

### Tillaga að próflestraráætlun

0. Lesa vel yfir prófið í hagrannsóknnum I. Það þarf að kunna skil á þeim hugtökum sem þar koma fyrir.

(a) Það þarf að kunna grunnatriði um línuleg líkön. Þ.e. hugtök sem tengjast mati, unbiased/efficient/consistent o.s.frv. Likelihood teoríu, ML-mat, Cramer-Rao ójöfnu, LR/Wald/LM prófanir, o.s.frv.

1. Kaffi 13-14.

Tímaraðir eru mikilvægt hugtak. Mikið af haggögnum eru á formi tímaraða og því mikilvægt að skilja hugtök sem þeim tengjast. Helstu hugtök eru síðstæðni ( $I(0)$ , stationarity og ergodic), sjálffylgni (autocorrelation). Ákveðið form á ósístæðni eru „unit-root“ ( $I(1)$ ). Það hugtak þarf að skilja. Í öllum tölfræðilegum prófum er verið að bera saman tvö líkön. Það er mikilvægt að skilja hvaða tvö líkön er verið að bera saman. Viðmiðunin, random-walk, er ekki endilega viðeigandi í öllum tímaröðum. Raðir eru ekki gerðar stationary með því að taka mismun. Það er hugsanlegt að fyrsti mismunur sé stationary. Hér þarf að lesa kaffa 13 og 14 í bók.

(b) Skilningur á þessum hugtökum er grundvöllur þess að hægt sé að setja upp tímaraðalíkan. Öll gagnagreining fer fram í gegnum líkön, DGP=data-generating-process. Það þarf að hugleiða hvaða kraftar gerðu það að verkum að gögnin eru eins og þau eru. Í vetur voru tekin fyrir spurious-regression=dellu-aðhvarfsdæmi. Það voru bæði hermunardæmi, fylgni

reiknuð milli tveggja random-walk og dæmi með raunverulegum gögnum (gögnin um lífshættuna og Ensku biskupakirkjuna). Hér þarf að lesa kafla 13 og 14 í bók. Einnig þarf að skilja hvað lá að baki verkefnunum á dæmablöðunum.

- (c) Það þarf að skilja hvað ECM hugtakið er. Í einni vítt má segja að þetta sé

$$\Delta X_t = \beta(X_{t-1} - \mu) + \varepsilon_t.$$

Ef  $\beta$  er neikvæð tala er það túlkað þannig að við spáum pösítívri breytingu á  $X_t$  ef  $(X_{t-1} - \mu)$  er neikvæð tala. Þ.e. krafturinn togar í áttina að  $\mu$ . Í ECM kemur önnur breyta í stað  $\mu$ .

- (d) Ýmsar útfærslur á tímaraðalíkönunum. Einvíðu AR, MA, ARMA og ARIMA útgáfurnar geta náð hreyfimyndri ýmissa hagraða. Það hreyfimyndur hefur vissa samsvörun við venjulegar diffurjöfnur.
- (e) Margvíðar útgáfur, VAR (VARMA) lýsa hreyfimyndri sem svarar til margvíðra diffurjafna. Umgengni við margvíð líkön er mun flóknari en umgengni við einvíð líkön.
- (f) Það þarf að skilja hvernig co-integration og ECM tengjast í VAR líkani. Það þarf að geta lesið úr VAR líkani sem hefur verið metið með aðferð Johansen.
- (g) ARCH/GARCH hugmyndin er að raðir sem virðast white-noise eru ekki endilega raðir af óháðum mælingum. Hugsanlegt er að stóru breytingarnar komi í gusum. Það er nauðsynlegt að þekkja þetta fyrirbæri.

2. Það eru til ýmis margvíð líkön sem ekki eru endilega tímaraðalíkön. Þ.e. í þeim er einhvers konar öðruvísi

tengslakerfi. Hér er átt við það sem stendur í kafla 12. Til að geta lesið þann kafla þarf að vera búið að meðtaka atriði sem koma fyrir í ýmsum köflum framar í bókinni. T.d. í köflum 7-9.

