

Välkommen kritik men slutsatserna håller

ELIAS COLLIN,
CHRISTIAN SANDSTRÖM
OCH KARL WENNBERG

REPLIKSVAR

I nummer 7/2021 av *Ekonomisk Debatt* presenterar Fridholm en replik på vår tidigare artikel "Utvärderingar av näringspolitik – en intressekonflikt mellan myndigheter, konsultföretag, politik och skattebetalare?" (*Ekonomisk Debatt* nr 4, 2021). I slutet av artikeln uppmantrar vi till en bredare diskussion om utvärderingar av innovations- och näringspolitik, varför Fridholms replik är välkommen. För kännedom rör en del av Fridholms invändningar även den debattartikel som publicerades i *DN* med anledning av våra resultat.

Elias Collin är statsvetare från Uppsala universitet och knuten till Ratio. eliascollin@gmail.com

Christian Sandström är biträdande professor i Digital Business vid Internationella Handelshögskolan i Jönköping. Hans forskning handlar om samspelet mellan teknisk utveckling, regleringar och företags konkurrenskraft. christian.sandstrom@ju.se

Karl Wennberg innehar Barbara Bergströms professur i Educational Leadership and Excellence vid Handelshögskolan i Stockholm samt är knuten till Linköpings universitet och Ratio. Wennbergs forskning kretsar kring bl a entreprenörskap, utbildningsekonomi och innovationspolitik. karl.wennberg@hhs.se

Fridholms inledande kritik handlar om att vår analys förväxlar lärande/formativa utvärderingar och effektutvärderingar, som vanligen är summativa. Den här distinktionen är viktig och vi är överens med Fridholm om att båda formerna av utvärderingar behövs. Vi är medvetna om att utvärderingar i vårt material innehåller såväl kvantitativa kontrafaktiska analyser som insatser där utvärderaren agerar "följeforskare" och i vissa fall fungerar som bollplank under genomförandet av ett projekt. Just därför är det ett så stort problem att lejonparten av de utvärderingar som görs faktiskt uttalar sig om "effekter" de i själva verket inte har undersökt. Utvärderingar som varken är utformade för eller syftar till att utreda effekter uttalar sig ändå om detta och myndigheter refererar till dessa utvärderingar som om så var fallet.

Fridholms andra kritik är att vår analys av relationen mellan utvärderande aktör och utvärderingsresultat (positivt eller negativt) inte beaktar vilken typ av utvärdering som genomförts. Om

det finns en systematisk skillnad mellan utvärderande aktörer (utförande myndigheter, expertmyndigheter, konsulter, lärosättesforskare) och typ av utvärdering kan det då snedvrیدا resultaten från våra skattningar. Fridholm visar dock inte någon egen sambandsanalys utan rapporterar enbart en figur där han kodat 209 utvärderingar av myndigheters innovationsinsatser 2009–19. För att formellt undersöka validiteten i argumentet att utvärderingstyp måste kontrolleras för har vi replikerat Fridholms kodning av "utvärderingstyp" för samtliga våra analyserade utvärderingar enligt det sätt han beskriver i sin replik. Detta med hjälp av två av varandra oberoende kodare. Vi använder denna nya variabel som en kontroll i logit-analysen som predicerar huruvida utvärderingar är positiva eller negativa (se tabell 1 i Collin m fl 2021, s 37). Alla tidigare resultat kvarstår i det närmaste identiskt även efter att utvärderingstyp enligt Fridholms kategorisering kontrollerats för, oavsett om vi introducerar fem nya dummyvariabler eller en kategorivariabel med sex olika värden. Detta visar att resultaten *inte* förändras även om vi formellt tar hänsyn till utvärderingarnas syfte och frågeställning på det sätt som Fridholm kan antas mena i sin invändning. Att resultaten kvarstår illustrerar huvudproblemet: utvärderingar uttalar sig om effekter och företrädesvis positiva sådana, utan att det finns belägg för det.

