

Zuzana Havránková zuzana.havrankova@gmail.com ▼

27. 1. 2021, 15:53

Komu: 61505008@fsv.cuni.cz

✉ BP rewards-performance: postup práce



Dobrý den pane kolego,

takový brainstorming, pokud něco nedává smysl nebo jsem zapomněla, pište. Nicméně dokud ještě sbíráte data, vytvořte si separátní file, třeba docx, pro poznámky, rozdělte si ho podle sekcí heterogeneity (třeba cross-section, time-series, developing countries, motivation, rewards, methods, cokoliv relevantního) a kopírujte si tam doslovní pasáže textu k jednotlivým sekcím (neparafrázuje, jenom kopírujte s referencí na paper). Tedy když natrefíte na něco, k čemu by se Vám mohlo potenciálně hodit do BP, udělejte copy-paste (posílám v příloze příklad k paperu meta-analysis.cz/skill). Bude se Vám to hodit až budete potřebovat argumentaci k diskuzi o tom, proč jste se rozhodl pro tu kterou promennou nebo proč Vaše výsledky dávají smysl.

Teď k postupu:

Skvělé papers jsou

<http://meta-analysis.cz/sigma/> (technicky, má obří dataset a není tak "hezky" napsaný)meta-analysis.cz/skill (slovně perfektní, víc než technicky, protože má malý dataset)<http://meta-analysis.cz/armington/> (taky dobrý paper)

máte tam pdf, kódy

já používám neobratně jednou program Stata a jednou R, je to proto, že v R moc neumím ale R umí víc triků než Stata. Tedy dávám Vám tady ty moje mixované kódy. Vždycky si nejdříve udělejte summary statistics. Pokud budete dělat PCC, co asi budete, tohle jsou dobré zdroje: <http://meta-analysis.cz/education/> a <http://meta-analysis.cz/competition/>. Tj jak vypadá průměr efektu a SE, jaká je jejich standardní chyba, jaké je minimum, jaké je maximum.

Pokud efekt je "elasticity" a standardní chyba je "se":

* Summary statistics

summary effect se

*Někdy je evidentní, že v samplu jsou nějaké nesmysly a že bude potřebné tyhle nesmysly alespoň malinko osekát, vtedy winsorizují na 1%. Pro jistotu si ale vždycky zobrazím, jak mi vypadá funnel plot (dole), bez jakéhokoliv omezení. Pokud nevypadá jako obrácený trichter (http://meta-analysis.cz/sigma/sigma.pdf str 40 obr A1 vlevo je krásný funnel plot, jak bych ho chtěla, PCC funnel vypadá nějak takhle <http://meta-analysis.cz/education/> s tím, že by měl být symetrický, ale jak vidíte, na tom obrázku mz chybí část vpravo, dávám paper do přílohy), tak se prvně koukám na podezřelé pozorování, které jsou naoko "mimo" toho trychtýře. Většinou se jedná třeba o "cloud" pozorování z jednoho paperu, kde je evidentní, že s tím něco nebude v pořádku a tedy tyhle chyby je potřeba si pohlídat hned na začátku, může to být jenom blbě zapsaná desetinná čárka nebo vysloveně studie, která je nějak nekonzistentní se zbytkem populace.

*Většinou, zejména ve větších samplech se nevyhnete tomu, že tam nějaké odlehlé pozorování budou a ta je potřeba winsorizace. Standardně se provádí na levelu 1%, 2.5%, max 5%. Jak se rozhodnout, jak velká winsorizace? Prvně winsorizují a koukám se, jak se mi změní summary statistics, pokud zvětšují winsorizaci a summary statistics se mi významně nemění, přestanu winsorizovat. Samozřejmě chci primárně dělat co nejmenší zásahy do datasetu a zvolit co nejmenší winsorizaci, ale zároveň chci mít stabilní výsledky, které nejsou "driven by outliers". Někdy je ve větší sample i "větší" bordel a až testy na publication bias (dole) ukážou, jestli jsou testy "stabilní/robustní" nebo ne. Stačí baterie standardních lineárních FAT-PET testů.

```
winsor2 elasticity, suffix(_w) cuts(5 95)
winsor2 se, suffix(_w) cuts(5 95)
```

* Pokud je o winsorizaci rozhodnuto a lineární testy (FAT-PET) vypadají konzistentně, přistoupím k nelineárním testům (taky dole).

