар таамаглал хийж байсан юм. Харин ангиллын модны хувьд навч буюу эцсийн зангилаа бүрийн хувьд хамгийн их давтагдсан ангиллын тусламжтай таамаглал хийдэг.

Регрессийн модыг ургуулахад ҮКН – ийн тусламжтай рекурсив хоёртын хуваалтыг ашигладаг бол ангиллын модонд үүнийг ашиглах боломжгүй юм. Учир нь ҮКН -ийг чанарын хувьсагчийн хувьд тооцох боломжгүй тул *ангиллын алдааны түвшинг* ашигладаг юм. Тиймээс тухайн нэг навчинд буй хамгийн их давтагдсан ангиллыг хуваарилах болно, харин ангиллын алдааны түвшин нь тухайн бүлгийн ажиглалтын тоо болон хамгийн их давтагдсан ангилалд хамаарахгүй ажиглалтуудын тооны харьцаа юм.

$$E = 1 - \max_{k}(\hat{p}_{mk}) \tag{3}$$

Энд $\hat{p}_{mk}$ нь m дүгээр бүлэгт k дугаар ангиллын эзлэх хувийг илэрхийлнэ. Гэвч энэхүү ангиллын алдаа хэмээх нь мод ургуулахад хангалттай мэдрэмтгий биш тул практикт 2 аргыг өргөнөөр ашигладаг.

Жинийн индекс нь дараах байдлаар тодорхойлогдох бөгөөд K ангиллуудын нийт хэлбэлзлийг хэмждэг. Хэрвээ $\hat{p}_{mk}$ нь 1 эсвэл 0 тэй тэнцүү бол Жинийн индекс бага гарах нь илэрхий байна. Индексийн бага утга нь тухайн зангилаанд нэг анги зонхилсон байгааг итгэх агаад энэ нь зангилааны цэвэр байдлыг хэмжих боломж бүхий Жинийн индексийн давуу тал юм.

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{4}$$

Дараагийн боломжит арга бол хөндлөн энтропи гэж нэрлэгдэх хэмжүүр :

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk} \tag{5}$$

Энд $0 \leq \hat{p}_{mk} \leq 1$ болон $0 \leq -\hat{p}_{mk} \log \hat{p}_{mk}$ нөхцөл биелнэ. Хэрвээ бүх $\hat{p}_{mk}$ – ийн утга 1 эсвэл 0 – тэй ойрхон бол хөндлөн энтропи бага утгатай байна. Жинийн индекстэй ойролцоо гарах энэхүү хэмжүүр нь мөн л m дүгээр зангилааны цэвэр байдлыг хэмжинэ.

Ангиллын модыг ургуулахдаа Жинийн индекс эсвэл Кроссентропи нь ихэвчлэн тодорхой хуваагдлын чанарыг үнэлэхэд ашиглагддаг ба аль аль нь зангилааны цэвэр байдалд ангиллын алдааны түвшнээс илүү мэдрэмтгий байдаг. Модыг тайрахад эдгээр гурван аргын аль нэгийг нь ашиглаж болно, гэхдээ эцсийн тайрсан модны зорилго нь таамаглалын нарийвчлал бол ангиллын алдааны түвшинг ашиглах нь илүү юм.

2.1.4Шийдвэрийн модны давуу ба сул тал

Шийдвэрийн модны арга зүй нь дараах давуу талуудтай. Үүнд:

- Бусад хүмүүст тайлбарлахад хялбар, ойлгомжтой. Шугаман регрессээс илүү хялбар гэж үздэг.
- Зарим хүмүүс шийдвэрийн мод нь бодит амьдрал дээрх хүний шийдвэр гаргалттай ижил хэмээн үздэг.
- Зураглахад хялбар бөгөөд мэргэжлийн бус хүмүүс ч төвөггүй ойлгоно.
- Чанарын шинжтэй таамаглалыг дамми үүсгэхгүйгээр хялбархан шийддэг.

Мөн дараах сул талуудтай. Үүнд:

- Бусад машин сургалтын аргатай харьцуулахад нарийвчлал бага, вариац өндөр

2.2 Баггинг

Өмнөх бүлэгт авч үзсэн шийдвэрийн мод нь харьцангуй хэлбэлзэл ихтэй байдаг. Өөрөөр хэлбэл сургалтын өгөгдлийг бид санамсаргүйгээр хуваагаад хоёуланд нь шийдвэрийн модыг ургуулбал үр дүн нь маш их зөрүүтэй гарна. Эсрэгээрээ ялгаатай өгөгдөл дээр энэхүү арыг давтвал хэлбэлзэл багатай ижил үр дүн гарна, мөн n болон p -ийн харьцаа их үед шугаман регресс бага хэлбэлзэлтэй байх хандлагатай. Бүүтстрап нэгтгэл буюу баггинг аргазүй нь статистик сургалтын аргуудын вариацыг бууруулах зорилготой ба шийдвэрийн модтой холбоотой асуудалд өргөнөөр ашигладаг.