Fridholms tredje kritik är att vår analys inte heller beaktar vilken typ av måttstock som används i de utvärderingar vi studerar. Måttstock förstås i det här fallet som: "mot vilken/vilka kriterier görs den normativa bedömningen" (Fridholm 2021, s 39). Det är förstås sant att den som väljer mycket enkla kriterier för när en intervention kan anses vara effektiv kommer klassa fler interventioner som positiva än om tuffare kriterier ställs upp. Att operationalisera utvärderingarnas respektive

måttstock är däremot problematiskt. Fridholm föreslår tre kategorier: måluppfyllelse, expertutlåtande och empiriska jämförelser (se Fridholm 2021 för mer exakta definitioner). Vi noterar att dessa tre kategorier inte är särskilt utmärkande inom gängse litteratur (se t ex Vedung 2009) men för den sakens skull inte nödvändigtvis är dåliga variabelvärden. Vad som däremot är problematiskt är att de tre kategorierna inte är ömsesidigt uteslutande – det är t o m troligt att utvärderingar både använder måluppfyllelse och empiriska jämförelser som måttstock vid en och samma utvärdering. Hur man i sådana fall kodar reliabelt är inte ett problem vi lyckats lösa. Vi har förståelse för att det är svårt att till fullo redogöra för sin metod i en kort replik men baserat på de förklaringar Fridholm ger av denna variabel och dess kodning var det inte möjligt att replikera kodningen på ett reliabelt sätt.

Vad vi däremot har gjort är att koda hur pass konkreta de mål en utvärdering har utgått ifrån har varit. Detta har vi gjort genom att kategorisera dem efter hur kvantifierbara de varit och således skilja mål som är 1) icke-kvantifierade, 2) kvantifierbara men inte mätbara och 3) mätbara i absoluta eller relativa termer. I vår initiala kodning (och även i den kompletterande kodning vi genomfört) är kategori 1 och 2 kraftigt överrepresenterade. När vi inorporerat denna kontrollvariabel i vår logit-analys får vi inga signifikanta resultat, varför den torde vara mindre viktig för vår analys.

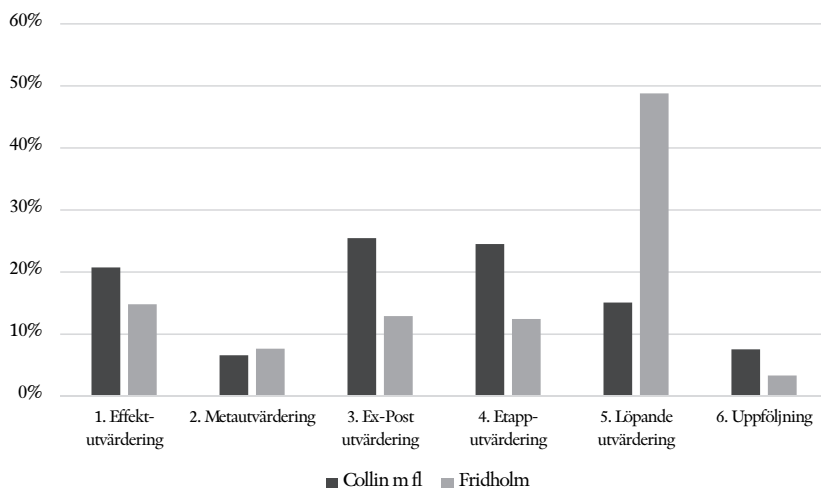
Vi är skeptiska till Fridholms kritik om att man även bör kategorisera olika utvärderingars ”måttstockar”. De exempel som Fridholm ger såsom måluppfyllelse, expertutlåtande och empirisk jämförelse är långt ifrån de enda mål som nämns i utvärderingslitteraturen. Att välja ut ett visst antal kategorier av måttstockar och sedan koda dessa uniformt över ett stort antal utvärderingar

blir lätt väldigt subjektivt. Vad vi gjort är att koda den måttstock som säger något om hur pass kvantifierade målen för det utvärderade projektet var, vilket är inkluderat som en kontrollvariabel i vår analys.

Vid kommunikation med Fridholm och jämförelse mellan de två dataseten så framkom även att Fridholms analys utgår från ett dataset som till 69 procent består av annorlunda utvärderingar än vårt dataset, 50 procent av hans data består av projektutvärderingar från Tillväxtverket (där många är ”löpande” utvärderingar inom ramprogrammet för EU stöd). Därav det annorlunda utseendet på staplarna. Figur 1 nedan jämför utvärderingarna i Fridholms dataset (n=209) med vårt eget (n=109). Fridholm påstår – utan tydliga argument – att vår analys baseras ”på ett mycket skevt underlag” (s 43). Samma kritik kan riktas mot hans eget urval som i huvudsak bygger på utvärderingar av Tillväxtverkets projekt. Detta illustrerar vikten av att myndigheters utvärderingar är fritt tillgängliga för analytiker, forskare och medborgare enligt EU:s *open data*-direktiv såväl som svensk offentlighetsprincip, samt att forskare öppet redovisar sitt material och sina resultat. Vi är tacksamma för den konstruktiva diskussionen med Fridholm under detta arbete.