PUBLICATION BIAS TESTS

Assuming the exogeneity assumption between estimates and their SEs

```
*****
```

* PUBLICATION BIAS - Funnel plot (Egger et al., 1997)

```
*****
```

```
gen precision = 1/se
```

```
twoway scatter precision elasticity, saving(funnel, replace)
```

```
*****
```

* PUBLICATION BIAS - FAT-PET (Stanley, 2005)

```
*****
```

```
xtset idstudy
```

*OLS

```
ivreg2 elasticity se, cluster(idstudy)
```

*Fixed-effects

```
xtreg elasticity se, fe vce(cluster idstudy)
```

*Between-effects

```
xtreg elasticity se, fe vce(cluster idstudy)
```

*Weighted by number of observations per study

```
ivreg2 elasticity se [pweight=weight*weight], cluster(idstudy)
```

*Weighted by precision (=1/se)

```
ivreg2 elasticity se [pweight=precision*precision], cluster(idstudy)
```

*you could also try test PB on subsamples, like

*>>top 5 journals in Economict or part of the data that is very well published

*>>on subsample of methods where endogeneity is accounted for (such as instrumental variables, 2SLS, 3SLS, GMM)

*>>on some interesting subsample :)

```
*****
```

* PUBLICATION BIAS - WAAP (Ioannidis et al., 2017)

```
*****
```

```
summarize elasticity [aweight=1/(se*se)]
```

```
gen waapbound = abs(r(mean))/2.8
```

```
reg tstat precision if se < waapbound, noconstant
```

```
*****
```

* PUBLICATION BIAS - TOP10 method (Stanley et al., 2010)

```
*****
```

```
summarize precision, detail
```

```
gen top10bound = r(p90)
```

```
summarize elasticity precision if precision > top10bound
```

```
*****
```

* PUBLICATION BIAS - Stem-based method in R (Furukawa, 2019), use 2 columns "elasticity" and "se" as an input

```
*****
```

```
/*
```

```

source("stem_method.R") #github.com/Chishio318/stem-based\_method
dataforward = read.table("clipboard-512", sep="\t", header=TRUE)
stem_results = stem(dataforward$elasticity, dataforward$se, param)
stem_results[["estimates"]]
*/
*****
* PUBLICATION BIAS - FAT-PET hierarchical in R (use "author", "elasticity", "se" as an input, author
can be a study name, such as "Author et al. (2020)")
*****
/*
library(bayesm)
datalabor = read.table("clipboard-512", sep="\t", header=TRUE)
str(datalabor)
study <- levels(datalabor$author)
nreg <- length(study); nreg
regdata <- NULL
for (i in 1:nreg) {
  filter <- datalabor$author==study[i]
  y <- datalabor$elasticity[filter]
  X <- cbind(1,
    datalabor$se[filter])
  regdata[[i]] <- list(y=y, X=X)
}
Data <- list(regdata=regdata)
Mcmc <- list(R=6000)
out <- bayesm::rhierLinearModel(
  Data=Data,
  Mcmc=Mcmc)
cat("Summary of Delta Draws", fill=TRUE)
summary(out$Deltadraw)
*/
*****
* PUBLICATION BIAS - Selection model (Andrews & Kasy, 2019)
*   online tool at https://maxkasy.github.io/home/metastudy/
*****
/*
use the same (winsorized) elasticity estimate and standard error as for the stem-based method
*/
*****
* PUBLICATION BIAS - Endogenous kink (Bom & Rachinger, 2020)
*   Stata code at https://sites.google.com/site/heikorachinger/codes
*****
/*
download the code and adjust, Stata only. You can check also http://meta-analysis.cz/discrete/discrete.do, code starts with "quietly{" and ends with "rename bs discrete
rename sebs standard_error".
*/
Relaxing the exogeneity assumption btw ESTIMATE and SE (ESTIMATE)
*****
* FAT-PET with IV
*****
* IV estimation, instrument is the number of observations used for estimation of the effect in the