Бие биеэсээ үл хамаарах, σ^2 вариацтай n ширхэг үл хамаарах $Z_1, \dots, Z_n$ ажиглалт өгсөн гэвэл дундаж нь болох $\bar{Z}$ – ийн вариац σ/n болно. Өөрөөр хэлбэл ажиглалтуудыг дундажлах нь вариацыг бууруулдаг. Таамаглалын вариацыг бууруулах, нарийвчлалыг сайжруулах уламжлалт арга нь эх олонлогоос маш олон сургалтын өгөгдөл салган авч, тус бүрд нь таамаглалын загвар боловсруулж, үр дүнг дундажлах явдал билээ. Өөрөөр хэлбэл $\hat{f}^1(x), \hat{f}^2(x), \dots, \hat{f}^B(x)$ гэсэн B ширхэг бие даасан сургалтын өгөгдөл гаргаж авсан гэвэл бага вариацтай статистик сургалтын загварыг дараах байдлаар олно:

$$\hat{f}_{avg}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^b(x) \tag{6}$$

Мэдээж бодит байдал дээр олон сургалтын өгөгдөл байх боломжгүй тул бүүтстрап ашиглан ганц сургалтын өгөгдлөөс олон тооны түүврийг гарган авдаг юм. B ширхэг ялгаатай бүүтстрап хийгдсэн сургалтын өгөгдөл бий болгон, түүнийгээ нэгтгэдэг. Тухайлбал, b ширхэг сургалтын өгөгдлийн тусламжтай таамаглал хийж нэгтгэнэ.

$$\hat{f}_{avg}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \tag{7}$$

Үүнийг л баггинг гэх бөгөөд шийдвэрийн модыг гүн ургуулах боломж олгон, тайрч засах шаардлагагүй болгодог. Мод бүр дангаараа гажуудал багатай, вариац өндөртэй байх боловч үүнийг дундажлах замаар бууруулдаг. Регрессийн модны хувьд дундажлах нь илэрхий юм. Харин ангиллын модны хувьд олонхын санал буюу B ширхэг таамаглалд хамгийн их тохиолдох ангиллыг тооцох аргыг ашигладаг.

Баггингаас гадна буй алдааны тооцоолол

Баггинг ашигласан загварын хувьд тестийн өгөгдлийн алдааг шууд тооцох арга байдаг буюу хөндлөн баталгаажуулалт (cross-validation) хийх шаардлагагүй. Баггинг арга мод нь ойролцоогоор нийт ажиглалтын гуравны хоёртой тэнцүү хэмжээний өгөгдлийг ашигладаг ба үлдсэн хэсгийн цүнхнээс гадна буй хэсэг гэж нэрлэдэг. Тестийн өгөгдлийн алдааг тооцохдоо цүнхний гадна байх өгөгдлийг мод бүрээр таамаглаж дундажлах замаар 1 таамаглал гарган авч бодит өгөгдөлтэй харьцуулан үздэг.

III БҮЛЭГ. ЭМПИРИК СУДАЛГААНЫ АРГА, АРГАЗҮЙ

Аргазүйн хэсэг нь бүхэлдээ Монте-Карло симуляцийн агуулга, үйл явцыг хамарна. Нэгдүгээрт, гурван өөр ӨҮП-аас хугацаан цуваны өгөгдлийг хэрхэн үүсгэх талаар авч үзэх болно. Хоёрдугаарт, бид эдгээр таамаглалын гүйцэтгэлд хугацааны цуваны таамаглалын урт, үлдэгдэл санамсаргүй хэмжигдэхүүний тархалтын онцлог болон баггинг аргазүйн бүүтстрап хийх тоо хэрхэн таамаглалын нарийвчлалд нөлөөлөхийг тооцно. Гуравдугаарт, бид ашиглах баггинг машин сурах алгоритмыг хэрхэн ашиглах талаар нарийвчлан авч үзэх болно.

Өгөгдөл үүсгэх процесс (ӨҮП)

Бид судалгаандаа практикт хамгийн түгээмэл ашиглагддаг, хэвийн тархсан үлдэгдэл санамсаргүй хэмжигдэхүүн бүхий авторегрессив (AR), threshold авторегрессив загвар (TAR) болон Ерөнхийлсөн авторегрессив нөхцөлт хетероскедастик (GARCH) процессуудыг авч үзэх болно. ӨҮП нэг бүрийн хувьд N=100 урттай хугацааны цувааг үүсгэх (сургалтын өгөгдөл) бөгөөд бүх тохиолдолд AR загвараар загварчилах болно. Түүврээс гаднах таамаглалын алдааг хэмжихдээ дараагийн 1, 6 болон 12 дахь бодит утгатай (тестийн өгөгдөл) харьцуулах болно. Бидний ӨҮП-ийн ерөнхий хэлбэр тэгшитгэл хэлбэрээр харагдаж байна.

1. Авторегрессив (AR) загвар

$$AR(2): y_t = 0.5y_{t-1} + 0.45y_{t-2} + \varepsilon_t$$
$$\varepsilon_t \sim N(0,1)$$

2. Threshold autoregressive (TAR(2)) загвар

$$y_t = \begin{cases} 0.1y_{t-1} + 0.09y_{t-2} + \varepsilon_t & \text{энд } |y_{t-1}| \leq 0 \\ 0.2y_{t-1} + 0.05y_{t-2} - \varepsilon_t & \text{энд } |y_{t-1}| > 0 \end{cases}$$

$$\varepsilon_t \sim N(0,1)$$

3. GARCH (2,2) загвар

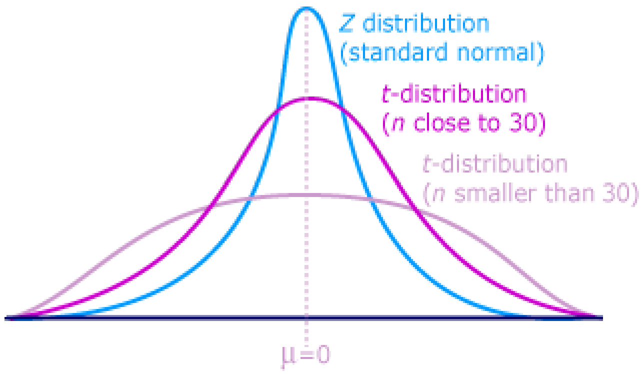
$$y_t^2 = 0.12y_{t-1}^2 + 0.04y_{t-2}^2 + 0.08\sigma_{t-1}^2 + -0.05\sigma_{t-2}^2 + \varepsilon_t$$
$$\varepsilon_t \sim N(0,1)$$

Дээрх ерөнхий хэлбэрийн тавил дээр үндэслэн бүүтстрапын тооны өөрчлөлт таамаглалын гажуудал болон MSE -д хэрхэн нөлөөлөхийг авч үзэхдээ 100 удаагийн бүүтстрап дахин түүвэрлэлтийг ашиглана. Үлдэгдэл санамсаргүй хэмжигдэхүүний хувьд стандарт хэвийн тархалт болон 20,40 чөлөөний зэрэгтэй хувилбаруудыг туршиж үзнэ. Коэффициэнтүүдийг нэгж язгууртай ойрхон эсвэл цагаан шуугиантай ойр гэх байдлаар өргөтгөх болно.

