**САНХҮҮ ЭДИЙН ЗАСГИЙН ИХ СУРГУУЛЬ** 

**ӨДРИЙН ХӨТӨЛБӨР**

ЭКОНОМИКСИЙН ТЭНХИМ

**ДИПЛОМЫН АЖЛЫН ТӨСӨЛ**

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

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

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

Огноо: 2020.01.07



Улаанбаатар. 2020

**АГУУЛГА**

[I БҮЛЭГ. СУДАЛГААНЫ АСУУДАЛ 1](#_Toc29300290)

[II БҮЛЭГ. СУДЛАГДСАН БАЙДАЛ 2](#_Toc29300291)

[III БҮЛЭГ. СУДАЛГААНЫ ЗОРИЛГО, ЗОРИЛТ 4](#_Toc29300292)

[IV БҮЛЭГ. СУДАЛГААНЫ ТААМАГЛАЛ 5](#_Toc29300293)

[V БҮЛЭГ. СУДАЛГААНЫ АРГАЗҮЙ 6](#_Toc29300294)

[5.1 Олон хүчин зүйлийн регрессийн шинжилгээ 6](#_Toc29300295)

[VI БҮЛЭГ. СУДАЛГААНЫ ӨГӨГДӨЛ 9](#_Toc29300296)

[VII БҮЛЭГ. СУДАЛГААНААС ХҮЛЭЭЖ БУЙ ҮР ДҮН 10](#_Toc29300297)

[VIII БҮЛЭГ. АШИГЛАСАН МАТЕРИАЛ 11](#_Toc29300298)

**ЗУРГИЙН ЖАГСААЛТ**

[Зураг 1 Тодорхой бус байдал ба аргачлалууд 3](#_Toc29284149)

**ТОВЧИЛСОН ҮГС, НЭР ТОМЪЁОНЫ ТАЙЛБАР**

ДНБ Дотоодын нийт бүтээгдэхүүн

ХҮИ Хэрэглээний үнийн индекс

ҮСХ Үндэсний статистикийн хороо

# СУДАЛГААНЫ АСУУДАЛ

Машин сургалтыг маш олон салбарт ашиглах болсон бөгөөд эдийн засгийн таамаглалд ч мөн ашиглаж байгаа юм. шийдвэрийн модны аргазүй, бүүсттрап, санамсаргүй ойгүүжүүлэлт гэх мэт алгоритмуудыг эдийн засагчид өргөн ашиглаж байгаа билээ. Ялангуяа Брэймэний “Баггинг аргазүй”(Breiman, 1996) хэмээх бүтээлд дурдсан түүний техникийн таамаглалын гажуудлыг нэмэгдүүлэхгүйгээр таамаглалын хэлбэлзлийг бууруулах шинж чанар нь машин сургалтын давуу талыг харуулж байгаа юм.

Таамаглалын модель сонголтын загваруудыг симуляцын аргаар үнэлэн, гүйцэтгэлийг хэмжих судалгаа түгээмэл байдаг билээ. Эдгээр судалгааны хэрэгцээ шаардлага нь загварын тодорхой бус байдал юм. Оновчтой таамаглалыг хийх асуудал нь загварын тодорхой бус байдалд бидний ашигладаг уламжлалт загварууд тийм ч сайн үр дүнг өгөхгүй бөгөөд судлаачид олон төрлийн аргуудыг ашигладаг. Эдийн засгийн таамаглалд үндсэн гурван төрлийн тодорхой бус байдал байдаг гэж үздэг бөгөөд загварын тодорхой бус байдлын тухайд бүх загвар буруу бөгөөд зөвхөн нэг загварыг сонгох нь учир дутагдалтай юм. Гэвч зарим загварууд нь муугүй ба эдгээрийг үр ашигтай нэгтгэж чадвал таамаглалын гүйцэтгэлийг үлэмж хэмжээгээр нэмэгдүүлдэг. (Box, G. E., & Draper, N. R, 1987) Чухамдаа судлаачид таамаглалуудыг нэгтгэх олон төрлийн арга ашигладаг бөгөөд үүнд симуляцийн арга техникийн нэг болох баггинг нь загварын тодорхой бус байдлыг устгадаг онцлогтой.

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 энэ чиглэлийн судалгаа шинжилгээний хоосон орон зай байгааг илэрхийлж байна.

