

Citeringar – att mäta och att mäta rätt

Vad betyder valet av citeringsmått för jämförelser av forskare i nationalekonomi? Vår analys av antalet citeringar i Scopus, Social Sciences Citation Index och Google Scholar för alla forskare med anknytning till IFN visar en stark positiv korrelation mellan de tre måtten vad gäller rangordningen mellan forskarna. Det finns dock vissa indikationer på att äldre forskare missgynnas i Scopus och att mer policyinriktade forskare gynnas i Google Scholar. Rangordningen kan också förändras om den baseras på h-index i stället för antalet citeringar. Citeringar bidrar med viktig information om en forskares kompetens och meriter, men de kan inte helt ersätta kvalitativa utvärderingar.

Kvantitativa mått används sedan länge för att mäta hur mycket forskning som åstadkoms och för att rangordna forskare, institutioner och tidskrifter. Fördelningen av resurser, beslut om anslag och tillsättandet av tjänster grundar sig också allt oftare på kvantitativa mått (Harzing och Alakangas 2016). Sedan år 2010 baseras fördelningen av en liten del av de direkta statliga forskningsanslagen till Sveriges lärosäten på vetenskapligt genomsnittligt mått med en bibliometrisk indikator. Flera lärosäten utnyttjar även själva bibliometriska mått för att fördela resurser mellan fakulteter och institutioner (Görnerup 2013).

Det mest basala måttet på en forskares eller institutions forskningsproduktion är antalet publikationer. Ofta görs en kvalitetsjustering baserat på publiceringskälla. För en artikel innebär detta att den tillskrivs ett värde motsvarande den genomsnittliga artikeln i tidskriften.

På senare tid har intresset ökat för att mer direkt mäta kvaliteten hos en viss studie i stället för att bara anta att den är lika med genomsnittet för artiklarna i den tidskrift där studien publicerats. Det vanliga sättet att göra detta på är att mäta hur många gånger studien citerats. Spelar det då någon roll vilket citeringsmått som används? Syftet med denna artikel är att belysa den frågan genom att jämföra resultaten när antalet citeringar hämtas från de tre mest använda citeringsdatabaserna: *Social Sciences Citation Index*, som ingår i *Web of Science*, *Scopus*, som drivs av Elsevier, och *Google Scholar*. Jämförelser görs både för antalet citeringar och *h-index*, vilket lägger större vikt vid kontinuerlig vetenskaplig produktion än relativt få mycket citerade verk.¹

Populationen för vår analys består av samtliga ordinarie och affilierade

SIMON EK OCH MAGNUS HENREKSON

Simon Ek har en masterexamen i nationalekonomi och är sekreterare för Arbetsmarknadsekonomin rådet. Han har tidigare arbetat som forskningsassistent vid Institutet för Näringslivsforskning (IFN).
simon.ek@ifn.se

Magnus Henrekson är professor och vd för Institutet för Näringslivsforskning (IFN). Han forskar huvudsakligen om entreprenörskapets ekonomi.
magnus.henrekson@ifn.se

¹ En forskares *h-index* är lika med x om forskaren har publicerat minst x artiklar som blivit citerade minst x gånger vardera.

Vi tackar Lars Calmfors för värdefulla synpunkter.

forskare vid Institutet för Näringslivsforskning (IFN) i Stockholm. Dessa forskares citeringar och rangordning baserad på citeringarna jämförs sedan för att se i vad mån de olika måtten ger en samstämmig bild.²

1. Vilka mått och databaser finns och används?

Tillvägagångssätt och problem vid mätning

Att mäta forskning är en grannliga uppgift. Om man enbart ser till kvantitet ger det incitament att publicera så många arbeten som möjligt, oavsett vad och var, och att dela upp sina resultat i ”minsta publicerbara enheter”. Fler publikationer kan inte automatiskt tolkas som ett större forskningsbidrag, utan det behövs kvalitetsjusteringar. Vad gäller tidskriftspubliceringar är två frågor särskilt viktiga: (1) Hur bör tidskrifter rangordnas och givet att en rangordning gjorts, (2) hur stora ska skillnaderna anses vara?

Skillnaderna mellan tidskrifter varierar beroende på vilket mått som används, och vissa mått är extrema (Henrekson och Waldenström 2011). Mest extremt är det så kallade KMS-måttet (Kalaitzidakis m fl 2003), vilket (åtminstone tidigare) förordats av European Economic Association. Exempelvis motsvarar en artikel i *American Economic Review* 588 artiklar i *Journal of Evolutionary Economics*.³ Detta kan jämföras med att *Association of Business Schools* (ABS) ger fyra poäng till en artikel i *AER* och två poäng till en artikel i *Journal of Evolutionary Economics* (Kremer m fl 2015).

Det finns en stark tendens hos människor att ”göra det vi mäter” (Holmström och Milgrom 1991). Det mått som används signalerar vilken slags forskning som anses värdefull, och de stora skillnaderna mellan olika mått gör att incitamenten blir beroende av vilket mått som väljs (Sauder och Espeland 2009; Mingers och Willmott 2013).

Tidskrifters kvalitet bestäms oftast av hur ofta dess artiklar citeras. Det mest använda kvalitetsmåttet på tidskrifter är *Journal Impact Factor* (JIF) som ingår i *Journal Citation Reports* (JCR). Kvaliteten avgörs enligt detta mått av hur många gånger dess artiklar i genomsnitt citeras i andra artiklar publicerade i tidskrifter som ingår i *Web of Science*. Citeringar har en mycket ojämn fördelning (Mutz och Daniel 2012), vilket innebär att en tidskrifts rang (och därmed dess artiklar) kan lyftas kraftigt tack vare ett fåtal extremt välciterade artiklar. Det finns också begränsningar i hur gamla citeringar och artiklar som inkluderas. Vissa verk får genomslag först flera år efter att de publicerats. Om verket är nydanande är detta t o m troligt, då resten av forskningsfältet först måste hinna ikapp. Ett mycket kort tidsfönster kan därför göra måttet missvisande. Om ett verk blir citerat redan innan det publicerats i sin slutliga form, bör då dessa citeringar också räknas? Mycket

² För en mer omfattande analys där vi också jämför citeringar från de olika databaserna för de 25 högst rankade svenska nationalekonomerna enligt *RePEc* hänvisas till Ek och Henrekson (2018). Där finns också ytterligare detaljer inklusive länkar avseende de olika måtten.