Үлдэгдэл санамсаргүй хэмжигдэхүүний тархалтын шинжилгээ.

Үлдэгдэл санамсаргүй хэмжигдэхүүний тархалтын шинж чанар нь таамаглалын гүйцэтгэлд хэрхэн нөлөөлөх нь сонирхолтой юм. Энгийн AR загварын хувьд бага вариацтай хэвийн тархалттай загвар нь таамаглалын бага гажуудалтай байна гэсэн хүлээлттэй байна. Харин бага чөлөөний зэрэгтэй стьюдент тархалт нь энгийн таамаглалын алдааг нэмэгдүүлэх хандлагатай, харин баггинг аргазүйн хувьд энэ асуудлыг хэрхэн шийдэхийг уг шинжилгээгээр авч үзнэ. Дараах зурагт хэвийн тархалт болон стьюдент тархалтын нягтын дүрслэлийг харуулав.

Зураг III-1 Хэвийн ба Т тархалт



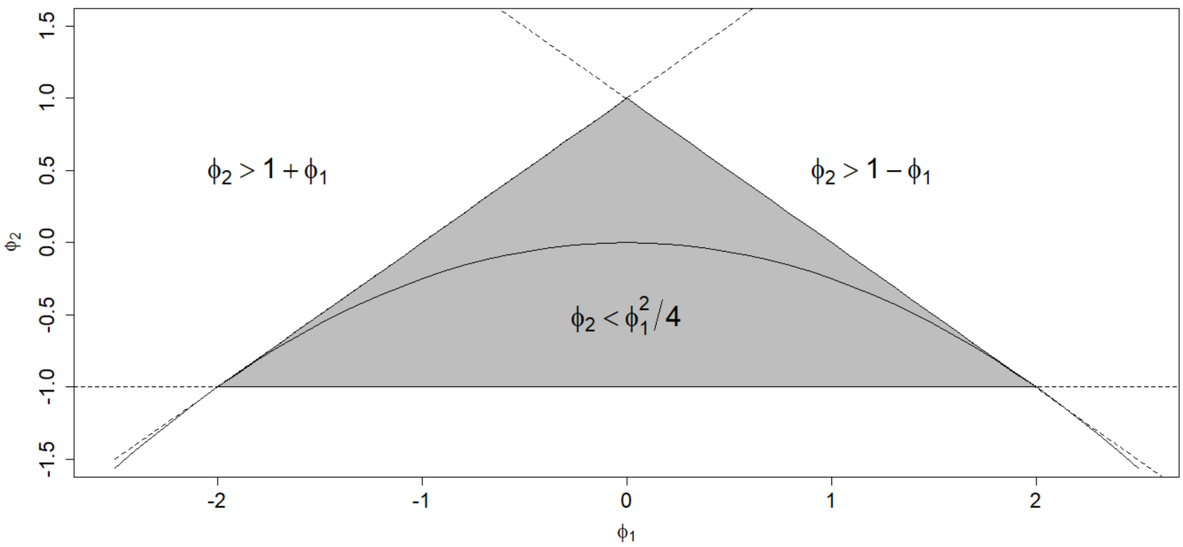
Процессын шинжилгээ.

Тухайн процессын хувьд нэгж язгууртай ойрхон, цагаан шуугиантай ойрхон байх асуудлыг авч үзнэ. AR(2) процессын хувьд дараах гурван нөхцөлийн уулзвар дээр тогтвортой байх нөхцөлийг авч үзэх боломжтой. (Зураг III-2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

1. $\phi_2 < 1 + \phi_1$
2. $\phi_2 < 1 - \phi_1$
3. $\phi_2 > -1$

Зураг III-2. Шийд тогтвортой байх гурвалжин



Энэхүү шинжилгээний хэсгийн зорилго нь ӨҮП-аар үүсгэсэн симуляцийн цуваа маань нэгж язгууртай ойрхон, цагаан шуугиантай ойрхон эсвэл бүтцийн өөрчлөлттэй байгаа тохиолдолд баггинг аргазүйн гүйцэтгэл хэрхэн өөрчлөгдөж буйг дүгнэх билээ.

