**САНХҮҮ ЭДИЙН ЗАСГИЙН ИХ СУРГУУЛЬ** 

**ӨДРИЙН ХӨТӨЛБӨР**

ЭКОНОМИКСИЙН ТЭНХИМ

**ДИПЛОМЫН АЖИЛ**

**Сэдэв:** *Эдийн засгийн таамаглалд машин сургалтын аргыг хэрэглэх нь*

Гүйцэтгэсэн: . . . . . . . . . . . . . . . . . . . . . О.Лхагвасүрэн /FA16B319/

Удирдагч багш : . . . . . . . . . . . . . . . . . . . . . . Д.Хашбаатар /Мастер/

Огноо: 2020.05.21



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. 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) – ийг ашиглан таамаглалын алдааны дундажын квадратыг тооцно. Симуляцаар үүсгэсэн болон бодит амьдрал дээр буй валютын ханшийн өгөгдөлийг шинжилгээнд ашиглана. Уламжлалт таамаглалын аргууд нь тухайн тохиолдолд сайн ажиллаж болох ч ерөнхий тохиолдолд сайн гүйцэтгэлтэй байдаггүй. Судалгааны үр дүнд баггинг аргазүйн ачаар таамаглалын загваруудын чадал сайжирсан бөгөөд валютын ханшийн өгөгдөл дээр харьцангуй сайн гүйцэтгэлийг өгч байна. Баггинг буюу бүүстрап агригэйшн хэмээх машин сургалтын алгорит нь бусад сургалтын аргуудын 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, C13*

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

**АГУУЛГА**

[I БҮЛЭГ. ОРШИЛ iv](#_Toc41416622)

[II БҮЛЭГ. СУДЛАГДСАН БАЙДАЛ 1](#_Toc41416623)

[2.1 Хугацааны цувааны таамаглалын загварууд 1](#_Toc41416625)

[2.2 Баггинг аргазүйг эдийн засгийн таамаглалд ашигласан судалгааны ажлууд 1](#_Toc41416626)

[2.3 Машин сургалтын алгоритм түүний хэрэглээ 3](#_Toc41416627)

[IV БҮЛЭГ. ОНОЛЫН УХАГДАХУУН БА ЗАГВАР 4](#_Toc41416628)

[2.1 Шийдвэрийн модны аргазүй 4](#_Toc41416629)

[4.1.1 Регресийн мод 4](#_Toc41416630)

[4.1.2 Ангиллын мод 6](#_Toc41416631)

[2.2 Баггинг 7](#_Toc41416632)

[V БҮЛЭГ. ЭМПИРИК СУДАЛГААНЫ АРГА, АРГАЗҮЙ 9](#_Toc41416633)

[2.1 Өгөгдөл боловсруулах үйл явц (ӨБҮ) 9](#_Toc41416634)

[VI БҮЛЭГ. ЭМПИРИК СУДАЛГАА 10](#_Toc41416635)

[VII БҮЛЭГ. ДҮГНЭЛТ, САНАЛ 10](#_Toc41416636)

[VIII БҮЛЭГ. ХАВСРАЛТ 10](#_Toc41416637)

[IX БҮЛЭГ. АШИГЛАСАН МАТЕРИАЛ 11](#_Toc41416638)

**Хүснэгтэн мэдээллийн жагсаалт**

**ЗУРГАН МЭДЭЭЛЛИЙН ЖАГСААЛТ**

[Зураг 1 Таамаглалын стратегүүд 1](#_Toc38311373)

[Зураг 2 Шийдвэрийн мод 2](file:///C:\Users\Lkhagvaa\Documents\GitHub\Bach-diploma\my_work\Uridchilsan%20hamgaalalt\Дипломын%20ажил.docx#_Toc38311374)

**ТОВЧИЛСОН ҮГС, НЭР ТОМЪЁОНЫ ТАЙЛБАР**

AЗ Авторегрессив загвар

АКД Алдааны квадратын дундаж

ХАЗ Хязгаарлалттай авторегрессив загвар

ӨБҮ Өгөгдөл боловсруулах үйл явц

ЕАНХ Ерөнхийлсөн авторегрессив нөхцөлт хетероскедастик

# 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 төлөөлхгүй гэдэг нь илэрхий. Таамаглалын загварыг түүврийн бус аргад үндэслэн рекурсив эсвэл роллинг таамаглалуудын тохирсон байдалд түшиглсэн сонгох нь түгээмэл байдаг. Үүний дараа түүврийг бүхэлд нь ашиглан парамтетрүүдийг 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 биеээсээ 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 загварууд нь тухайн нэг даалгаврыг сайн гүйцэтгэх боловч ерөхний тохиолдолд хангалттай сайн үр дүнг өгдөггүй.