```

primary study, try to find a strong instrument, could be some function of nobs like $\ln(\text{nobs})$, $1/\sqrt{\text{nobs}}$, $1/\text{nobs}$, $1/(\text{nobs}*\text{nobs})$..)

generate instrument = $1/\sqrt{\text{nobs}}$

ivreg2 elasticity (se=instrument), cluster(idstudy)

* PUBLICATION BIAS - Caliper test (Gerber & Malhotra, 2008)

check <http://meta-analysis.cz/skill/skill.do> (there should be smaller intervals around $t=1.96$ but we did not have enough observations in those intervals)

PUBLICATION BIAS using p-uniform (van Aert & van Assen, 2020) - code for Stata & R

Je to asi nejvíce valuable robustness check, protože nepotřebuje mít žádné předpoklady o vztahu mezi efektem a jeho standardní chybou, bere v úvahu jenom p-values a to jenom ty signifikantní efekty a předpokládá nějakou distribuci těch p-values pokud efekt existuje a pokud neexistuje a signifikance je tvořena p-hackingem. Mrkněte na krátké video tady

https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pcurve.html, které tohle vysvětluje. p-uniform* vychází z p-curve Uriho Simonsohna.

generate data for p-uniform (here we generated median values for each study)

bysort idstudy: egen elasticity_med = median(elasticity)

bysort idstudy: egen se_med = median(se)

bysort idstudy: egen tstat_med = median(tstat)

preserve

sort idcoeff

collapse (lastnm) elasticity_med se_med tstat_med (p50) nobs_med=nobs, by(idstudy)

save "pcurve.dta", replace

restore

/* code for R

library(installr)

library(puniform)

library(metafor)

library(dmetar)

library(haven)

library(tidyverse)

library(meta)

data <- read_dta("pcurve.dta")

data\$variance_med <- data\$se_med*data\$se_med

preferred model (check if you have the correct side, side can be left or right)

puni_star(yi = data\$sigma_med, vi = data\$variance_med, side="right", method="ML", alpha = 0.05, control=list(max.iter=1000,tol=0.1, reps=10000, int=c(0,2), verbose=TRUE))

alternative but does not give you publication bias estimate

puni_star(yi = data\$sigma_med, vi = data\$variance_med, side="right", method="P", alpha = 0.05, control=list(max.iter=1000,tol=0.1, reps=10000, int=c(0,2), verbose=TRUE))

funky alternative

puni_star(tobs = data\$tstat_med, ni = data\$nobs_med, alpha = 0.05, side = "right", method = "ML", control=list(stval.tau=0.5, max.iter=10000,tol=0.1, reps=10000, int=c(0,2), verbose=TRUE))

*/

HETEROGENEITY

*Heterogeneita je o vysvětlujících proměnných. Pokud mi vyjde v publication bias testech publication bias přítomen, musím ho zahrnout i do části o heterogeneitě. Publication bias totiž může být nakonec jenom náhoda, output nějaké konkrétní "metodologie" nebo něco podobného, nemusí to být vysloveně o "selekci" výsledků. Udělám si testy na corelaci a collinearitu. Je potřeba počítat s tím, že tady bude značná korelace, ale důležitější bude VIF (variance inflation factor), který by měl být menší než 10. Tedy pokud to bude s VIF špatný, zkuste buď spojit nějaké variables (jako třeba metodologické, pokud to dává smysl), nebo zkuste použít jinou transformaci (třeba u publication year, mid-yr of data), přinejhorším variable vyhodíte. Většinou, když se odšéfuje VIF, odšéfuje se i korelace. Pokud je VIF OK a zůstane u něčeho vysoká korelace, 0.8 je ještě OK a neřešila bych to, ale nemělo by se to stát. Většinou se o tyto "drobné" problémy postará BMA s použitím dilution prior, dilution prior je právě na to.