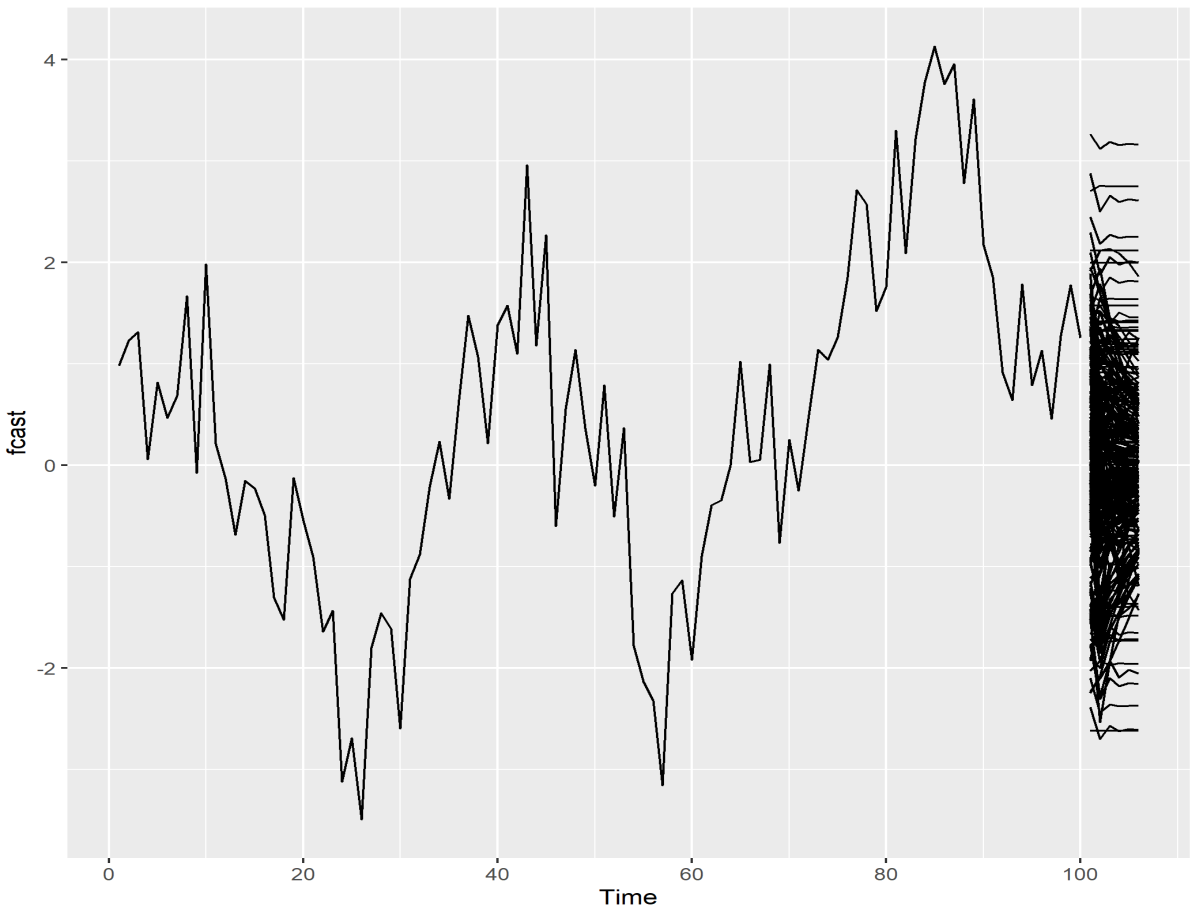
Бүүтстрап дахин түүвэрлэлтийн шинжилгээ.

Симуляцийн дахин түүвэрлэлтийн хувьд аль болох их хэмжээний бүүтстрап хийхийг шаарддаг билээ. Практикт бүүтстрапын тоог 500, 1000 болон түүнээс дээш байхыг илүүд үздэг. Харин бид энэхүү судалгаанд 100 удаагийн дахин түүвэрлэлтийн үр дүнг авч үзнэ. Жишээлбэл нэг цуваны хувьд 250 удаа бүүтстрап хийж таамаглалд дүрсэлбэл дараах дүр зураг харагдана. Дараах AR(2) ӨҮП бүхий симуляцийн цуваны хувьд таамаглалын замуудыг зурагт харууллаа.

$$AR(2): y_t = 0.5y_{t-1} + 0.45y_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim N(0,1)$$

Зураг III-3. Таамаглалын замууд



Үндсэн алгоритмыг дараах байдлаар тодорхойлж болох юм.

1. Харгалзах ӨҮП бүрд 100 урттай симуляцийн цуваа үүсгэх
2. Тухайн цувааг АР(2) загвараар загварчилж, дараагийн h үеийн таамаглалыг хийх замаар y_{t+h}^{org} гаргаж авах
3. Тухайн цувааг блок бүүтстрап хийх замаар B ширхэг цуваа гарган авч АР(2) загвараар таамаглал хийж үр дүнг дундажлах $y_{t+t}^{bag} = \frac{\sum_{i=1}^{B} y_{t+h}^{B}}{B}$
4. Гуравдугаар алхмын үр дүн бодит утгатай харьцуулах
5. 1-4 дэх алхмыг 100 удаа давтах
6. Нэгдсэн MSE болон гажуудлыг хэмжих

Шинжилгээний хэсгийг нэгтгэн нэг хүснэгтэд багтаавал :

Хүснэгт III-1. Шинжилгээний хувилбар

Шинжилгээний төрөл	Хувилбар
Таамаглалын уртын шинжилгээ	1
	6
	12
Тархалтын шинжилгээ	N (1,0)
	T (df=15)
	T (df=30)
Процессын шинжилгээ	Нэгж язгууртай ойрхон
	Цагаан шуугиантай ойрхон
	Бүтцийн өөрчлөлттэй

IV БҮЛЭГ. ЭМПИРИК СУДАЛГАА

Монте-карло симуляцийн үр дүнд бид өгөгдөл үүсгэх гурван процесүүдийн хувьд баггинг аргазүй ялгаатай замаар нөлөөлж байгааг олж тогтоолоо. Бид AR(2), TAR(2) болон GARCH(2) өгөгдөл үүсгэх процессуудыг дагах гурван төрлийн 100 ширхэг симуляцийн загварыг AR(2) загвараар таамаглаж үр дүнг таамаглалын алдааны квадратуудын дунджаар харьцуулав. Гурван төрлийн шинжилгээ ашигласан бөгөөд үүнд үндсэн шинжилгээ, тархалтын шинжилгээ болон процессын шинжилгээ зэрэг болно.