³ KMS-måttet användes av Lindqvist (2003) när han rangordnade svenska professorer i nationalekonomi.

talar för att så borde vara fallet, särskilt med tanke på den ofta långa publiceringseftersläpningen.

Det vanligaste och mest vedertagna sättet att mäta vetenskaplig produktion är att använda kvalitetsviktade publiceringsmått. Även om det finns variation avgränsar de flesta kända måtten sig till verk publicerade i vetenskapliga tidskrifter med *peer review*. Man beräknar sedan en kvalitetsvikt för varje tidskrift som inkluderas. Dessa vikter kan justeras för bl a självciteringar, tidskriftens sid- eller artikelantal och ibland även för författarnas ålder. Varje artikel publicerad i en viss tidskrift tilldelas samma vikt och en forskares (institutionens) poäng räknas sedan ut genom att (1) multiplicera varje publikation med motsvarande tidskrifts vikt och (2) summerna de viktade poängen för personens (institutionens) publikationer.

En relativt ny idé är att också ta hänsyn till citeringars kvalitet när vikter skapas för tidskrifter. Exempel på mått som gör det är *Eigenfactor* (Bergstrom 2007) baserat på data från *SSCI* och *SCImago Journal Rank* (Gonzales-Pereira m fl 2010) som använder *Scopus*. Båda dessa mått bygger delvis på att (1) en tidskrifts kvalitet bestäms av hur välciterade dess artiklar är och att (2) citeringar av tidskriftens publikationer viktas utifrån detta kvalitetsmått.

Det finns också hybridrangordningar. Association of Business Schools (ABS) tidskriftsrangordning utnyttjar både data över citeringar och ett system med fackgranskning utförd av en vetenskaplig kommitté för varje forskningsområde.

Under senare år har våra nordiska grannländer tagit fram s k auktoritetsregister för forskningskanaler. Dessa definierar vilka forskningskanaler (t ex tidskrifter och förlag) som ska medräknas när beslut tas om bl a finansiering och hur kanalerna värderas. Mest känd är troligen den ”norska listan” (Register over vitenskaplige publiseringskanaler) som sköts av Norsk Senter for Forskningsdata. I Sverige sker ett liknande utvecklingsarbete kring ett auktoritetsregister inom ramen för projektet SwePub, där Kungliga biblioteket samverkar med Vetenskapsrådet och Sveriges universitets- och högskoleförbund.

De nordiska registren har en betydligt jämnare kvalitetsjustering jämfört med andra publiceringsmått och skapar således helt andra incitament.⁴ Flera svenska universitet har börjat använda den norska listan. De publiceringskanaler som platsar på listan klassificeras i endera av två nivåer: forskningskanaler (nivå 1) och forskningskanaler som är ledande inom sitt fält (nivå 2). Ett tydligt exempel på att fördelningen är mindre ojämn är att både *Quarterly Journal of Economics* och *Scandinavian Journal of Economics* klassificeras som nivå 2 (och får då samma poäng). Den förra ligger på första plats inom nationalekonomi i *JCR* med en *JIF* på 6,66 medan den senare ligger på plats 102 med en *JIF* på 1,44. *Ekonomisk Debatt* finns också med på listan (nivå 1). Att sikta mot toptidskrifter blir således inte längre lika självklart.

⁴ Med ojämn kvalitetsjustering menar vi att det finns en stor variation i tilldelade kvalitetsvikter mellan tidskrifter.

Dock ska poängteras att syftet med den norska listan är att mäta vetenskaplig forskningsaktivitet på lärosätes- eller institutionsnivå; den är inte tänkt som ett verktyg för att utvärdera enskilda forskare (Henning 2013).

Ett problem med tidskriftjusterade publiceringsmått är att det finns stor variation mellan artiklar inom tidskrifter vad gäller citeringsfrekvens; även i de högst rankade tidskrifterna publiceras artiklar med låg eller försumbar citeringsfrekvens (Laband och Tollison 2003; Oswald 2007). Detta är ett skäl till att användandet av kvalitetsmått för tidskrifter för att utvärdera enskilda artiklar eller forskare fått mycket kritik (se t ex Callaway 2016). Det finns också indikationer på att sambandet mellan en enskild artikels och motsvarande tidskrifts citeringsfrekvens försvagats sedan 1990-talet (Lozano m fl 2012).

Därför är citeringsmått som mäter verkligt genomslag viktiga (givet att ett arbete inte citeras som avskräckande exempel). Alternativet är att mäta antalet publikationer eller att använda tidskriftsjusterade publiceringsmått. Vid det senare överlämnas kvalitetsbedömningen till en tidskrifts granskare och redaktörer som beslutar om artikeln ska få ingå i tidskriften och därmed tillskrivas ett värde motsvarande dess genomsnittliga genomslag. När det gäller forskning om Sverige – särskilt mer policynära sådan – är det inte alltid optimalt att låta utländska tidskrifter, i stället för citeringar från andra forskare (eller utredare), bedöma forskningens värde.

En del av de fasta forskningsmedlen i Sverige fördelas baserat på respektive lärosätes forskningsproduktion, vilken mäts med en bibliometrisk indikator baserad på citeringsdata från en databas som sköts av Vetenskapsrådet och som bygger på samma grundmaterial som *Web of Science*. Den svenska indikatorn tar hänsyn till och normerar för det faktum att det genomsnittliga antalet publiceringar och citeringar i *Web of Science* varierar mellan discipliner. Normeringen görs dels för att kompensera ämnesområden med svag representation i *Web of Science*, dels för att kunna jämföra antalet citeringar mot genomsnittet i världen för samma forskningsfält ett givet år (Henning 2013).⁵

Precis som med publiceringsmått finns en rad problem med citeringsmått. Ett är de stora skillnaderna mellan discipliner i praxis rörande författarskap, publicering och citering (LSE 2011). Detta gäller även mellan olika nationalekonomiska fält (Card och DellaVigna 2013). Det kan också finnas olika citeringsmönster över tid; exempelvis tycks teoretiska arbeten till en början ha mindre, men över tid mer uthålligt, genomslag jämfört med empiriska studier (Kim m fl 2006).