Шийдвэрийн модны аргазүй, бүүсттрап, 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 макроэдийн 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 ажлын зорилго –** Энэхүү судалгааны гол зорилго нь таамаглалын загваруудыг симуляцийн аргаар шинжлэн, харьцуулах билээ. Уламжлалт загвар Баггинг аргаар таамаглаж бусад аргуудтай харьцуулах замаар тодорхой үр дүнд хүрэх билээ. Энэ зорилгод харгалзан дараах зорилтууд дэвшигдэж байна.

* Өмнө хийгдсэн таамаглалын модель тодорхой бус байдалтай холбоотой судалгааны ажлууд болон онолын загваруудыг судлах
* Оновчтой таамаглалын стратегүүдийг сонгон авч, арга зүйг судлах
* Сонгосон аргазүйгээ ашиглан макроэкономиксийн засгийн голлох үзүүлэлтийн талаарх таамаглал хийх
* Загвар бүр ­­­дэх үнэлгээний үр дүнг харьцуулж дүгнэлт, санал боловсруулах

**Судалгааны обьект –** Судалгаанд хоёр төрлийн өгөгдлийг ашиглах бөгөөд тодорхой тархалтад захирагдах өгөгдөл (AR, TAR, GARCH) болон бодит амьдрал дээр буй валютын ханшийн өгөгдлийг сонгон авсан байгаа.

# СУДЛАГДСАН БАЙДАЛ



## 

## Хугацааны цувааны таамаглалын загварууд

Хугацааны цуваан таамаглалын загваруудыг үндсэнд нь шугаман ба шугаман бус гэж ангилдаг. Шугаман загваруудыг хялбаршуулсан, ашиглахад хялбар байдгаас түгээмэл байдаг. Өргөн хэрэглэгддэг шугаман аргачлалын жишээ бол АРИМА загварууд (Box et al. 1994) ба АРМА (ARMA) загварууд юм. Эдгээр хоёр аргачлал нь хугацаан цувааны өнгөрсөн утгууд болон өмнөх таамаглалын алдааны шугаман хамаарлыг харуулдаг. Өөр нэг түгээмэл арга нь Холт-Винтерийн арга бөгөөд үүнийг бид экспоненциаль гөлийлгөх арга гэдгээр түгээмэл мэдэх билээ. Энэ аргазүй нь хугацаан цувааны өнгөрсөн утгуудын жигнэсэн утгуудыг ашигладаг, хугацааны үе бүрт коэффициэнтийн утга экспоненциал хамаарлаар буурдаг тул саяны утгуудад илүү жин өгөх онцлогтой. Бүтцийн хувьд хандлага болон улирлын нөлөөг агуулсан байдаг тул олон төрлийн хугацааны цуваан өгөгдөлд тохиромжтой гэх давуу талтай.

Дээр дурдсан хоёр аргыг багтаасан ихэнх хугацааны цуваан загваруудын хувьд тухайн цувааны шинж чанарыг тодорхойлдог тэгшитгэлийн чухал параметрүүд нь мэдэгдэхгүй байдаг бөгөөд цаг хугацааг шаардсан туршилт, алдаа, статистик тестүүдийг ашиглах замаар тогтоогддог. Түүнчлэн загварын шугаман бүтэц нь боломжит шугаман бус байдлыг илэрхийлэх чадваргүй бөгөөд энэ нь гүйцэтгэлийн хангалтгүй байдалд хүргэдэг ажээ. Энэ бэрхшээлийг шийдэхийн тулд шугаман бус АРМА загварын хөгжүүлсэн боловч шугаман бус хэлбэрийг тодорхойлохын тулд мөн л олон төрлийн нэмэлт статистикийн тестүүд, алдааны дүрмүүдийг ашигладаг. Иймд асуудлыг хиймэл мэдрэлийн сүлжээг (ANNs) ашигласнаар шийдвэрлэх онол хөгжсөн бөгөөд эдгээр нь ямар ч шугаман бус зан авирыг барьж чаддаг загварчлалын ерөнхий бүтэц юм.