- (a) SUR=seemingly-Unrelated-Regression. Hvað er það sem seemingly unrelated? Hvernig á að meta þannig líkön á skilvirkan hátt? Hvað er vandamálið við skilvirkt mat á SUR?
- (b) Simulatan-kerfi var eitt sinn fyrirferðarmikið hugtak í hagrannsóknnum. Það þarf að skilja hvað er það sem er simultan. Hvers vegna er ekki hægt að meta líkan á því formi sem maður hefur áhuga á. Hvers vegna gefur regression af magni á verð ekki consistent mat á responsi í magni við verðbreytingu?
- (c) Hvaða aðferðir koma til greina við að meta simultan líkön. Hér þarf að kunna eina aðferð (2STLS) og vita af öðrum (t.d. 3STLS). Hvaða skilyrði þurfa að vera uppfyllt til að hægt sé að meta svona líkan?
- (d) Panel-data líkön. Hvers vegna viljum við hafa margar mælingar á hverjum einstakling? Eru skýristærðir exogen? Hver er munurinn á FE-líkani og RE-líkani? Hvað er venjulega verið að prófa þegar talið er um Hausman próf?
- (e) GMM=Generalised-Method-of-Moments. Hvað er Method-of-Moments? Hvað er það sem er general? Hér getur verið gott að lesa yfir ljósrit úr Hamilton og skilja verkefni. R-forrit og excel-skjal sem sýnt var í námskeiðinum eiga að skýra hugmyndina. GMM má nota við mjög margar tegundur af líkönum og útfærslurnar mjög mismunandi.

3. Ólínuleg líkön, kaflar 6 og 11. Hvað þýðir það að líkan sé ólínulegt? Hvers vegna er t.d. LS-aðferðin óaðgengilegri í ólínulegum líkönum. Tæknilegur vandi í umgengni við ólínulegu líkön er meiri en í línulegum. Tækni sem kynnt var er aðferð Newton. Hún byggir á Taylor útvíkkun og einhvers konar fixed-point rökfræði. Hér dugar að þekkja til aðferðar Newton og geta tekið eitt skref í henni. Höfundarnir erum með sniðuga útfærslu á henni sem þeir kalla artificial regression. Í vetur var sýnt dæmi um útfærslu á artificial regression. Það dugar að þekkja til aðferðar Newton, artificial regression er tæknilegt aukaatriði.

Aðalatriði er að skilja hvernig á að túlka viðkomandi líkan. Fjölskyldan af ólínulegum líkönum er mjög stór og því eru venjulega mjög góðar ástæður fyrir vali á tilteknu formi.

- (a) Logit/probit (binomial líkan) líkön henta þegar skýra á 0/1 (já/nei) breytu. Hvers vegna er margvitt (multinomial-dreifing) probit líkan oft skynsamlegra en margvitt logit líkan? Hvernig á að meta þessi líkön?
- (b) Líkön þar sem háða breytan er jákvæð einkennir t.d. líkön um biðtíma (duration/survival). Hvaða líkindadreifingar koma til greina í slíkum líkönum? Þegar safnað er gögnum um duration/biðtíma er oft mikið um censoring. Hvað er censoring? Hvernig fer censoring inn í likelihood-fallið? Gögn fyrir venjuleg censoring geta einnig komið fyrir í venjulegum regression líkönum. Slíkt er kallað tobit-líkan. Það þarf að skilja hvaða vandamál tengjast tobit-líkönum.
- (c) Talningargögn koma víða fyrir í hagrannsóknum.

Poisson líkan er oft notað sem viðmiðunarlíkan. Af hverju? Hvað þýðir hazard rate? Hvað er overdispersion? Hvernig mætti leiðrétta ályktun fyrir overdispersion?