Det finns flera metoder för att harmonisera citeringar för att därmed kunna jämföra mellan forskningsfält och discipliner.⁶ Den svenska resursfördelningsmodellen bygger på en sådan metod. Men en fråga är då hur

⁵ Utöver vetenskaplig produktion tar fördelningsmodellen hänsyn till lärosätenas förmåga att attrahera externa forskningsmedel. Se Kesselberg (2015) och Henning (2013).

⁶ Se Waltman och van Eck (2013) och Bornmann och Marx (2015) för jämförelser av olika normaliseringsmetoder. Den senare studien jämför olika metoder med kvalitativa bedömningar. Bornmann m fl (2016) gör ett försök att normalisera citeringar i *Google Scholar*.

man bör göra med innovativ forskning som spänner över flera fält. Harmonisering mellan forskningsfält kan också ge incitament att försöka klassificera ett verk på ett för forskaren så gynnsamt sätt som möjligt.

Ytterligare en aspekt är om justeringar för antalet författare av varje verk bör användas. Om ingen medförfattarjustering görs ges incitament att försöka vara delaktig i så många välciterade publikationer som möjligt, utan att det spelar någon roll hur stort det enskilda bidraget till verket är.

En annan fråga, precis som med tidskrifter, är om citeringarna bör kvalitetsjusteras eller om man ska begränsa vilka slags citeringar som ska räknas. Att citeras av andra forskare – snarare än t ex av utredare – skulle kunna vara en starkare indikator på forskningens vetenskapliga bidrag. Samtidigt visar citeringar som görs av utredare att forskningen har praktiskt värde. Olika citeringsdatakällor tillämpar olika avgränsningar och insamlingsmetoder. I både *Scopus* och *Web of Science* ingår t ex enbart vetenskapligt publicerade arbeten. Således exkluderas citeringar till och från utredningar eller liknande. Detta leder oss in på frågan om hur viktigt valet av citeringsdatabas är för utfallet i en jämförelse mellan forskare.

Databaser över citeringar och Hirschs h-index

Web of Science (WoS) startades 1964 och är den i särklass äldsta och mest kända citeringsdatabasen för fackgranskade verk. Databasen sträcker sig tillbaka till år 1900. Vi fokuserar på verk som ingår i *Social Sciences Citation Index (SSCI)*, som är en del av *WoS*. I *SSCI* ingår ca 3 200 tidskrifter och ca 8,5 miljoner verk. För att komma med i *WoS* måste ett verk ha publicerats av en av plattformens utvalda tidskrifter. Tidskrifter som redan ingår och som kan komma att ingå i *WoS* utvärderas årligen. Tidskrifterna ska uppfylla krav på punktlighet och följa internationella redaktionella normer för språk, innehåll och citeringar.

Tidigare var *WoS* den enda omfattande databasen för citeringar av akademiska artiklar. År 2004 introducerades både *Scopus* och *Google Scholar*. *Scopus* drivs av det nederländska förlaget Elsevier och omfattar närmare 23 000 fackgranskade tidskrifter från över 5 000 förlag.⁷ Utöver detta inkluderas verk från vissa branschtidskrifter, artiklar som blivit accepterade men inte publicerade ännu och konferensbidrag. Databasen innehåller även böcker som publicerats fr o m 2005. Tidigare ingick endast publikationer och citeringar fr o m 1996, men sedan 2015 har arbete lagts på att utvidga databasen bakåt i tiden till 1970 (Harzing och Alakangas 2016). Syftet är att ge en mer rättvisande bild för äldre forskare. *Scopus* inkluderar endast citeringar i publikationer som ingår i *Scopus*.

WoS och i allt högre grad *Scopus* åtnjuter hög respekt inom akademien.⁸

⁷ För att en tidskrift ska ha en chans att inkluderas i *Scopus* krävs bl a att dess artiklar är fackgranskade, att utgivningen är regelbunden och har pågått i minst två års tid samt att sammanfattningar och referenser är tillgängliga på engelska. Vidare måste riktlinjer för granskningsprocessen och publiceringsetik vara allmänt tillgängliga.

⁸ Dock har hävdats att det finns en tendens till ideologisk bias vad gäller vilka tidskrifter som accepteras i *SSCI* (Klein och Chiang 2004).

Google Scholar (GS) har däremot stundtals fått utstå hård kritik vad gäller träffsäkerhet. GS är en sökmotor för akademiska publikationer och hämtar citeringar från alla relevanta källor den kan hitta (Mingers och Yang 2017). Bland annat Jacsó (2010) menar att GS är olämplig för bibliometriska studier av forskares och tidskrifters genomslag. I vissa fall tillskriver sökmotorns algoritmer publikationer (och därmed citeringar) till författare som inte finns, vilket gör att den egentliga författaren får en lägre citeringsfrekvens. Likaså förekommer dubbelräkning av verk och citeringar.⁹

Samtidigt har GS blivit känd och välanvänd just för att den beaktar många olika slags källor. Harzings (2007) mjukvara *Publish or Perish* använder GS för sina beräkningar, vilket har bidragit till sökmotorns popularitet. Vidare är det lätt att registrera en egen profil och sökningar är gratis.

Tidigare forskning finner generellt en relativt hög överensstämmelse mellan *Scopus* och *WoS*, medan GS skiljer sig åt och har betydligt bredare täckning för de flesta forskningsfält. GS genererar också fler citeringar. Kvaliteten på den insamlade informationen från GS är dock lägre även om *WoS* och *Scopus* också lider av problem, inte minst att författare med vanliga namn ofta förväxlas (Harzing och Alakangas 2016; Waltman 2016; Mingers och Yang 2017).