## 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 таамаглагчидыг 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 хувьсагчидыг хамруулан таамаглал хийж энэхүү аргазүйн гүйцэтгэлийг үнэлсэн байдаг. Нарийвчилж авч үзвэл тэд баггинг хийсэн LARS-ийг ДНБ-ий өсөлт, ХҮИ-ийн инфляци, IBR (АНУ дахь Холбооны сангийн ханштай дүйцэх банк хоорондын бэлэн мөнгөний ханш) -ын динамик хүчин зүйлийн загвар, Ridge регресс, LARS, Бэйсийн VAR зэрэгтэй харьцуулсан. 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 хувьсагчидын сонголтын талаар тодорхой бус үед Рао Блэквеллийн теорем ба Багингын дагуу таамаглалын загвараа сонгож гүйцэтгэлийг харьцуулах судалгаа хийсэн. Тэд загвар сонголт, параметрийн тооцоо хийхэд зориулж янз бүрийн схемүүдийн тархалтын шинж чанарыг судалж үзсэн: Akaike мэдээллийн шалгуурыг ашиглан түүвэрт тулгаарлан 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 тааруухан 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 Яанг (2006) нар баггинг алгоритмийн хэрэглээг хугацааны цуваа, бүр цаашилбал чанарын хувьсагч, квантилыг таамаглахад ашиглаж өргөтгөсөн. Чанарын таамаглал хийхийн тулд квантил тооцоолол, давамгай санал бүхий бинар таамаглалыг ашигласан бөгөөд ийм тохиолдолд баггинг аргазүй сайн гүйцэтгэлтэй байх боломжтойг харуулсан. Эмпирик хэрэглээний хувьд тэд сарын давтамжтай S&P500 болон NASDAQ хувьцааны индексийн өгөөжийг ашиглан үр дүнг танилцуулсан. (Lee, T.-H., & Yang, Y, 2006)

Иноуе, Килиан (2008) нар АНУ-ын ХҮИ-ийн хугацаан цуваан таамаглал хийхдээ багинг аргазүйг авч үзсэн байна. Тэд инфляцийн таамаглахдаа динамик шугаман регрессд баггинг тохирч болох талаар судалж үзсэн. Ингэхдээ хэд хэдэн загваруудыг баггингтай болон баггинггүй харьцуулж үзсэн бөгөөд үүнд регрессийн загварууд, фактор загвар болон агшаасан регрессийн загваурууд (ЛАССО-той) хамаарна. Тэдний эмпирик үр дүн нь баггинг нь инфляцийг 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 параметрын регрессийн таамаглалын загваруудыг авч үзсэн. Судалгааны зорилго нь эдийн засгийн онол болон регрессийн функцээс гарч ирэх чухал эерэг эсвэл монотоник шинж чанартай эдийн засгийн олон төрлийн хязгаарлалтуудыг 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 холбогдолыг харуулсан. Баггинг хязгаарлалттай үнэлэгчид болон таамаглалын давуу талыг Монте-Карлогийн өргөн цар хүрээтэй симуляцаар харуулсан болно. Санхүүгийн хөрөнгийн хураамжийг урьдчилан таамаглах эмпирик судалгаанд баггинг аргазүйг ашигласнаар том урт хугацааны таамаглалын алдаа гаргах боломжийг багасгаж, хязгаарлалт бүхий таамаглалын үнэлгээг илүү тогтвортой болгож байсан. (Lee, T.-H., Tu, Y., & Ullah, A, 2015)

Жин, Су, Уллах (2014) түүврийн бус таамаглал хийхдээ баггин 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. 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 харьцангүй үр ашигтай байгааг ажигласан бөгөөд түүврийн бус таамаглал хийхдээ баггинг аргазүйг ашиглаж илүү тогтвортой үр дүнд хүрэх боломжтой санал болгосон. 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 бусад таамаглагч хувьсагчдыг гүйцэтгэлээрээ давж байгааг олж мэдсэн юм. Уламжлалт нэг шугаман урьдчилсан болон параметрийн бус прогнозын 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 инфляци 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)

## Машин сургалтын алгоритм түүний хэрэглээ

### Шинэ 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 (энэ сэтгүүлд, 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   -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.

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 2016) Жан нар. (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.

Хэл нь мэдээллийн өөр нэг шинэ хүчирхэг эх сурважийн нэг билээ. Хиймэл дагуулын зургын 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.