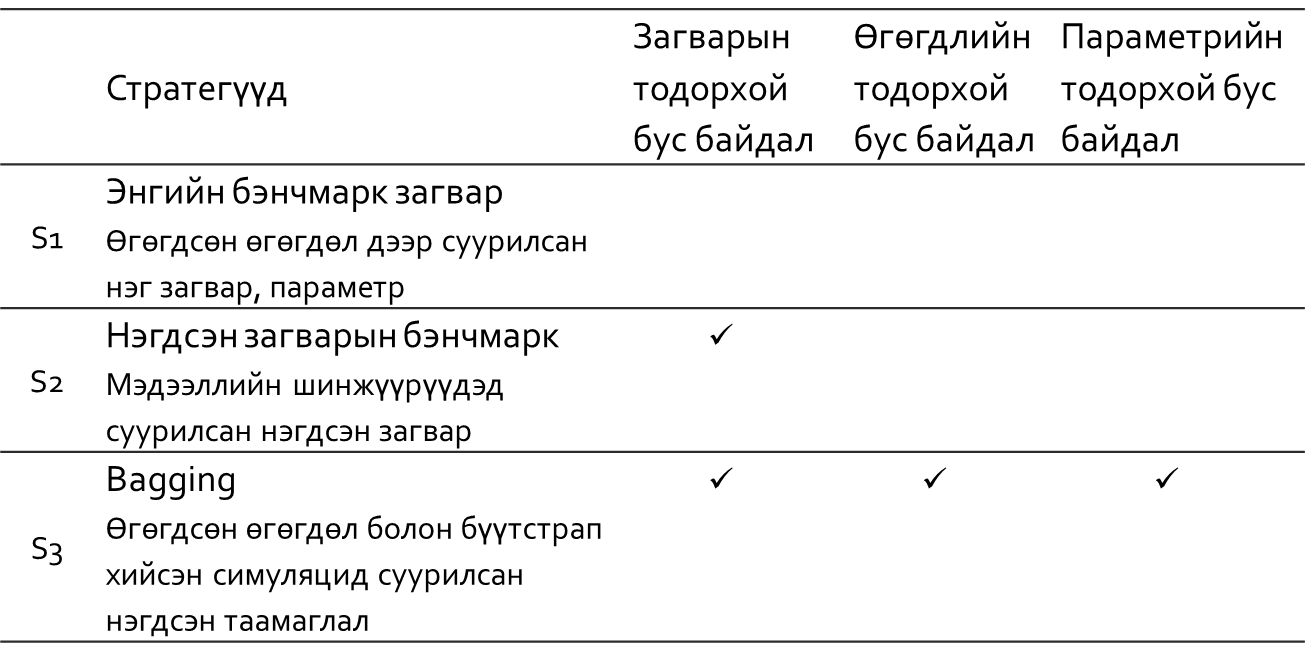
# СУДЛАГДСАН БАЙДАЛ

Энэхүү бүлгээр загварын тодорхой бус байдал таамаглалд ямар асуудлыг авчирдаг, түүнийг шийдвэрлэх ямар арга замууд байдаг тухай авч үзэх болно.

Баггинг аргазүйг Брэймэний “Баггинг аргазүй”(Breiman, 1996) хэмээх бүтээлд дурдсан ба таамаглалын гажуудлыг нэмэгдүүлэхгүйгээр хэлбэлзлийг бууруулах аргазүй хэмээн тодорхойлсон байдаг. Тэрээр шийдвэрийн модны суурийг ашиглан тавин удаа бүүтстрап симуляци хийсэн бөгөөд шинжилгээний үр дүнд 2 зүйлийг онцолсон байна. Нэгдүгээрт бэнчмарк загвар оновчтой биш үед баггинг аргазүй нь таамаглалын нарийвчлалыг сайжруулдаг. Дараагийнх нь эхнийхтэй ижил буюу бэнчмарк загварын (шугаман регрессийн) таамаглал оновчтой бол баггинг аргазүйн таамаглалын алдаа харьцангуй их байна.

Баггинг аргазүйг хугацааны цуваан өгөгдөлд яагаад ашигладаг талаар судалгааны ажлыг Хиндман, Бэргмэр болон Петропаулас хийсэн бөгөөд энд гурван төрлийн тодорхой бус байдлыг авч үзсэн бөгөөд эдгээр тодорхой бус байдалд зургаан төрлийн таамаглалын стратегиуд хэрхэн ажиллаж байгаад үндэслэн харьцуулан дүгнэлт өгсөн байна. Тодорхой бус байдлын хувьд өгөгдлийн тодорхой бус байдал, загварын тодорхой бус байдал болон параметрийн тодорхой бус байдалд хэрхэн үнэлж байгааг дараах хүснэгтээр харуулав. (Petropoulos, F., Hyndman, R. J., & Bergmeir, C., 2018)