Utöver att studera antalet citeringar för varje författare använder vi också Hirschs (2005) *h*-index baserat på citeringar och verk från respektive databas eller sökmotor. Måttet har fått stor uppmärksamhet och brett erkännande.¹⁰ Syftet är att korrigera för effekten av *one-hit wonders*, dvs enstaka mycket citerade verk, genom att lägga större vikt vid kontinuerlig vetenskaplig produktion.

En författares *h*-index är x om minst x av författarens samtliga publicerade verk (N) blivit citerade minst x gånger vardera och resterande verk ($N - x$) blivit citerade färre än x gånger vardera. En författare kommer då som minst att ha x^2 citeringar, men mer än så om $(N - x) > 0$ (givet att de ytterligare verken har citerats någon gång) eller om vissa verk fått fler än x citeringar. Måttet ger ingen extra vikt till verk som fått ett mycket stort antal citeringar. I stället premierar det forskare som publicerar många studier och där dessa blir förhållandevis väl citerade. Men *h*-indexet kan också vara alltför restriktivt. Exempelvis kommer två författare med fem publikationer och minst fem citeringar per publikation att båda få $h = 5$, oavsett det totala antalet citeringar för respektive forskare.

2. Vår undersökning

Uppgifter om antalet citeringar enligt *SSCI*, *Scopus* och *GS* har samlats in för de 68 forskare som var eller varit anställda vid eller affilierade till IFN mellan februari 2016 och maj 2017. Data samlades in i maj 2017.

⁹ I forskningssyfte lyckas Delgado López-Cózar m fl (2014) manipulera antalet citeringar i sina egna profiler i GS genom att publicera fabricerade verk av en fiktiv författare på internet.

¹⁰ Mingers och Yang (2017) argumenterar för att *h*-index är ett av de bästa måtten för att mäta forskningsbidrag.

Citeringarna är mycket ojämnt fördelade. Endast elva av de 68 forskarna har fler citeringar än det genomsnittliga antalet i samtliga databaser, och de fem till sju (beroende på mått) mest citerade forskarna har hälften av det totala antalet citeringar. *Scopus* har en något jämnare fördelning i toppen, medan *GS* är jämnast i den nedre delen av fördelningen.

Jämförelse mellan citeringar

I figur 1 presenteras parvisa spridningsdiagram för de tre citeringsmått. Alla forskare utom en har lika många eller fler citeringar i *GS* än i *Scopus* och *SSCI*.¹¹ Det skattade sambandet anger att en ytterligare citering i *SSCI* är förknippad med 5,9 citeringar i *GS*, medan en citering i *Scopus* motsvarar 4,6 citeringar i *GS*. Tidigare slutsatser i den bibliometriska forskningen om att *GS* täcker in en mycket bredare grupp av citeringar bekräftas således för vår grupp av forskare.

Förklaringsgraden i sambandet mellan citeringar i *GS* och de andra databaserna är hög ($R^2 = 0,85$ för *SSCI*; $R^2 = 0,76$ för *Scopus*). Det betyder att även om citeringarna är fler i *GS* finns en stark koppling mellan att ha fler citeringar i *GS* och fler citeringar enligt de två andra måtten. *GS* kan således fungera förhållandevis väl som *proxy* för att mäta graden av inomakademiska citeringar. Dock finns ett fåtal forskare vars citeringsmönster avviker från det skattade sambandet. Det gäller särskilt en forskare som har näst flest citeringar enligt *GS* (strax under 10 700) och strax över 400 i *SSCI*. Baserat på antalet *SSCI*-citeringar är antalet *GS*-citeringar mer än fyra gånger större än vad regressions sambandet förutsäger.

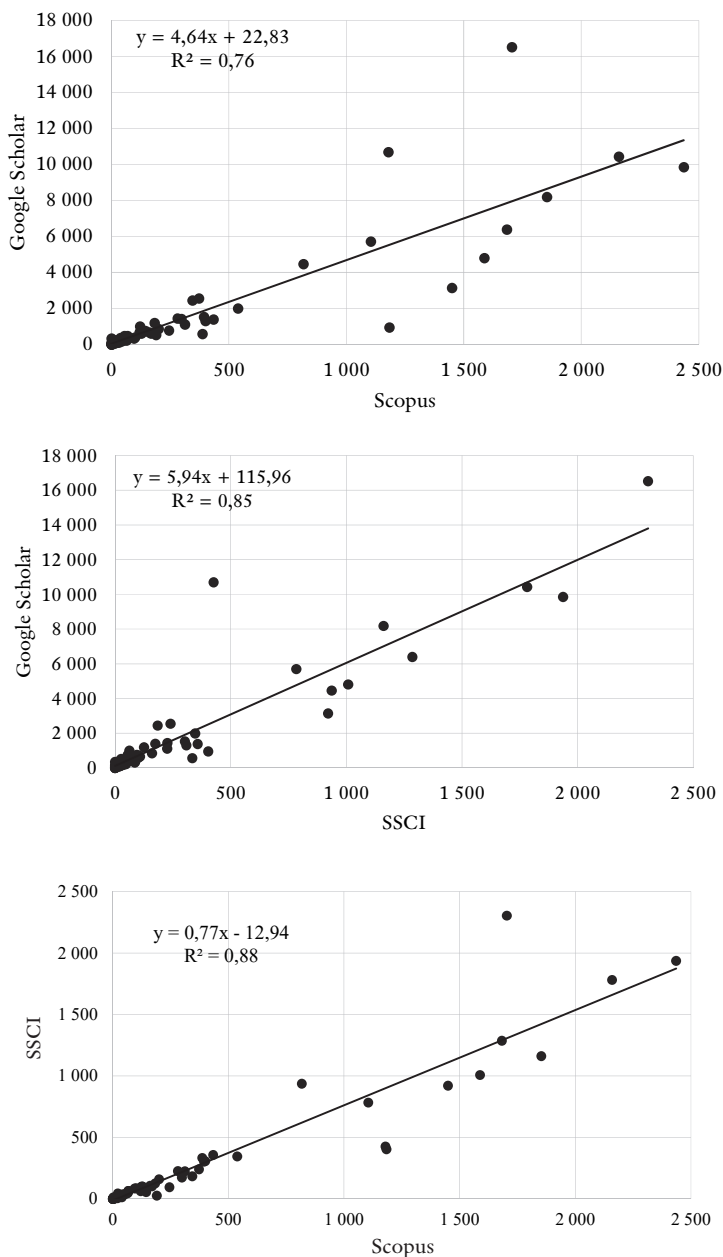
För en av de tio mest välciterade forskarna är i stället antalet *GS*-citeringar avsevärt lägre (eller citeringar i *SSCI* högre) än det skattade sambandet. Citeringarna i *GS* borde enligt skattningen vara ca 1,8 gånger så många (ca 5 500 jämfört med ca 3 100). Diskrepansen blir än större i jämförelsen mellan *GS* och *Scopus* (det skattade sambandet predikterar ca 6 500 *GS*-citeringar). Där är det också fler bland dem med störst antal citeringar enligt båda måtten som har klart färre citeringar i *GS* (eller högre i *Scopus*) än vad som kan förväntas. Det mest extrema exemplet är en forskare med nästan 1 200 citeringar i *Scopus* som bara har drygt 900 i *GS*.