4.1 Үндсэн шинжилгээ

Хүснэгт IV-1. Үндсэн шинжилгээний үр дүн

Загвар	h = 1		h = 6		h = 12	
	AR	Баггинг	AR	Баггинг	AR	Баггинг
AR	1.05625	6.79293	3.03006	7.60872	5.33877	8.01010
TAR	1.02132	1.00763	0.87752	0.87366	0.97734	0.97772
GARCH	3.73005	3.60248	3.47399	3.47395	3.18152	3.17648

Эх сурвалж: Судлаачийн тооцоолол

Үндсэн шинжилгээний хувьд үлдэгдэл санамсаргүй хэмжигдэхүүн стандарт хэвийн тархалттай, бүүтстрап дахин түүвэрлэлтийн тоо 40 бөгөөд дараах гурван ӨҮП-ийг дагах симуляцийн цуваг ашиглах болно. Үндсэн шинжилгээний гол зорилго нь AR загварын болон баггинг ашиглаж нэгтгэн дундажласан таамаглалын утга бодит утгаасаа хэр ялгаатай байгааг шинжлэх, өөрөөр хэлбэл MSE -ийн тусламжтай баггинг аргазүй таамаглалын нарийвчлалд хэрхэн нөлөөлж буйг хэмжих.

AR(2) ӨҮП-ийн тусламжтай симуляцийн цуваа үүсгээд түүнийгээ AR(2) загвараар таамаглан, үр дүнг баггинг аргазүйтэй харьцуулах нь учир дутагдалтай юм. Өөрөөр хэлбэл энэ энэ ӨҮП-ийн хувьд баггинг аргазүйн гүйцэтгэл сайн байхгүй нь илэрхий. Харин бусад ӨҮП -ийн хувьд (TAR, GARCH) AR(2) оор загварчилсан нь таамаглалын тодорхой бус байдлыг харуулах бөгөөд баггинг аргазүйн гүйцэтгэлийг сорих жинхэнэ талбар юм.

Хүснэгт IV-1 -д харуулсанчлан таамаглалын алхам бүрийн хувьд ӨҮП тус бүрийн хувьд AR(2) болон баггинг аргазүйн таамаглалын гүйцэтгэлийг харьцуулсан байна. Мэдээж таамаглах хугацааны урт ихсэх тусам MSE утга өсөх нь мэдээж. Хүснэгтэд мөрийн дагуу ӨҮП бүрт харгалзах MSE үүдийг харуулсан бөгөөд таамаглалын 1, 6 болон 12 дахь үеийн AR болон баггинг загваруудыг таамаглалын алдааг харуулсан болно.

AR ӨҮП – харгалзах MSE ийн утга h буюу таамаглах үеийн уртын хэмжээ нэмэгдэх тусам AR загварын хувьд ихэсч, баггинг загварын хувьд тогтвортой харагдаж байна. Баггинг аргазүйн MSE нь TAR, GARCH ӨҮП-ийн хувьд сайн гүйцэтгэлийг өгч байна. Энэ нь загвараа буруу тодорхойлсон тохиолдолд баггинг аргазүй нь сайн сонголт байж болно гэдгийг харуулж байна.

4.2 Тархалтын шинжилгээ

Тархалтын шинжилгээний гол зорилго нь үлдэгдэл санамсаргүй хэмжигдэхүүний тархалт таамаглалын гүйцэтгэлд хэрхэн нөлөөлөхийг авч үзэх билээ. Энэхүү шинжилгээнд стандарт тархалт ба 20, 40 чөлөөний зэрэгтэй Т тархалтыг авч үзнэ. Т тархалтын хувьд чөлөөний зэрэг бага байх тусам таамаглалын гүйцэтгэлд сөргөөр нөлөөлөх хандлагатай буюу MSE харьцангүй их гарах хүлээлттэй байна.

Хүснэгт IV-2. Тархалтын шинжилгээний үр дүн

Загвар	$N \sim (0,1)$		$T \sim (df = 20)$		$T \sim (df = 60)$	
	AR	Баггинг	AR	Баггинг	AR	Баггинг
AR	1.05625	6.79293	1.22117	6.76804	1.11804	7.29705
TAR	1.02132	1.00763	1.26151	1.24028	1.04342	1.03679

Эх сурвалж: судлаачийн тооцоолол

AR ӨҮП-ийн хувьд мөн л адил сайнгүй үр дүн гарсан байна. Гэхдээ энэ шинжилгээний зорилгын дагуу тархалтын өөрчлөлтийн нөлөө таамаглаж байсантай ижил байна. 20 чөлөөний зэрэг бүхий Т тархалтын хувьд таамаглалын алдаа хэвийн тархалттай харьцангуй өссөн байгааг харж болно. Чухамдаа энэ байдалд баггинг аргазүйн үр ашиг хэр байгаа нь сонирхолтой. TAR ӨҮП -ийн хувьд баггинг аргазүйн гүйцэтгэл харьцангуй сайжирсан харагдаж байна. Тухайлбал, $df = 20$ тохиолдолд хоёр аргазүйн таамаглалын зөрүү хамгийн их байна. Энэ нь загвар буруу тодорхойлогдсон, үлдэгдэл санамсаргүй хэмжигдэхүүний тархалт хэвийн бус хуульд захирагддаг тохиолдолд баггинг аргазүй үр ашигтай байх хандлагатайг илтгэж байгаа юм.