Vi delar Fridholms åsikt att fler och bättre effektutvärderingar bör genomföras, samt att utvärderare behöver uttrycka sig tydligare rörande vad de syftar på när de redovisar ”effekter”. Som Riksrevisionen (2020) tydligt visar så förekommer i dag övertolkning av resultat från myndigheter och beslutsfattande politikernas sida när resultaten från olika utvärderingar kommuniceras. Vi har inte på något sätt hävdad att alla andra utvärderingar är ”illegitima” och vi ämnar inte bemöta Fridholms påstående om ”oetisk argumentation”. Att anklaga någon för oetiskt förfarande

Figur 1
Typer av utvärderingar i vårt och Fridholms dataset



Källa: Egna beräkningar.

är allvarligt. Vad vi har poängterat är problemet som uppstår när effektutvärderingar (*impact evaluations* i EU-terminologi) så sällan genomförs och beslutsfattare omtolkar andra former av utvärderingar som ”bevis på effekt”. Om EU lägger 100-tals eller 1000-tals mdr euro och Sverige 30-tals mdr kronor på olika former av näringslivsstöd med oklara effekter kan detta ha snedvridande konkurrenseffekter och därmed hota innovationsförmågan (Aghion m fl 2020).

Vi vill understryka att vår studie inte syftar till att säga någonting om utvärderingarnas faktiska kvalitet. En formativ, lärande, utvärdering kan vara mycket väl genomförd och skapa gott lärande för den myndighet som utvärderas. Detta utan att säga något om de effekter den utvärderade interventionen har haft. Sådana formativa utvärderingar skapar goda förutsättningar för förbättrad verksamhet, men metoderna som används kan sällan nyttjas till att säga någonting om en interventions reella effekter. Problemet infinner sig snarare när utvärderingar som syftar till lärande uttalar sig om effekter utan att det finns belägg för detta och/eller när myndig-

heter använder dessa utvärderingar för att övertyga politiker om värdet i deras arbete.

I vår ursprungliga text har vi har aldrig påstått att konsultföretag skulle vara inkompetenta. Vad vi däremot påvisar är att incitamentsstrukturen inom vilken utvärderingar i dag genomförs påverkar utvärderingarnas resultat. Såväl forskare som konsulter är försatta i en beroendeställning vis-à-vis de myndigheter som utvärderas. Detta illustreras av våra resultat där den överväldigande majoriteten av de utvärderingar som görs landar i positiva slutsatser. Kompetenta utvärderare behöver få verka i en incitamentsstruktur inom vilken det är möjligt att göra kritiska, opartiska och metodologiskt stringenta utvärderingar. I det långa loppet kommer duktiga utvärderare, myndigheter och medborgarna tjäna på detta.

REFERENSER

- Aghion, P, C Antonin och S Bunel (2020), *The Power of Creative Destruction: Economic Upheaval and the Wealth of Nations*, Harvard University Press, Cambridge MA.
- Collin, E, C Sandström och K Wennberg (2021), ”Utvärderingar av näringspolitik

– en intressekonflikt mellan myndigheter, konsultföretag, politik och skattebetalare?”, *Ekonomisk Debatt*, årg 49, nr 4, s 30–41.

Fridholm, T (2021), ”Collin m fl har inte underlag för sina slutsatser”, *Ekonomisk Debatt*, årg 49, nr 7, s 39–44.

Riksrevisionen, (2020), ”Effektutvärderingar av näringspolitiken – bristande tillförlitlighet”, RiR 2020:30, Riksrevisionen, Stockholm.

Vedung, E (2009), *Utvärdering i politik och förvaltning*, Studentlitteratur, Lund.

Rättelse

I ”Utvärderingar av näringspolitik – en intressekonflikt mellan myndigheter, konsultföretag, politik och skattebetalare?” *Ekonomisk Debatt* 4/2021 (årgång 49) så redogjorde vi för att den aktuella perioden för urvalet av utvärderingar var 2009–19 när det egentligen ska vara 2005–20. För att så nära som möjligt jämföra våra data och kodning med Fridholm har vi även genomfört en ny

kodning av 140 utvärderingar 2009–19, från vår ursprungliga dataserie av utvärderingar 2005–20 där analysen tidigare baserades på ett 50 procentigt stickprov. I detta urval är skillnaden i utvärderingar av kategori 4 och 5 med Fridholms data något mindre än vad som illustreras i figur 1, men de multivariata resultaten kvarstår i princip oförändrat i effektstorlek och riktning.