Vi gör också en – om än mer subjektiv – bedömning av forskarnas inriktning. Utan namns nämnande finner vi ett tydligt mönster: De som har större fokus mot policyanalys ligger högre i *GS* än i *SSCI* och vice versa för dem vars forskning är renodlat inomvetenskaplig.

Den forskare som har flest citeringar enligt både *SSCI* och *GS* får betydligt färre citeringar i *Scopus* än förväntat. Givet hans citeringar enligt *GS* (ca 16 500) borde citeringarna i *Scopus* vara närmare 3 600. I realiteten är de bara drygt 1 700. I jämförelsen mellan *GS* och *SSCI* avviker han mindre från det skattade linjära sambandet. Denne forskare är bland de äldsta i samplet, och de få citeringarna i *Scopus* reflekterar att databasen ännu i

¹¹ I de resultat vi redovisar inkluderas självciteringar i måtten från *SSCI*. Analyserna har också gjorts där självciteringar exkluderas. Detta påverkar inte slutsatserna.

Figur 1
Jämförelse av antalet
citeringar enligt de
tre olika citeringskäl-
lorna



Källa: Google Scholar (2017), Scopus (2017) och SSCI (2017).

liten utsträckning inkluderar arbeten publicerade före 1996. I takt med att så sker kan vi förvänta oss att denna skillnad för de äldre forskarna försvinner.¹²

¹² Redan i början av 2016 samlade vi in data för denne forskare. Det visar sig att han haft en mycket stor ökning i antal citeringar i *Scopus*, mycket större än i de andra databaserna, vilket endast kan förklaras av att *Scopus* uppdaterats med äldre verk.

Vi har också testat att exkludera de tolv mest citerade forskarna – vilka ligger betydligt högre än resterande forskare enligt de tre måtten – från samtliga analyser. Sambandet mellan *Scopus* och *GS* blir då något starkare, vilket är rimligt eftersom de äldre mest citerade forskarna, för vilka skillnaderna var stora, då exkluderas. Sambandet mellan *SSCI* och *GS* blir i stället klart svagare ($R^2 = 0,66$) och förklaringsgraden mellan *SSCI* och *Scopus* blir något starkare ($R^2 = 0,92$). Riktningkoefficienterna blir också klart lägre när *GS* jämförs (på *y*-axeln) mot de andra två måtten.

Rangordningen efter citeringar

Figur 2 visar rangordningen av forskarna enligt de olika citeringskällorna mot varandra. När det gäller rangordningen av forskare utifrån citeringar är kopplingen mellan *GS* och de andra måtten starkare än vad vi förväntade oss *a priori*. Förklaringsgraden är hög när *GS* jämförs med både *Scopus* ($R^2 = 0,89$) och *SSCI* ($R^2 = 0,88$). Högst är den dock mellan *SSCI* och *Scopus* ($R^2 = 0,96$), vilket är en betydligt starkare koppling än vid den tidigare jämförelsen av antalet citeringar. Men skillnaderna är stora för ett fåtal forskare. I jämförelsen mellan *GS* och de andra databaserna sticker en forskare ut med en skillnad i rang på 25 steg mellan *GS* och såväl *Scopus* som *SSCI*. Forskaren har tidigare arbetat som utredare på hög nivå.

Den genomsnittliga absoluta skillnaden i rang är störst mellan *GS* och *SSCI* (5,2 steg), något lägre mellan *GS* och *Scopus* (5,1 steg) och lägst mellan *SSCI* och *Scopus* (2,6 steg).