Санхүүчид болон эдийн засагчид Компюстат (дэлхийн хэмжээнд идэвхтэй, идэвхигүй байгаа компаниудын санхүүгийн, статистик, зах зээлийн мэдээллийн мэдээллийн сан) дээрх байгууллагын санхүүгийн мэдээлэлд ихээхэн н айддаг. 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 туршилтанд 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.

Үйлчилгээний үнэлгээг таамаглах

# ОНОЛЫН УХАГДАХУУН БА ЗАГВАР

## Шийдвэрийн модны аргазүй

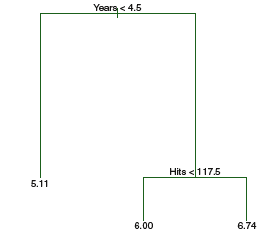
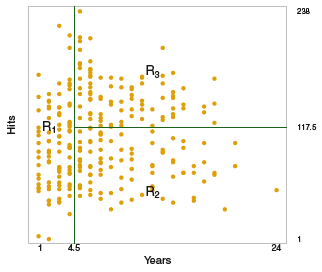
Өгөгдсөн ажиглалтын тусламжтай үзэгдлийг таамаглахдаа бид хуваасан бүлгүүдийн дундаж болон моодыг ашигладаг. Өгөгдлийг тодорхой хэсгүүдэд хуваах дүрмийг дээрээс доош салаалах модоор хураангуйлдаг бөгөөд үүнийг шийдвэрийн мод хэмээдэг. Шийдвэрийн модонд суурилсан аргазүй нь энгийн, тайлбарлахад хялбар 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.

### Регресийн 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 мөчирөөр 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 тоголж буй жил, өнгөрсөн жил гүйцэтгэсэн цохилтын тоо нөлөөлдөг гээд шийдвэрийн модыг дүрсэлвэл доорх мод үүснэ. Зурагт регрессийн мод өгөгдөлд хэрхэн тохирч байгааг болон модыг оройгоос салбарлуулан хуваах дүрмийг харж болно. нь өмнө дурдаж байсан эцсийн зангилаа буюу модны навчийг илэрхийлэх ба модны эцэст байрлана. Дотоод зангилаа болох “” болон “” нь өгөгдлийг гурван бүлэгт хувааж байгаа юм. Ингээд таамаглалын загвар бэлэн боллоо. Тухайлбал лигт 4.5 болон түүнээс дээш жил тоголсон, өнгөрсөн улирал 117.5 – оос бага цохилт хийсэн тамирчиний цалин 3 дугаар навч буюу 6 хэмээн таамаглагдаж байна.

Зураг Шийдвэрийн мод



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

Энд нь дүгээр бүлэгт буй хамааран хувьсагчидын 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 болон түүнийг болон гэсэн 2 хэсэгт хуваах 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.

#### 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 cost Complexity тайралт эсвэл хамгийн сул холбоос тайралт хэмээн нэрлэдэг. Боломжит бүх модыг сонгохын оронд бид сөрөг биш утгатай параметрээр индексжүүлсэн дэд моднуудыг авч үзнэ. дэд модонд харгазах нь дараах тэгшитгэлийг хамгийн бага утгатай байхаар сонгогддог.

Энд нь модны эцсийн зангилаа, нь дүгээр эцсийн зангилаанд харгалзах тэгш өнцөгт, нь тэгш өнцөгт буюу дүгээр сургалтын өгөгдлийн дундаж буюу таамаглагдсан утга. Чухамдаа параметр нь сургалтын өгөгдөлийн тохирсон байдал болон дэд модны Complexity хоёрын ацан шалааг зохицуулдаг. Хэрвээ бол болох буюу дээрх тэгшитгэл нь сургалтын өгөгдлийн алдааг хэмжих юм. Бид -ийн утыг cross- validation болон валидэйшн аргуудын тусламжтай олох ба энэ утгад харгазах дэд моднуудыг өгөгдлөөс сонгон авна. Алхамуудыг Алгоримт 8.1-д харуулав.

Алгоритм 8.1 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 функцын тусламжтай тайрч засан хамгийн шилдэг дэд модны цувааг олж авна.
3. K-дахин cross validation – г ашиглан – ийн утгыг олно. Энэ нь сургалтын өгөгдлийг K ширхэг бүлэгт хуваана гэсэн үг. байх бүлэгийн хувьд:
   1. 1 болон 2 дугаар алхамыг К дугаар бүлгээс бусал хэсэгт давтан гүйцэтгэнэ
   2. функцийн хувьд өгөгдлөөс үлдсэн K дугаар бүлгийг ашиглан таамаглалын алдааны квадратыг үнэлнэ

функцын тусламжтай үр дүнг дундажлах замаар дундаж алдааг хамгийн бага байлгах – г сонгон авна.