*Зураг 1 Тодорхой бус байдал ба аргачлалууд*



Макроэкономиксийн томоохон үзүүлэлтүүдийг таамаглах нь их хэмжээний бэрхшээлийг авчирдаг. Ихэвчлэн нэгдсэн загварыг ашигладаг бөгөөд жинлэх аргазүйгээсээ хамаарч бага зэргийн ялгаа байдаг. АНУ-ын ажил эрхлэлтийн түвшинг баггинг (динамик ерөнхий тэнцвэрт регрессийн загвар) болон авторегрессив тархсан хожимдол (ARDL) бүхий загварыг ашиглан таамагласан байна. Судалгааны үр дүнд баггинг аргазүй нь харьцангуй бага MSFE-тэй байсан. Мөн таамаглалын нарийвчлалыг сайжруулж байсан нь үр дүнг хэлж болно.

**Шийдвэрийн модны давуу ба сул тал:**

* Шийдвэрийн мод нь үр дүнг энгийн хүмүүст тайлбарлахад хялбар
* Практикт хувь хүний шийдвэр гаргалттай ижил
* Чанарын хувьсагчийг ашиглахад хялбар

**Сул тал:**

* Бусад аргуудтай харьцуулахад нарийвчлал бага

# СУДАЛГААНЫ ЗОРИЛГО, ЗОРИЛТ

Энэхүү судалгааны гол зорилго нь таамаглалын загваруудыг симуляцийн аргаар шинжлэн, харьцуулах билээ. Ялангуяа эдийн засгийн голлох үзүүлэлтүүдийг 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 ТААМАГЛАЛ

Судалгааны эх сурвалж, судлагдсан байдлаас үндэслэн дараах таамаглалыг дэвшүүлэв.

**Таамаглал 1.** Таамаглалын загвар тодорхой бус үед баггинг аргыг ашигласан таамаглал нь илүү нарийвчлалтай байна

**Таамаглал 2.** Таамаглалын загварын тодорхой бус тохиолдолд баггинг аргазүй нь илүү хэлбэлзэл багатай байна.

# СУДАЛГААНЫ АРГАЗҮЙ

Энэ бүлэгт бэнчмарк загвар буюу уламжлалт шугаман регрессийн аргазүйн талаар авч үзнэ.

## Олон хүчин зүйлийн регрессийн шинжилгээ

Хамааран хувьсагчид нөлөөлдөг бусад олон хүчин зүйлийг хянах боломжийг олгодог учир бусад хүчин зүйл тогтмол нөхцөлд олон хүчин зүйлийн регрессийн шинжилгээ илүү тохиромжтой байна. Энэ нь туршилтын бус өгөгдөл ашиглах шаардлагатай болсон тохиолдолд эдийн засгийн онолын таамаглал шалгах болон бодлогын нөлөөг үнэлэхэд чухал нөлөөтэй. Учир нь олон хүчин зүйлийн регрессийн шинжилгээ нь харилцан хамааралтай байж болох олон тайлбарлагч хувьсагчийг багтаадаг тул энгийн регрессийн шинжилгээгээр илрүүлж чаддаггүй асуудлын учир шалтгааныг олж мэдэхэд илүү үр дүнтэй байдаг.

Олон хүчин зүйлийн регрессийн шинжилгээнд Гаусс – Марковын үндсэн 5 таамаглал биелсэн тохиолдолд ХБКА – аар үнэлэхэд үнэлэгч нь BLUE1 шинжийг хангана. Харин сонгодог шугаман регрессийн таамаглалд “үлдэгдэл хэвийн тархалттай байх” хэмээх 6 дахь таамаглалыг нэмж оруулдаг. Параметрүүдийн үнэлэгчдийн тархалт нь үлдэгдэл санамсаргүй хэмжигдэхүүний тархалтаар шууд тодорхойлогдох учир хэвийн тархалтын таамаглал биелж байхад ХБКА – ын үнэлэгч хамгийн сайн гажуудалгүй үнэлэгч байна.

Хөндлөн өгөгдөл ашиглан олон хүчин зүйлийн регрессийн шинжилгээ хийхэд дараах хэд хэдэн асуудлыг анхаарах шаардлагатай. Үүнд:

1. Мультиколлинеарын асуудал,
2. Хетероскедастикийн асуудал,
3. Орхигдсон хувьсагчийн гажуудал
4. Эндогенийн асуудал
5. Хэмжилтийн алдаа зэрэг багтана.