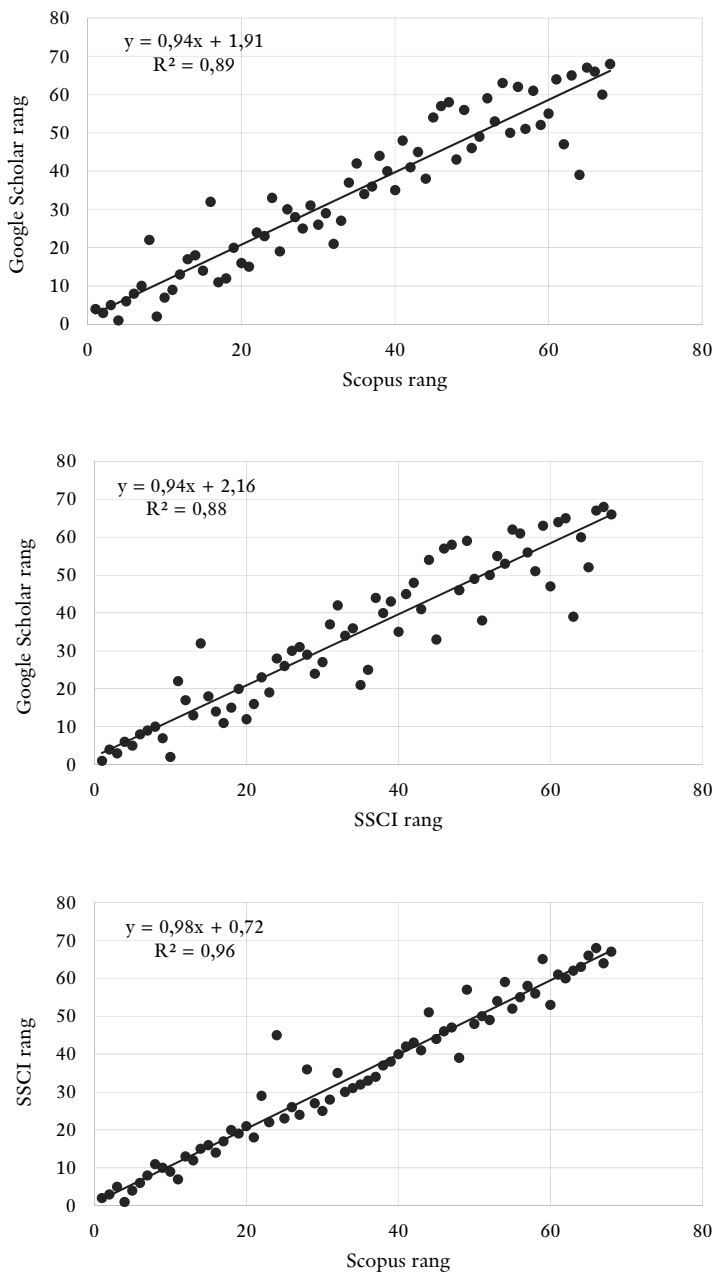
Skillnader i h-index mellan olika citeringskällor

h-index påverkas inte av att en forskare har någon eller några extremt väl-citerade artiklar, utan det krävs många verk som blir citerade för att få ett högre *h*. Skillnaderna mellan forskarna blir därmed mindre när *h*-index jämförs. Det högsta värdet på *h*-index i våra data är 53 och baseras på citeringar från *GS*. Det innebär att denne forskare publicerat minst 53 verk som vart och ett blivit citerat minst 53 gånger. Motsvarande siffra är 24 för *SSCI* och 28 för *Scopus*.

Figur 3 redovisar de parvisa sambanden för *h*-index enligt de tre citeringsmåtten. Då framgår att det även för *h*-index finns stora skillnader mellan måtten för några av forskarna, vilket indikerar att det inte bara är någon enstaka artikel som ligger bakom skillnaderna. Ett annat tecken på detta är att förklaringsgraden bara blir aningen högre än när citeringar jämförs i figur 1. Den tidigare slutsatsen att de mer policyinriktade forskarna får större genomslag i *GS* jämfört med de andra databaserna stärks också.

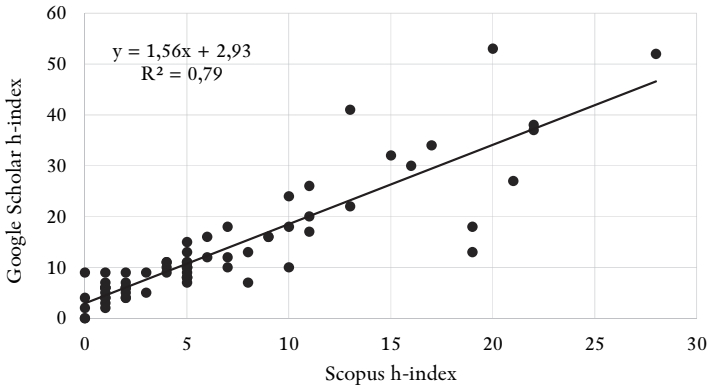
Vidare analys visar att det inte är självklart att rangordningen enligt *h*-index överensstämmer med rangordningen efter antal citeringar för alla forskare. Detta gäller för samtliga tre citeringsmått. Det betyder att utfallet av en jämförelse kan se mycket olika ut beroende på vilken vikt man vill lägga vid totalt genomslag jämfört med ett större antal artiklar som är relativt välciterade. I praktiken handlar det om att ta ställning till huru-

Figur 2
Jämförelse av rangordningen enligt de tre olika citeringskällorna

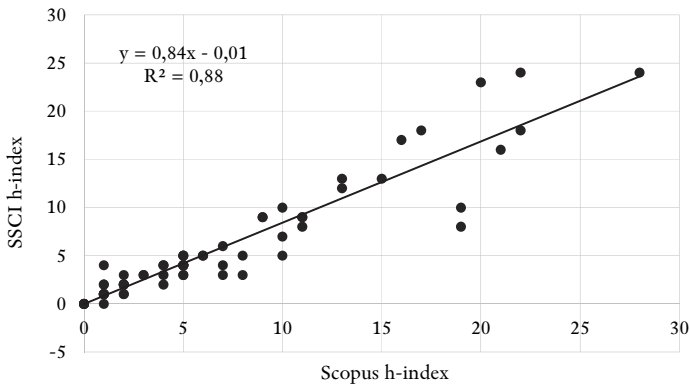
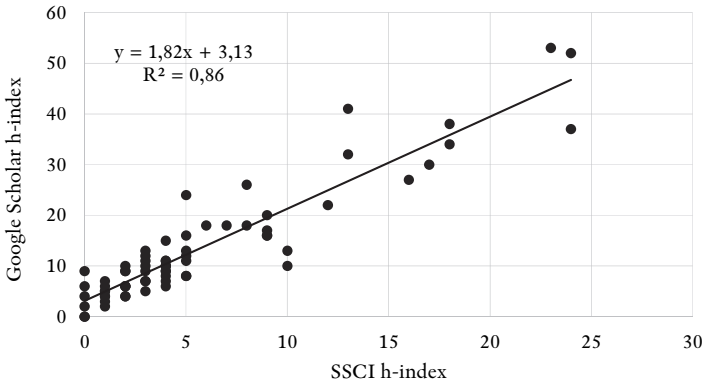


Källa: Google Scholar (2017), Scopus (2017) och SSCI (2017).

vida man värderar de forskare högst som publicerar förhållandevis lite men lyckas göra banbrytande arbeten, eller om man vill premiera forskare med stor produktion av hög och jämn kvalitet.