correlate varlist

collin varlist

*HETEROGENEITY - **Bayesian model averaging code for R** (já data do R kopíruji ale Vy určitě budete umět i načíst *.dta file:). První model "labor" jsem už přestala používat, protože "labor1" s dilution prior je superior, ostatné modely jsou "robustness checks", stejně take FMA je robustness check

/*

library(BMS)

*ctrl+C from labor.xlsx, sheet("data.r")

datalabor = read.table("clipboard-512", sep="\t", header=TRUE)

labor = bms(datalabor, burn=1e5, iter=3e5, g="UIP", mprior="uniform", nmodel=50000, mcmc="bd", [user.int](#)=FALSE)

labor1 = bms(datalabor, burn=1e5, iter=3e5, g="UIP", mprior="dilut", nmodel=50000, mcmc="bd", [user.int](#)=FALSE)

labor2 = bms(datalabor, burn=1e5, iter=3e5, g="BRIC", mprior="random", nmodel=50000, mcmc="bd", [user.int](#)=FALSE)

labor3 = bms(datalabor, burn=1e5, iter=3e5, g="hyper=BRIC", mprior="random", nmodel=50000, mcmc="bd", [user.int](#)=FALSE)

#pro kazdy model, tady jenom "labor"

coef(labor, [order.by.pip](#) = F, exact=T, include.constant=T)

image(labor, yprop2pip=FALSE, [order.by.pip](#)=TRUE, do.par=TRUE, do.grid=TRUE, do.axis=TRUE, cex.axis = 0.7)

summary(labor)

plot(labor)

print(labor\$stopmod[1])

plotComp("UIP and Uniform"=labor1,"BRIC and Random"=labor2,"HQ and Random"=labor3, add.grid=F,cex.xaxis=0.7)

library(corrplot)

datalabor = read.table("clipboard-512", sep="\t", header=TRUE)

col<- colorRampPalette(c("red", "white", "blue"))

M <- cor(datalabor)

corrplot.mixed(M, lower = "number", upper = "circle", lower.col=col(200), upper.col=col(200), tl.pos = c("lt"), diag = c("u"), tl.col="black", tl.srt=45, tl.cex=0.85, number.cex = 0.5,

```
cl.cex=0.8, cl.ratio=0.1)
```

```
*/
```

```
*****
```