1. Сонгогдсон – д харгалзах дэд моднуудыг олж авна.

### Ангиллын мод

Ангиллын мод нь регрессийн модтой бараг л ижилхэн боловч тоон биш чанарын хувьсагчийг таамаглах болно. Регрессийн модыг санавал бид нэг ижил эцсийн зангилаад хамаарах ажиглалтын хувьд хамааран хувьсагчийн тооцож, дундажлаж замаар таамаглал хийж байсан. Харин ангиллын модны хувьд навч буюу эцсийн зангилаа бүрийн хувьд хамгийн их давтагдсан ангилалын 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 ангилалыг хувуурилах болно, харин ангиллын алдааны түвшин нь тухайн бүлгийн ажиглалтын тоо болон хамгийн их давтагдсан ангилалд хамаарахгүй ажиглалтуудын тооны харьцаа юм.

Энд нь дүгээр бүлэгт дугаар ангилалын 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 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.

Дараагийн боломжит арга бол хөндлөн энтропи гэж нэрлэгдэх хэмжүүр :

Энд болон нөхцөл биелэнэ. Хэрвээ бүх – ийн утга 1 эсвэл 0 – тэй ойрхон бол хөндлөн энтропи бага утгатай байна. Жинийн индекстэй ойролцоо гарах энэхүү хэмжүүр нь мөн л 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 түвшингээс илүү мэдрэмтгий байдаг. Модыг тайрахад эдгээр гурван аргуудын аль нэгийг нь ашиглаж болно, гэхдээ эцсийн тайрсан модны зорилго нь таамаглалын нарийвчлал бол ангиллын алдааны түвшинг ашиглах нь илүү юм.

#### Шийдвэрийн модны сул ба давуу тал

* Давуу тал
  + 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.
* Сул тал
  + Бусад машин сургалтын аргуудтай харьцуулахад нарийвчлал бага, вариац өндөр

## Баггинг

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

Бие биеэсээ үл хамаарах, вариацтай ширхэг үл хамаарах ажиглалт өгсөн гэвэл дундаж нь болох – ийн вариац 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 гэсэн ширхэг бие даасан сургалтын өгөгдөл гаргаж авсан гэвэл бага вариацтай статистик сургалтын загварыг дараах байдлаар олно:

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

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 ширрэг таамаглалд хамгийн их тохиолдох ангилалыг тооцох аргыг ашигладаг.

**Цүнхнээс гадна буй алдааны тооцоолол**

Баггинг ашигласан загвырын хувьд тестийн өгөгдлийн алдааг шууд тооцох арга байдаг буюу cross-validation, 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.

# ЭМПИРИК СУДАЛГААНЫ АРГА, АРГАЗҮЙ

Аргазүйн хэсэг нь бүхэлдээ Монте-Карло симуляцийн агуулга, үйл явцыг хамарна. Нэгдүгээрт, гурван өөр ӨБҮ-аас хугацаан цувааны өгөгдлийг хэрхэн үүсгэх талаар авч үзэх болно. Хоёрдугаарт, бид эдгээр таамаглалын гүйцэтгэлд хугацааны цувааны таамаглалын урт, үлдэгдэл санамсаргүй хэмжигдэхүүний тархалт, загварын эрэмбэ хэрхэн нөлөөлөхийг тооцно. Гуравдугаарт, бид ашиглах баггинг машин сурах алгоритмыг хэрхэн ашиглах талаар нарийвчлан авч үзэх болно.

## Өгөгдөл боловсруулах үйл явц (ӨБҮ)

Бид судалгаандаа практикт хамгийн түгээмэл ашиглагддаг, хэвийн тархсан үлдэгдэл санамсаргүй хэмжигдэхүүн бүхий авторегрессив (АЗ), theroshold авторегрессив загвар (TAR) болон Ерөнхийлсөн авторегрессив нөхцөлт хетероскедастик (ЕАНХ) процессүүдийг

# ЭМПИРИК СУДАЛГАА

# ДҮГНЭЛТ, САНАЛ

# ХАВСРАЛТ

# АШИГЛАСАН МАТЕРИАЛ

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.

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.

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

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.

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.

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.

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