**Мультиколлинеарын асуудал**

Тайлбарлагч хувьсагч хоорондоо өндөр хамааралтай болсон үед мультиколлинеарын асуудал үүсдэг ба үнэлэгчийн эрчимтэй байх шинж чанарыг алдуулдаг. Иймд тайлбарлагч хувьсагчдын коэффициентуудын вариац их болж коэффициентын ач холбогдол буурна. Уг асуудлыг илрүүлэхийн тулд хоёр нөхцөлийг шалгаж үзнэ.

Үнэлгээний тэгшитгэлийг бүхэлд нь үлдэгдлийн стандарт алдаанд хувааж нормчлох буюу жинлэсэн хамгийн бага квадратын аргыг ашиглах замаар хетероскедастикийн асуудлыг засаж үнэлгээг дахин хийнэ.

**Орхигдсон хувьсагчийн алдаа**

Эмпирик судалгааны эконометрик үнэлгээний хувьд тайлбарлагдагч хувьсагчид нийгэм, эдийн засаг, соёл, улс төрийн орчны зүгээс нөлөөлж болох хувьсагчдыг бүгдийг нь багтаах боломжгүй учир орхигдсон хувьсагчийн алдаа үүсэх магадлал өндөр юм. Энгийн регрессийн шинжилгээтэй харьцуулахад илүү олон хувьсагчийг үнэлгээнд оруулах боломжтой учир олон хүчин зүйлийн регрессийн шинжилгээнд орхигдсон хувьсагчийн алдаа гарах магадлал харьцангуй бага боловч мэдээллийн ба өгөгдлийн хязгаарлалтаас шалтгаалан үнэлгээнд оруулах боломжгүй хувьсагчид их байдаг.

**Эндогенийн асуудал**

Тайлбарлагч хувьсагч болон үлдэгдэл утгын хооронд өндөр хамааралтай үед энэхүү асуудал үүсдэг. Энэ нь эконометрик үнэлгээний хазайлтгүй шинж чанарыг алдагдуулдаг бөгөөд тайлбарлагч болон тайлбарлагдагч хувьсагчдын хамаарлын тогтвортой байдлыг алдагдуулах гол хүчин зүйл болно.

**Хэмжилтийн алдаа**

Үнэлгээнд ашиглагдаж буй эдийн засгийн болон институцийн хувьсагчдын тохиромжгүй хэмжилт нь хэмжилтийн алдаа гарах үндэс болно. Хэмжилтийн алдааны үр дагаварт үнэлгээ гажуудалтай, нийцгүй болно.

Орхигдсон хувьсагчийн гажуудал, хэмжилтийн алдаа, эндогенийн асуудлыг засах нэг арга нь туслах хувьсагчийн арга буюу хоёр шатат ХБКА юм. Энэ талаар дэлгэрүүлэн авч үзье.

**Туслах хувьсагчийн арга буюу хоёр шатат ХБКА**

Туслах хувьсагчийн аргын тухай анхны ойлголтыг Филип Г. Райт , түүний хүү Сэвалл Райтын хамт 1928 онд бичсэн “Амьтан ба ургамлын гаралтай тосны тариф” хэмээх номондоо танилцуулжээ. (Stock & Trebbi, 2003)

Өмнө нь дурдсанчлан статистик болон эконометрикийн шинжилгээнд орхигдсон хувьсагчийн алдаа, хэмжилтийн алдааг болон эндогенийн асуудлыг засварлах зорилгоор туслах хувьсагчийн аргыг ашигладаг. Уг аргыг ашигласнаар үнэлгээний нийцтэй шинж чанарыг хадгалах боломжтой юм. Үнэлгээнд ашиглагдах туслах хувьсагч нь орхигдсон хувьсагч болон бусад тайлбарлагч хувьсагчтай хамааралгүй, зөвхөн эндоген тайлбарлагч хувьсагчтай хамаарал бүхий экзоген хувьсагч байх ёстой.

Тэгшитгэл зөвхөн нэг туслах хувьсагчтай тохиолдолд ХБКА – аар регрессийг үнэлэх боломжтой. Харин үнэлгээний тэгшитгэлд нэг бус нэлээд олон тооны туслах хувьсагчийг ашиглах шаардлага тохиолддог. Олон туслах хувьсагч бүхий тухайн тохиолдолд регрессийн тэгшитгэлийг үнэлэх тооцооллын арга нь хоёр шатат ХБКА юм.