4.3 Процессын шинжилгээ

Энэхүү шинжилгээний хэсгийн зорилго нь ӨҮП-аар үүсгэсэн симуляцийн цуваа маань нэгж язгууртай ойрхон, цагаан шуугиантай ойрхон эсвэл бүтцийн өөрчлөлттэй байгаа тохиолдолд баггинг аргазүйн гүйцэтгэл хэрхэн өөрчлөгдөж буйг дүгнэх билээ. Шинжилгээний үр дүнг доорх хүснэгтэд харуулав.

Хүснэгт IV-3. Процессын шинжилгээний үр дүн

Загвар	Цагаан шуугиантай ойр		Нэгж язгууртай ойрхон		Бүтцийн өөрчлөлттэй	
	AR	Баггинг	AR	Баггинг	AR	Баггинг
AR	1.0675352	1.0483674	1.056252	6.792932	1.107217	4.611568

Эх сурвалж: судлаачийн тооцоолол

Тухайн ӨҮП нь цагаан шуугиантай ойр үед баггинг аргазүйн таамаглалын гүйцэтгэл харьцангуй сайн байна. Гэвч энд нэг зүйлийг анхаарах ёстой нь ӨҮП маань өөрөө AR бөгөөд таамаглалыг AR загвараар хийж байгаа бөгөөд ийм нөхцөлд баггингийн үр дүн сайн гарахгүй нь мэдээж. Мөн нэгж язгууртай ойр болон бүтцийн өөрчлөлттэй үед блок бүүтстрап хийх нь тохиромжгүй гэдгийг харуулж байна.

ДҮГНЭЛТ, САНАЛ

Энэхүү судалгааны ажлаар загварын тодорхой бус асуудлыг шийдэх машин сургалтын баггинг аргазүйн талаар авч үзлээ. Таамаглалын загварын тодорхой бус байдал гэдэг нь тухайн цувааг яг ямар загвараар загварчлах талаар мэдэхгүй, өөрөөр хэлбэл жинхэнэ загварыг мэдэхгүй байгаа нөхцөлийг үзэж болно. Судлагдсан байдлаар бид таамаглалын загваруудын хөгжлийн талаар товч дурдсан бөгөөд бидний өргөнөөр ашигладаг Бокс-Жинкинсийн загварын давуу болон сул талын талаар авч үзсэн юм. Эдгээр загвар нь тухайн нэг асуудлыг сайн шийдэж болох ч ерөнхий тохиолдолд таамаглалын муу үр дүн өгөх хандлагатай байдаг бөгөөд судлаачид нэгтгэх, дундажлах гэх мэт олон төрлийн арга техникийг хослуулан ашигладаг.

Хугацааны цуваан таамаглалд сүүлийн жилүүдэд ашиглагдах болсон баггинг буюу бүүтстрап нэгтгэл хэмээх аргазүйн гүйцэтгэлийг симуляцийн аргаар үнэлж, харьцуулах нь энэ судалгааны гол зорилго байсан юм. Симуляцийн судалгаанд өгөгдөл үүсгэх AR, TAR болон GARCH гэсэн гурван төрлийн процессыг авч үзсэн ба таамаглалын уртын (үндсэн) шинжилгээ, тархалтын (үлдэгдэл санамсаргүй хэмжигдэх) болон процессын шинжилгээг хийсэн болно.

Таамаглах хугацааны урт нэмэгдэх тусам оновчтой таамаглах магадлал буурна. Өөрөөр хэлбэл таамаглалын MSE өсөх нөлөөтэй байдаг. Энэхүү асуудлыг баггинг аргазүйн хувьд авч үзэхэд (үндсэн шинжилгээ) загвараа буруу тодорхойлсон бусад загвараас харьцангуй тогтвортой үр дүнг өгч байсан билээ. Тархалтын шинжилгээний хувьд үлдэгдэл санамсаргүй хэмжигдэхүүний тархалтын шинж чанар таамаглалын гүйцэтгэлд хэрхэн нөлөөлж буйг судлах байсан юм. Буруу сонгогдсон загварыг баггингийн тусламжтай сайжруулахад үлдэгдэл санамсаргүй хэмжигдэхүүн нь хэвийн бус үед илүү сайн үр дүнтэй байсан болно. Харин нэгж язгууртай ойрхон хугацааны цуваан таамаглалд баггинг аргазүйг ашиглах нь тохиромжгүй гэх дүгнэлтийг процессын шинжилгээг ашиглан хийж байна.

Нэгтгэн дүгнэвэл, энэхүү судалгааны ажлаар TAR, GARCH загвараар үүсгэсэн симуляцийн хувьд баггинг аргазүй таамаглалын гүйцэтгэлийг сайжруулах боломжтойг харуулсан билээ. Тухайлбал, тухайн хугацааны цуваа цагаан шуугиантай ойролцоо, үлдэгдэл санамсаргүй хэмжигдэхүүн нь хэвийн тархалтад захирагддаггүй үед баггинг аргазүй эерэг үр дүнг өгч байна. Энэ нь таамаглалын загвар илүү нарийн, нүсэр болох тусам баггинг аргазүйн гүйцэтгэл сайжирна гэж үзсэн бусад судалгааны ажилтай тохирч байгаа юм.

Дипломын ажилд үндэслэн дараах санал зөвлөмжүүдийг дэвшүүлж байна:

1. Бодит өгөгдөл дээр хийгдсэн эмпирик судалгаа хийх нь ач холбогдолтой
2. Баггинг аргазүйг илүү нарийвчилсан, илүү том загварууд дээр ашиглах хэрэгтэй.
3. Мөн бүүтстрап дахин түүвэрлэлтийн тоог нэмэгдүүлэх нь таамаглалын гүйцэтгэлд эерэгээр нөлөөлдөг боловч тооцооллын хувьд бэрхшээлтэй байдаг. Иймд энэ төрлийн программ хангамжийг хөгжүүлэх нь ач холбогдолтой юм.