4. Tölvutæknin hefur gert ýmis reiknifrek töl tæknilega möguleg. Höfundar bókarinnar hafa mikin áhuga á reikningi og því eru þessi atriði á víð og dreif í textanum.
- (a) Í námskeiðinu var kynnt bootstrapping aðferðafræðin. Hvað er bootstrapping? Í hvaða tilfellum er hún áhugaverð?
- (b) Sýnt var dæmi um MSM=method-of-simulated moments. Hugmyndin er að þegar erfitt er að reikna:

$$\int x^r f(x) dx,$$

þá megi nálgast þetta integral með því að herma gildi úr dreifingunni og nota

$$\frac{1}{S} \sum_{j=1}^S x_j^r, \text{ sem nálgun.}$$

Er þessi aðferð alltaf nothæf þegar MM er nothæf?

- (c) Minnst var á Bayes aðferðir í fyrirlesturum og sýnt einfald dæmi. Grundvöllur Bayes aðferða byggist á Bayes heimspekinni um að það megi nota líkur til að mæla vissu. Tæknilega eru aðeins til formúlur fyrir nokkur einföld líkön. Nauðsynlegt er að beita hermun til að fá mynd af posterior dreifingunni.

Gangurinn er ca.

Geng út frá líkani,

$f(y|\theta)$ ,  $y$  er mæld breyta,  $f$  er þéttifall,  $\theta$  parameter,

Set fram fyrirfram skoðanir, mína hlutdrægni,

með þéttifalli „a priori“ dreifingarinnar,

$\pi(\theta)$ ,

safna mælingum

$y_1, \dots, y_n$ ,

Reikna „a posteriori“ dreifingu,

$\pi(\theta|y_1, \dots, y_n) \propto \pi(\theta)L(\theta|y_1, \dots, y_n)$ ,

þar sem  $L(\theta|y_1, \dots, y_n)$  er likelihood-fallið.

Ályktunin byggir síðan á að skoða a posteriori dreifinguna,  $\pi(\theta|y_1, \dots, y_n)$ . Vandinn er að þetta getur verið mjög flókin dreifing og aðeins í nokkrum tilfellum er hægt að para saman form á  $\pi(\theta)$  og  $f(y|\theta)$  þannig að þægileg formúla komi út. Þess vegna er gerð forrit sem herma a posteriori dreifinguna og henni er síðan lýst með empírískum aðferðum. Mikið notað í dag.

5. Í allri gagnagreiningu er nauðsynlegt að skilgreina og skilja líkan sem hentar því vísindalega vandamáli sem verið er að skoða. Þegar líkan hefur verið metið liggur alltaf fyrir var á því hvort „rétt“ (nothæft), líkan hafi verið skilgreint. Í kafla 15 eru reifuð ýmis atriði sem snúa að prófunum (diagnostics) á líkönum. RESET prófið fyrir fallformi er einfalt í notkun og því gott að hafa í huga. Skoðun á 3ja og 4ða mómenti eru líka gagnlega þegar verið er að athuga forsendur um normaldreifingu.

Þar er einnig nefnt hugtakið kernel-estimation. Þegar skilgreina og meta þarf líkan með endanlegum gögnum

Þá þarf alltaf annað hvort að festa parametrískt form eða setja einhvers konar smoothness þvingu. Kernel-estimation er óparametrísk aðferð þar sem sett hefur verið smoothness skilyrði, þ.e. metin kúrfa má ekki beygja of mikið.

## **Almennt um próflesturinn**

Það þarf að skilja bókina. Með skilning er átt við að skilja hugtökin sem tengjast eðli, mati og prófunum á líkönum. Tæknileg atriði eins og t.d. artificial-regression er auka-atriði (kemur sér vel ef einhver vill forrita aðferðir). Það þarf að átta sig á að hagrannsóknir eru að langmestu leyti „observational“ vísindi og ekki experimental vísindi. Þess vegna eru atriði eins og instrumental aðferðir og Hausman próf í panel-data líkönum einkennandi á því hvernig econometría er frábrúðgin margri annari tölfræðimeningu. Mikið af hagrænum gögnum er safnað í tíma og því eru tímaraðalíkon mikilvæg. Það að kíkja í aðrar bækur er oft gagnlegt. Ég held að reynslan sanni að hin bók er alltaf betri. Hún skýrir sama hlutinn út með pínulítið öðrum hætti og getur því fyllt upp í götin.

**Í prófinu eru öll skrifleg hjálpargögn leyfð.**