* 1. **Шийдвэрийн модонд суурилсан аргазүй**

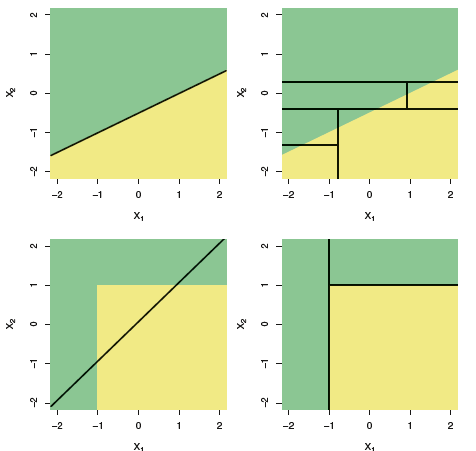
Шийдвэрийн модны аргазүй нь регресс (regression) болон ангилалд (classification) ашиглагддаг. Шийдвэрийн модны онолыг авъя үзвэл дараах зураг 2-т харуулснаар болон нь модны навч бөгөөд гэсэн гэсэн 2 дотоод зангилаа байна. Навч дотоод зангилаа нь мөчрөөр холбогдоно.

*Зураг 2 Шийдвэрийн мод*

|  |  |
| --- | --- |
|  |  |

Зураг 3-т шугаман болон шугаман бус нөхцөлд шугаман болон шийдвэрийн модны аргазүйн үнэлгээний үр дүнг харууллаа.

*Зураг 3* *Шугаман регрэсс ба шийдвэрийн мод*



# СУДАЛГААНЫ ӨГӨГДӨЛ

Машин сургалтын аргыг ашиглахдаа макроэкономиксийн голлох үзүүлэлтүүдийг ашиглаж өгөгдлийн санг бүрдүүлнэ. Ингэхдээ нобелийн шагналт, нэрт эдийн засагч Милтон Фрэйдманы хэлснээр хамгийн чухал үзүүлэлтүүд болох эдийн засгийн өсөлт, үнийн тогтвортой байдал, ажилгүйдлийн түвшин гэх хүршгүй гурван оргил гэсэн үзүүлэлтүүд дээр судалгааг хийнэ.

Уг судалгаанд эдийн засгийн өсөлт, үнийн тогтвортой байдал болон ажилгүйдлийн түвшинг ДНБ, инфляци мөн ажил эрхлэлт гэсэн хувьсагчуудаар илэрхийлж, ҮСХ-оос бэлтгэн хүргэдэг хоёрдогч мэдээллийн сангаас сонгон авна.

* Дотоодын нийт бүтээгдэхүүн (Gross domestic product)

Дотоодын нийт бүтээгдэхүүн (ДНБ) нь тухайн нутаг дэвсгэрийн хүрээнд дотоод, гадаадын аж ахуйн нэгж, байгууллага, иргэдийн нэг жилийн хугацаанд шинээр бий болгосон нийт нэмэгдэл өртгийн хэмжээ юм. МУ-ын хувьд 1990 оноос хойш тооцож эхэлсэн бөгөөд улирлын давтамжтайгаар тооцож гаргадаг.

* Хэрэглээний үнийн индекс (Consumer price index)

ҮСХ-оос 1991 оноос хойш жил бүр хийгдэж буй инфляцын түвшин буюу хэрэглээний үнийн индекс нь хэрэглэгчдийн худалдаж авсан бараа, үйлчилгээний нэр төрөл, чанар өөрчлөлтгүй тогтвортой байхад үнэ дунджаар хэрхэн өөрчлөгдөж буйг хэмжих үзүүлэлт бөгөөд нийт өрхөөр тооцсон үнийн өөрчлөлтийг харуулах үзүүлэлт юм.

* Ажилгүйдлийн түвшин ба хөдөлмөр эрхлэлт (unemployment rate & labour force)

Ажилгүйдлийн түвшин нь ажиллах хүчний хэдэн хувийг ажилгүйчүүд эзэлж байгааг харуулдаг үзүүлэлт юм. ҮСХ-оос 2006 оны 3 сараас хойш сар бүр гаргаж буй үзүүлэлт юм.

Өгөгдлийг шинжлэх программ хангамжийн тухайд Rstudio программыг ашиглана.

# СУДАЛГААНААС ХҮЛЭЭЖ БУЙ ҮР ДҮН

Судлагдсан байдалд үндэслэн симуляцийн аргаар үнэлсэн энэхүү таамаглалын алдаа бидний уламжлалт аргаар үнэлсэн таамаглалаас илүү нарийвчлалтай, алдаа бага байна хэмээн хүлээж байгаа билээ. Энэхүү үр дүн нь таамаглалын загварын тодорхой бус байдал бий болсон нөхцөлд биелэнэ гэдгийг онцлох шаардлагатай. Өөрөөр оновчтой таамаглалыг баггинг аргаар симуляцын хийх замаар үнэлэх нь үр ашигтай.

# АШИГЛАСАН МАТЕРИАЛ

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*, 123-140.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association*, 511-522.

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.