АШИГЛАСАН МАТЕРИАЛЫН ЖАГСААЛТ

Audrino, F., & Medeiros, M. C. (2011). Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and Bagging. *Journal of Applied Econometrics, 26*(6), 999–1022.

Bartolomei, S. M., & Sweet, A. L. (1989). A note on a comparison of exponential smoothing methods for forecasting seasonal series. *International Journal of Forecasting, 5*, 111-116.

Bernheim, Douglas, Daniel Bjorkegren, Jeffrey Naecker, and Anatonio Rangel. (2013). Non-Choice Evaluations Predict Behavioral Responses to Changes in Economic Conditions. *NBER Working paper*.

Blumenstock, Joshua E., Gabriel Cadamuro. (2015). Predicting Poverty and Wealth from Mobile Phone Metadata. *Science, 350*(6264), 1073–76.

Blumenstock, Joshua Evan. (2016). Fighting Poverty with Data. *Science, 353*(6301), 753–54.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis:Forecasting and control.* San Francisco: Holden Day.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* New York: John Wiley & Sons.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *26*(2), 123-140.

Brown, R. G. (1959). *Statistical forecasting for inventory control.* New York: McGraw-Hill.

Brown, R. G. (1963). *Smoothing, forecasting and prediction of discrete time series.* Englewood Cliffs: Prentice-Hall.

Cholette, P. A., & Lamy, R. (1986). Multivariate ARIMA forecasting of irregular time series. *International Journal of Forecasting, 2*, 201– 216.

Dietvorst, Berkeley J., Joseph P Simmons, and Cade Massey. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, 114–126.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*, 1057– 1072.

Franses, P. H., & Romijn, G. (1993). Periodic integration in quarterly UK macroeconomic variables. *International Journal of Forecasting, 9*, 467– 476.

Gardner Jr., E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting, 4*, 1-38.

Glaeser, Edward L., Scott Duke Kominers,Michael Luca, and Nakhil Naik. (2016). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. *Economic Inquiry*.

Henderson, J. Vernon, Adam Stor eygard,and David N. Weil. (2012). Measuring Economic Growth from Outer Space. *American Economic Review, 102*(2), 994–1028.

Hill, G., & Fildes, R. (1984). The accuracy of extrapolation methods: An automatic Box–Jenkins package SIFT. *Journal of Forecasting, 3,* 319– 323.

Hirano, K., & Wright, J. H. (2017). Forecasting with model uncertainty: Representations and risk. *Econometrica, 85*(2), 617–643.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages. *International Journal of Forecasting*(20), 5-13.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005). Prediction intervals for exponential smoothing state space models. *Journal of Forecasting, 24*, 17-37.

Inoue, A., & Kilian, L. (2008). How useful is Bagging in forecasting economic time. *Journal of the American Statistical Association*, *103*(482), 511–522.

Jin, S., Su, L., & Ullah, A. (2014). Robustify financial time series forecasting with Bagging. *Econometric Reviews, 33*(5-6), 575–605.

Johnston, F. R., & Harrison, P. J. (1986). The variance of leadtime demand. *Journal of Operational Research Society, 37*, 303– 308.

Kang, Jun Seok, Polina Kuznetsova, Michael. (2013). Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews. *EMNLP 2013: 2013 Conference on Empirical Methods in Natural Language.*

Kim, J. H. (2003). Forecasting autoregressive time series with bias-corrected parameter estimators. *International Journal of Forecasting, 19*, 493–502.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, Ziad Obermeyer. (2015). Prediction Policy Problems. *American Economic Review, 105*(5), 491–495.

Kling, J. L., & Bessler, D. A. (1985). A comparison of multivariate forecasting procedures for economic time series. *International Journal of Forecasting, 1*, 5 –24.

Landsman, W. R., & Damodaran, A. (1989). A comparison of quarterly earnings per share forecast using James-Stein and unconditional least squares parameter estimators. *International Journal of Forecasting*, 491– 500.

Ledolter, J., & Abraham, B. (1984). Some comments on the initialization of exponential smoothing. *Journal of Forecasting, 3*, 79– 84.

Lee, T.-H., & Yang, Y. (2006). Bagging binary and quantile predictors for time series. *Journal of Econometrics, 135*(1), 465–497.

Lee, T.-H., Tu, Y., & Ullah, A. (2015). Forecasting equity premium: Global historical average. *Journal of Business and Economic Statistics, 33*(3), 393–402.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions—Five years of experience. *Journal of Business and Economic Statistics, 4*, 25–38.

Liu, T. -R., Gerlow, M. E., & Irwin, S. H. (1994). The performance of alternative VAR models in forecasting exchange rates. *International Journal of Forecasting, 10*, 419– 433.

Lobell, David B. (2013). The Use of Satellite Data for Crop Yield Gap Analysis. *Field Crops Research*, 56–64 .

McClain, J. G. (1988). Dominant tracking signals. *International Journal of Forecasting, 4*, 563– 572.

Miller, D. M., & Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting, 19*, 669– 684.

Moritz, Benjamin, Tom Zimmermann. (2016). Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns. *Available at SSRN 2740751*.

Newbold, P., & Bos, T. (1989). On exponential smoothing and the assumption of deterministic trend plus white noise datagenerating models. *International Journal of Forecasting, 5*, 523– 527.

Newbold, P., Agiakloglou, C., & Miller, J. (1994). Adventures with ARIMA software. *International Journal of Forecasting, 10*, 573– 581.

Osborn, D. (1990). A survey of seasonality in UK macroeconomic variables. *International Journal of Forecasting, 6*, 327– 336.

Pegels, C. C. (1969). Exponential smoothing: Some new variations. *Management Science, 12*, 311-315.

Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 545-554.