Figur 3
Jämförelse av h-index
enligt de tre olika
citeringskällorna



Källa: Google Scholar (2017), Scopus (2017) och SSCI (2017).

3. Slutsatser och reflektioner

De viktigaste argumenten för att använda kvantitativa mått på forskningsproduktion och vetenskapligt genomslag är att de är standardiserade och transparenta, att man slipper förlita sig på potentiellt subjektiva utvärde-

rare och att man kan undvika effekter av att utvärderare kan vara bekanta med vissa forskares bidrag sedan tidigare. Resurstilldelningen till svenska universitet är redan i dag delvis baserad på hur ofta artiklar av lärosätets forskare citeras. Med tanke på de stora anslag som ges till forskning är det viktigt att anslagen går dit där de gör mest nytta och får störst effekt. Å ena sidan är således de potentiella fördelarna med att mäta forskning stora ifall det ger en effektivare resursallokering. Å andra sidan är alla mått behäftade med svagheter. Hur som helst är det uppenbart att det är viktigt att använda så rättvisande och ändamålsenliga mått som möjligt.

Vi har studerat om valet av citeringsdatabas spelar någon större roll för hur stort en forskares vetenskapliga bidrag uppskattas vara relativt andra forskare baserat på alla forskare knutna till IFN. Vi bedömer att resultaten skulle stå sig väl även om urvalet utvidgades till andra akademiskt verksamma nationalekonomer i Sverige.

Vår analys visar att *SSCI* och *Scopus* stämmer väl överens för de allra flesta. *GS* ger nästan undantagslöst ett betydligt större antal citeringar än de två andra måtten (totalt antal citeringar för alla forskare är fem till sex gånger fler), men kopplingen mellan att ha fler citeringar i *GS* och att ha fler i *SSCI* och *Scopus* är stark. Samtidigt är skillnaderna mellan måtten stora för några av forskarna. Så är också fallet när vi jämför *h*-index. Om en utvärdering skulle göras enbart baserat på ett av måtten skulle utfallet kunna se mycket olika ut för dessa forskare.

Äldre forskare tillgodoses inte alla sina verk och citeringar i *Scopus* då databasen inte är heltäckande före 1996. Detta är dock ett övergående problem eftersom *Scopus* utvidgar sin databas bakåt i snabb takt. Vidare får mer policyorienterade forskare ett större genomslag enligt *GS* jämfört med dem som sysslar med frågor som främst är av inomvetenskapligt intresse, vilket är i linje med tidigare forskning om databasernas täckningsgrad och avgränsningar.

Den som använder en databas för att utvärdera forskning bör ta hänsyn till vad den faktiskt mäter och vilka avgränsningar som görs. Då det för vissa forskare finns stora skillnader mellan de olika måtten bör flera mått användas. *Scopus* eller *SSCI* är bäst för att värdera det inomvetenskapliga genomslaget för en forskares arbeten. För *Scopus* talar att den har bättre täckning än *SSCI* och i takt med att *Scopus* blir mer heltäckande bakåt i tiden ökar dess relativa försteg.

Då *GS* även beaktar en forskares genomslag på policyanalys och kvalificerad samhällsdebatt ger detta mått information utöver vad som beaktas av *SSCI* och *Scopus*. Som Björklund (2014) pekar på finns en konflikt mellan att lägga tid på att producera toppforskning och den tredje uppgiften – att delta i ekonomisk-politisk debatt och med utredningar. Vilken avvägning som här görs bör styra vilken vikt som läggs vid olika mått.

För samtliga mått kan också ett *h*-index beräknas, vilket ger kompletterande information genom att det erbjuder ett alltmer vedertaget sätt att mäta både kvantiteten och kvaliteten i någons forskning.

Vilken roll bör då de mått vi diskuterat i denna artikel spela vid tjänsteställningar och anslagsfördelning? Först bör konstateras att det inte existerar några perfekta substitut till självständiga kvalitativa sakkunnigbedömningar av forskares vetenskapliga bidrag. Detta innebär bl a att sakkunniga inte helt sonika kan anta att kvaliteten på en forskares arbeten direkt kan avläsas genom att se i vilken tidskrift arbetena publicerats. Från senare år finns några nedslående exempel på utlåtanden som i stället för att utvärdera kvaliteten på forskares arbeten bedömer tidskrifterna där de publicerats (eller snarare bara återger professionens samlade bedömning av de aktuella tidskrifterna). Detta betyder inte att det är irrelevant i hur högt rankade tidskrifter en forskare publicerar sig. Det bidrar med ett kvantitativt och transparent mått.

Citeringsmått bidrar med ytterligare information och här finns starka skäl att använda både antalet citeringar i *Scopus* (alternativt *SSCI* även om vi förordar *Scopus*) och *GS* i kombination med *h*-index för de två måtten. Samtidigt som citeringsmåtten är transparenta ställer det krav på att utvärderare som använder dem väger in att det finns skillnader i citeringsfrekvens mellan fält, att ett verk ibland blir välciterat på grund av att det visade sig vara fel, att det inte beaktas hur högt rankad den tidskrift är i vilken verket citeras och att det inte sker någon justering för medförfattare.

Huvudslutsatsen av vår analys kan kort sammanfattas som att citeringsmått – både absolut antal och *h*-index – inte kan ersätta kvalitativa utvärderingar men att de bidrar med viktig information om en forskares kompetens och meriter. Såväl *Scopus* (alternativt *SSCI*) som *GS* ger värdefull information och båda måtten har ett självständigt värde.

Bergstrom, C T (2007), "Eigenfactor: Measuring the Value and Prestige of Scholarly Journals", *College and Research Libraries News*, vol 68, s 314–316.

