

Vad för slags experiment bör ekonomer syssla med?

För ett par år sedan publicerade Armin Falk och James Heckman en uppmärksammat artikel med titeln "Lab Experiments Are a Major Source of Knowledge in the Social Sciences" i tidskriften *Science*. Författarna – båda uppburna nationalekonomer – argumenterade där för att både fältexperiment och laboratorieexperiment i grunden brottas med samma problem vad avser generaliserbarhet och extern validitet – och att det därför är omöjligt att säga att det ena skulle vara bättre än det andra.

Det som slår en när man läser både Falk/Heckman och förepråkare för fältexperiment – som till exempel John List och Steven Levitt – är egentligen att både fältstudier och experiment påminner mycket om teorimodeller. De har alla samma grundproblem – de förutsätter artificiella villkor och har alla problem med en "trade-off" mellan intern och extern validitet. Ju mer artificiella villkor, ju större intern validitet, men också mindre extern validitet. Ju mer vi riggar experiment/fältstudier/modeller för att undvika "confounders", ju mindre påminner villkoren om det verkliga "målsystemet". Sätillvida tror jag också att Falk/Heckman har rätt i sina synpunkter på att föra diskussionen om fältet-eller-experimentet i termer av realism – det handlar inte om det, utan i grunden om hur ekonomer via olika isoleringsstrategier i olika "nomologiska maskiner" försöker få kunskap om kausala relationer. Till skillnad från både Falk/Heckman och fältexperimentförspråkare som List och Levitt är jag tveksam till generaliserbarheten/extrapolerbarhet hos dem alla, eftersom risken är stor för att kausala mekanismer ser olika ut i olika kontexter och att bristande homogenitet/stabilitet/invarians gör att vi i vilket fall som helst inte har några fullgoda exportlicenser till "verkliga" samhällen/ekonomier.

Om man i huvudsak betraktar experiment/fältstudier som heuristiska verktyg så är nog skiljelinjen mellan exempelvis Falk/Heckman och List/Levitt väldigt svår att se.

Men om vi ser experiment/fältstudier som teoritest eller modeller som till syvende och sist pretenderar att säga något om verkliga "målsystem" är problemet med extern validitet centralt (och var under lång tid också ett avgörande skäl till att beteendekonomen hade problem att få sina forskningsresultat publicerade).

Antag att du har undersökt hur kinesiska lantarbetares arbetsprestationer A påverkas av B ("treatment"). Hur kan vi extrapolera/generalisera till nya sampel utanför ursprungspopulationen (exempelvis Sverige)? Hur vet vi att eventuella replikationsförsök "lyckats"? Hur vet vi när dessa replikationsförsöks resultat kan sägas berättiga de inferenser som görs i ursprungspopulationens sampel? Om, t ex, $P(A|B)$ är den konditionala densitetsfunktionen för ursprungssamplet, och vi är intresserade av att göra en extrapolerande prediktion av $E[P(A|B)]$, hur kan vi veta att det nya samplets densitetsfunktion är identisk med den ursprungliga? Om vi inte kan ge några riktigt bra argument för att så är fallet, säger ju inferenser byggda på $P(A|B)$ egentligen inget om vad som gäller för målsystemets $P'(A|B)$.

Som jag ser det är det här pudelns kärna. Externvaliditet/extrapolerbarhet/generaliserbarhet bygger på att vi skulle kunna göra inferenser baserade på $P(A|B)$ som är exporterbara till andra populationer för vilka $P'(A|B)$ gäller. Visst, kan man övertygande visa att P och P' är tillräckligt lika, är problemen kanske överkomliga. Men att godtyckligt bara introducera funktionella specifikationsrestriktioner av typen invarians/stabilitet/homogenitet är åtminstone för en vetenskapsteoretisk realist långt ifrån tillfredsställande. Och ofta är det tyvärr detta jag ser när jag tar del av (neoklassiska) ekonomers olika modeller/experiment/fältstudier.

Jag menar med detta inte att empiriska metoder *per se* skulle vara så problematiska att de inte kan användas. Tvärtom ställer jag mig i grunden – dock inte förbehållslöst – positiv till det ökade användandet av experiment/fältstudier. Inte minst som alternativ till ett i mina ögon fullständigt ofruktbart snickrande på "brolösa" axiomatisk-deduktiva teorimodeller. Min kritik handlar mer om aspirationsnivåer och vad vi med våra medierande epistemologiska verktyg/metoder inom samhällsvetenskapen tror oss kunna (be)visa.

Många "experimentalister" hävdar att det är enkelt att replikera experiment under olika förutsättningar och att man därför enkelt också kan testa undersökningsresultatens robusthet. Men är det verkligen så enkelt? Om vi i det anförda exemplet ovan kör ett test och finner att våra prediktioner inte stämmer – vad ska vi då dra för slutsats? Att B funkar i Kina men inte i Sverige? Eller att B funkar i ett underutvecklat agrarsamhälle men inte i ett postmodernt tjänstesamhälle? Att de funkade när fältstudien gjordes år 2008 men inte år 2010? Populationsval är i praktiken nästan aldrig enkla. Hade problemet om extern validitet handlat om att sluta från sampel till

population vore detta inget avgörande problem. De verkligt intressanta slutningarna görs väl ändå regelmässigt från specifika lab/experiment/fält till specifika verkliga situationer/institutioner/strukturer som vi är intresserade av att förstå eller (kausalt) förklara. Och då blir populationsproblemet svårare att tackla.

Alla – både labbar och experimentalister -bör betänka följande rader från David Salsburgs **The Lady Tasting Tea** (Henry Holt 2001:146):

In Kolmogorov's axiomatization of probability theory, we assume there is an abstract space of elementary things called 'events' ... If a measure on the abstract space of events fulfills certain axioms, then it is a probability. To use probability in real life, we have to identify this space of events and do so with sufficient specificity to allow us to actually calculate probability measurements on that space ... Unless we can identify Kolmogorov's abstract space, the probability statements that emerge from statistical analyses will have many different and sometimes contrary meanings.

Eller varför inte fundera över vetenskapsteoretikern Nancy Cartwrights synpunkter på värdet av randomiserade experiment i *The Lancet* 23/4 2011:

But recall the logic of randomized control trials ... [T]hey are ideal for supporting 'it-works-somewhere' claims. But they are in no way ideal for other purposes; in particular they provide no better bases for extrapolating or generalising than knowledge that the treatment caused the outcome in any other individuals in any other circumstances ... And where no capacity claims obtain, there is seldom warrant for assuming that a treatment that works somewhere will work anywhere else. (The exception is where there is warrant to believe that the study population is a representative sample of the target population – and cases like this are hard to come by.)