Rapach, D. E. (2010). Bagging or combining (or both)? An analysis based on forecasting US employment growth. *Econometric Reviews*, 511-533.

Riise, T., & Tjøstheim, D. (1984). Theory and practice of multivariate ARMA forecasting. *Journal of Forecasting, 3*, 309– 317.

Satchell, S., & Timmermann, A. (1995). On the optimality of adaptive expectations: Muth revisited. *International Journal of Forecasting, 11*, 407–416.

Sendhil Mullainathan, Jann Spiess. (2017). Machine Learning: An Applied Econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.

Snyder, R. D. (1985). Recursive estimation of dynamic linear statistical models. *Journal of the Royal Statistical Society, 4*, 235– 243.

Spencer, D. E. (1993). Developing a Bayesian vector autoregressive forecasting model. *International Journal of Forecasting, 9*, 407– 421.

Sweet, A. L., & Wilson, J. R. (1988). Pitfalls in simulation-based evaluation of forecast monitoring schemes. *International Journal of Forecasting, 4*, 573–579.

West, C. (2004). Can Out-of-Sample Forecast Comparisons help Prevent Overfitting? *Journal of Forecasting*, 115-139.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica, 68*, 1084–1097.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*, 324–342.

Yar, M., & Chatfield, C. (1990). Prediction intervals for the Holt–Winters forecasting procedure. *International Journal of Forecasting, 6*, 127– 137.

Yule, G. U. (1927). On the method of investigating periodicities in disturbed series, with special reference to Wo¨ lferTs sunspot numbers. *Philosophical Transactions of the Royal Society London, Series A, 226*, 267– 298.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New York: Wiley.

ХАВСРАЛТ

Хавсралт 1. Симуляцад ашигласан R программ дээр бичсэн код:

Simulation code.R

Lkhagvasuren
2020-06-06

```
library(tseries)
library(forecast)
library(fGarch)

fitting = function(series ,h){
  model = ar.ols(series,order.max = 4)
  fcast = predict(model, n.ahead = h)$pred
  return(fcast) }

# AR(2) -------------------------------------------------------------------

n  <-  100
h <- 12

nsim  <-  500
# ar  <- c(.5,.45)
# ar  <- c(.01,.005 )
ar <- c(.8, .1)

# Boosttrap
nb  <-  100
# nb  <-  20
# nb  <-  60

# Sanamsargui hemjigdehuun
sd <- 1
# sd <- 5
# sd <- 10

b <- round( n^(1/3))

order <- c(2,0,0)

set.seed(3)
errors_sim <- matrix(nrow = nsim, ncol = h)
errors_bag <- matrix(nrow = nsim, ncol = h)

for(i in 1:nsim){
  ar.sim <- arima.sim( list(order= order, ar = ar),
                       innov = rt(n+h, df = 40)  ,  # t dist,
                       n = n + h)
  train <- head( ar.sim, n)
  test <- tail(ar.sim, h)
  errors_sim[i,] <- fitting(train, h) - test
  errors_bag[i,] <- colMeans(tsbootstrap(train, nb = nb, b=b, type = 'block
',
```

```
                                                  statistic = fitting, h= h)$statist
ic) - test}

MSE_ar <- colMeans(errors_sim^2)
MSE_bag <- colMeans(errors_bag^2)


# TAR(2) ---------------------------------------------------------------------------

### y[t-1] > g == y[t] = a0 + a1 * y[t-1] + a2 * y[t-2] + e[t]
### y[t-1]<= g == y[t] = b0 + b1 * y[t-1] + b2 * y[t-2] + e[t]

tar2.sim = function(y0, e , n, p1, p2, th){
  y = rep(0 ,n)
  a0 = p1[1]; a1 = p1[2]; a2 = p1[3]
  b0 = p2[1]; b1 = p2[2]; b2 = p2[3]
  for(t in 3:n){
    if(y[t-1] > th) y[t] = a0 + a1 * y[t-1] + a2 * y[t-2] + e[t]
    else y[t] = b0 + b1* y[t-1] + b2 * y[t-2] + e[t]}
  return(y)}

set.seed(3)
errors_sim <- matrix(nrow = nsim, ncol = h)
errors_bag <- matrix(nrow = nsim, ncol = h)


for(i in 1:nsim){
  tar.sim = tar2.sim(0, e = rt(n+h,df=40),
                     n = n+h,
                     c(0.1,.09,.07),
                     c(0.2 , .05 , .04), th = 0)
  train <- head( tar.sim, n)
  test <- tail(tar.sim, h)
  errors_sim[i,] <- fitting(train, h) - test
  errors_bag[i,] <- colMeans(tsbootstrap(train, nb = nb, b=b, type = 'block
',
                                                  statistic = fitting, h= h)$statist
ic) - test}

MSE_ar <- colMeans(errors_sim^2)
MSE_bag <- colMeans(errors_bag^2)

# GARCH(2,2) -----------------------------------------------------------------------

spec = garchSpec(model = list(alpha = c(0.2, 0.1),
                              beta = c(0.3,.1)),
                 cond.dist = 'norm')

order <- c(2,0,0)

set.seed(3)
errors_sim <- matrix(nrow = nsim, ncol = h)
errors_bag <- matrix(nrow = nsim, ncol = h)

for(i in 1:nsim){
```

```
  garch.sim = garchSim(spec, n = n + h, extended = F)$garch
  train <- head( garch.sim, n)
  test <- tail(garch.sim, h)
  errors_sim[i,] <- fitting(train, h) - test
  errors_bag[i,] <- colMeans(tsbootstrap(train, nb = nb, b=n/10, type = 'bl
ock',
                                         statistic = fitting, h= h)$statist
ic) - test}


MSE_ar <- colMeans(errors_sim^2)
MSE_bag <- colMeans(errors_bag^2)
```