***HETEROGENEITY - Frequentist model averaging code for R** (Hansen---tohle je takový primitivní FMA kód, zatím nikdo nic lepšího nenapsal, proto pokud ho budete používat, seřadte si sloupce proměnných dle Posterior Inclusion Probability z "labor1" modelu)

```
*****
```

```
/*
library(foreign)
library(xtable)
library(LowRankQP)
datalabor=read.table("clipboard-512", sep="\t", header=TRUE)
datalabor <- na.omit(datalabor)
x.data <- datalabor[,-1]
const_<-c(1)
x.data <- cbind(const_,x.data)

x <- sapply(1:ncol(x.data),function(i){x.data[,i]/max(x.data[,i])})
scale.vector <- as.matrix(sapply(1:ncol(x.data),function(i){max(x.data[,i])}))
Y <- as.matrix(datalabor[,1])
output.colnames <- colnames(x.data)
full.fit <- lm(Y~x-1)
beta.full <- as.matrix(coef(full.fit))
M <- k <- ncol(x)
n <- nrow(x)
beta <- matrix(0,k,M)
e <- matrix(0,n,M)
K_vector <- matrix(c(1:M))
var.matrix <- matrix(0,k,M)
bias.sq <- matrix(0,k,M)

for(i in 1:M)
{
  X <- as.matrix(x[,1:i])
  ortho <- eigen(t(X)%*%X)
  Q <- ortho$Vectors ; lambda <- ortho$values
  x.tilda <- X%*%Q%*%(diag(lambda^-0.5,i,i))
  beta.star <- t(x.tilda)%*%Y
  beta.hat <- Q%*%diag(lambda^-0.5,i,i)%*%beta.star
  beta[1:i,i] <- beta.hat
  e[,i] <- Y-x.tilda%*%as.matrix(beta.star)
  bias.sq[,i] <- (beta[,i]-beta.full)^2
  var.matrix.star <- diag(as.numeric(((t(e[,i])%*%e[,i])/(n-i))),i,i)
  var.matrix.hat <- var.matrix.star%*%(Q%*%diag(lambda^-1,i,i)%*%t(Q))
  var.matrix[1:i,i] <- diag(var.matrix.hat)
  var.matrix[,i] <- var.matrix[,i]+ bias.sq[,i]
}

e_k <- e[,M]
```

```

sigma_hat <- as.numeric((t(e_k)%*%e_k)/(n-M))
G <- t(e)%*%e
a <- ((sigma_hat)^2)*K_vector
A <- matrix(1,1,M)
b <- matrix(1,1,1)
u <- matrix(1,M,1)
optim <- LowRankQP(Vmat=G,dvec=a,Amat=A,bvec=b,uvec=u,method="LU",verbose=FALSE)
weights <- as.matrix(optim$alpha)
beta.scaled <- beta%*%weights
final.beta <- beta.scaled/...tor
std.scaled <- sqrt(var.matrix)%*%weights
final.std <- std.scaled/...tor
results.reduced <- as.matrix(cbind(final.beta,final.std))
rownames(results.reduced) <- output.colnames; colnames(results.reduced) <- c("Coefficient",
"Sd. Err")
MMA.fls <- round(results.reduced,4)
MMA.fls <- data.frame(MMA.fls)
t <- as.data.frame\(MMA.fls\$Coefficient/...Err\)
MMA.fls$pv <- round( (1-apply(as.data.frame\(apply\(t,1,abs\)\)), 1, pnorm))*2,3)
MMA.fls$names <- rownames(MMA.fls)
names <- c(colnames(datalabor))
names <- c(names,"const_")
MMA.fls <- MMA.fls[match(names, MMA.fls$names),]
MMA.fls$names <- NULL
MMA.fls
*/

```

Tedy primárně si udělejte pořádek v datech, odhadněte publication bias. Jak budete mít hotové testy pro publication bias, budeme vědět, jakým směrem se práce bude pobírat (jaká story tam je). V těchto datech si myslím že není taková velká motivace pro to, aby tam nějaký bias byl, ale uvidíme. Tipla bych si, že zajímavější část bude heterogeneity

Upozornění :

Není-li v této zprávě výslovně uvedeno jinak, má tato e-mailová zpráva nebo její přílohy pouze informativní charakter. Tato zpráva ani její přílohy v žádném ohledu Univerzitu Karlovu k ničemu nezavazují. Text této zprávy nebo jejích příloh není návrhem na uzavření smlouvy, ani přijetím případného návrhu na uzavření smlouvy, ani jiným právním jednáním směřujícím k uzavření jakékoliv smlouvy a nezakládá předmluvní odpovědnost Univerzity Karlovy. Obsahuje-li tento e-mail nebo některá z jeho příloh osobní údaje, dbejte při jeho dalším zpracování (zejména při archivaci) souladu s pravidly evropského nařízení GDPR.

Disclaimer:

If not expressly stated otherwise, this e-mail message (including any attached files) is intended purely for informational purposes and does not represent a binding agreement on the part of Charles University. The text of this message and its attachments cannot be considered as a proposal to conclude a contract, nor the acceptance of a proposal to conclude a contract, nor any other legal act leading to concluding any contract; nor does it create any pre-contractual liability on the part of Charles University. If this e-mail or any of its attachments contains personal data, please be aware of data processing (particularly document management and archival policy) in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council on GDPR.

Přílohy



education.pdf – PDF, 973 kB

[↓ Stáhnout](#) [Zobrazit >](#)



notes.docx – Word, 614 kB

 **Stáhnout** [Zobrazit >](#)

Stáhnout všechny přílohy