Björklund, A (2014), "Nationalekonomisk toppforskning i Sverige – omfattning, lokalisering och inriktning", *Ekonomisk Debatt*, årg 42, nr 5, s 6–19.

Bornmann, L, A Thor, W Marx och H Schier (2016), "The Application of Bibliometrics to Research Evaluation in the Humanities and Social Sciences: An Exploratory Study Using Normalized Google Scholar Data for the Publications of a Research Institute", *Journal of the Association for Information Science and Technology Science*, vol 67, s 2778–2789.

Bornmann, L och W Marx (2015), "Methods for the Generation of Normalized Citation Impact Scores in Bibliometrics: Which Method Best Reflects the Judgement of Experts?", *Journal of Informetrics*, vol 9, s 408–418.

Callaway, E (2016), "Beat It, Impact Factor!

Publishing Elite Turns against Controversial Metric", *Nature*, vol 535, s 210–211.

Card, D och S Della Vigna (2013), "Nine Facts about Top Journals in Economics", *Journal of Economic Literature*, vol 51, s 144–161.

Delgado López-Cózar, E, N Robinson-García och D Torres-Salinas (2014), "The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators", *Journal of the Association for Information Science and Technology*, vol 65, s 446–454.

Ek, S och M Henrekson (2018), "Citeringar – en jämförelse av tre olika mått", IFN Policy Paper nr 81, Institutet för Näringslivsforskning, Stockholm.

Google Scholar (2017), databas, scholar.google.se.

González-Pereira, B, V P Guerrero-Bote och F Moya-Aneón (2010), "A New Approach to the Metric of Journals' Scientific Prestige: The SJR Indicator", *Journal of Informetrics*, vol 4, s 379–391.

REFERENSER

- Görnerup, E (2013), "Från departement till doktorand – på vilka grunder fördelas de direkta statsanslagen för forskning?", rapport, Svenskt Näringsliv, Stockholm.
- Harzing, A W (2007), *Publish or Perish*, webbsida, harzing.com/resources/publish-or-perish.
- Harzing, A W och S Alakangas (2016), "Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison", *Scientometrics*, vol 106, s 787–804.
- Henning, K (2013), "Bedömning av publikationskanaler i de norska och danska bibliometriska modellerna – konsekvenser för humaniora och samhällsvetenskap", rapport, Göteborgs universitetsbibliotek, Göteborg.
- Henrekson, M och D Waldenström (2011), "How Should Research Performance Be Measured? A Study of Swedish Economists", *Manchester School*, vol 79, s 1139–1156.
- Hirsch, J E (2005), "An Index to Quantify an Individual's Scientific Research Output", *Proceedings of the National Academy of Sciences*, vol 102, s 16569–16572.
- Holmström, B och P Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization*, vol 7, s 24–52.
- Jacsó, P (2010), "Metadata Mega Mess in Google Scholar", *Online Information Review*, vol 34, s 175–191.
- Kalaitzidakis, P, T P Mamuneas och T Stengos (2003), "Rankings of Academic Journals and Institutions in Economics", *Journal of the European Economic Association*, vol 1, s 1346–1366.
- Kesselberg, M (2015), "Forskningsresurser baserade på prestation – tilldelning och omfördelning av basanslag till forskning och utbildning på forskarnivå baserat på indikatorer 2009–2014", rapport 2015:5, Universitetskanslersämbetet, Stockholm.
- Kim, E H, A Morse och L Zingales. (2006), "What Has Mattered to Economics since 1970?", *Journal of Economic Perspectives*, vol 20, s 189–202.
- Klein, D B och E Chiang (2004), "The Social Sciences Citation Index: A Black Box – With an Ideological Bias?", *Econ Journal Watch*, vol 1, s 134–165.
- Kremer, R D, A Laing, B Galliers och A Kiem (2015), *ABS Academic Journal Guide 2015*, Chartered Association of Business Schools (ABS), London, charteredabs.org/academic-journal-guide-2015.
- Laband, D N och R D Tollison (2003), "Dry Holes in Economic Research", *Kyklos*, vol 56, s 161–174.
- Lindqvist, T (2003), "Nationalekonomisk forskning i Sverige – publiceringar och ranking av forskare och institutioner", *Ekonomisk Debatt*, årg 31, nr 3, s 21–30.
- Lozano G A, V Larivière och Y Gingras (2012), "The Weakening Relationship between the Impact Factor and Papers' Citations in the Digital Age", *Journal of the American Society for Information Science and Technology*, vol 63, s 2140–2145.
- LSE (2011), *Maximizing the Impacts of Your Research: A Handbook for Social Scientists*, London School of Economics and Political Science, London.
- Mingers, J och H Willmott (2013), "Taylorizing Business School Research: On the 'One Best Way' Performative Effects of Journal Ranking Lists", *Human Relations*, vol 66, s 1051–1073.
- Mingers, J och L Yang (2017), "Evaluating Journal Quality: A Review of Journal Citation Indicators and Ranking in Business and Management", *European Journal of Operational Research*, vol 257, s 323–337.
- Mutz, R och H D Daniel (2012), "Skewed Citation Distributions and Bias Factors: Solutions to Two Core Problems with the Journal Impact Factor", *Journal of Informetrics*, vol 6, s 169–176.
- Oswald, A J (2007), "An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers", *Economica*, vol 74, s 21–31.
- Sauder, M och W N Espeland (2009), "The Discipline of Rankings: Tight Coupling and Organizational Change", *American Sociological Review*, vol 74, s 63–82.
- Scopus (2017), databas, www.scopus.com.
- SSCI (2017), databas, www.webofknowledge.com.
- Waltman, L (2016), "A Review of the Literature on Citation Impact Indicators", *Journal of Informetrics*, vol 10, s 365–391.
- Waltman, L och N J van Eck (2013), "A Systematic Empirical Comparison of Different Approaches for Normalizing Citation Impact Indicators", *Journal of Informetrics*, vol 7, s 833–883.