

Bra forskning är jämnt fördelad över svenska lärosäten

**Rapport till KK-stiftelsen av docent Ulf Sandström
4 juli 2018 (version 4:1)**

Många tar för givet att forskning med svagt genomslag är koncentrerad till vissa småskaliga universitet högskolor. Detta motiverar närmare undersökning eftersom det förhållandet att svensk forskning i väsentlig grad skulle förbättras om verksamheten flyttades från de regionala högskolorna till universiteten behöver i så fall beläggas med fakta. Om det är så att de stora universiteten dragit ifrån och gör bättre resultat än vad som framgick av en tidigare undersökning (Sandström 2015) borde detta kunna förklaras av att forskningsresurserna kanaliserats till dessa lärosäten. Men frågan är hur det egentligen ser ut? Har de stora dragit ifrån och har de små förlorat i samma mån?

Metod och genomförande: Rapporten grundar sig på svenska publiceringar under perioden 2012-2015 med citeringar fram till och med 2017. Rapporten tillämpar bibliometriska metoder: Undersökningen använder i huvudsak en s.k. percentilmodell vilken på ett rimligt sätt ställer samman produktionen från enskilda forskare till ett sammanlagt värde när citeringsfrekvens har lagts till. Modellen för undersökning är således densamma som Sandström (2015).

Resultat: Undersökningens huvudsakliga resultat är att kvalitetsskillnader mellan universiteten å ena sidan och de nya universiteten och högskolorna å den andra förefaller ha uttraderats. Detta trots att de finansieringsströmmar som initierades av alliansregeringen mellan 2006 och 2014 mestadels gick till de stora universiteten och specialhögskolorna (se Sandström et al. 2010; Sandström & Wold 2015).

Tumregeln är att de gamla och breda universiteten gör knappa 60 procent av produktion och citeringar, de specialiserade universiteten (tekniska högskolor, KI och SLU) gör drygt 30 procent och de nya universiteten och högskolorna fördelas relativt jämnt på de återstående tio procenten. Tendensen är dock att nya universitet och högskolor får en allt mindre del av percentilpoängen. Detta förklaras i första hand av att universiteten relativt sett ökar i de lägre decentilerna, de tar över personal från sjukhus, institut och företag vilka tidigare tog hand om stora delar av de lågproduktiva forskarna.

Det som omedelbart skulle kallas "forskning med svagt genomslag" är forskning i den undre halvan av decentilerna, dvs. från decentil 6 till decentil 10. Högskolorna och de nya universiteten är klart underrepresenterade i denna kategori medan universiteten har överrepresentation. En tolkning är att högskolorna är yngre och är på väg att

etablera sig som forskningsproducenter och gör bra resultat till följd av att de saknar barlast. För detta talar också det förhållandet att nya universitet och högskolor presterar ungefär likvärdiga resultat även om de nya universiteten har ett visst försteg när vi avgränsar oss till verkan av forskningskvalitet.

Kostnaderna för forskning med svagt genomslag kan definieras på olika sätt. Nollciterade artiklar är ett sätt att operationalisera detta begrepp. Här har vi valt att använda artiklar som inte når över 50 % mest citerade per tidskriftsklass. Med den definitionen framkommer att stora delar av forskningen med svagt genomslag är koncentrerad till de gamla universiteten.

Bibliometrisk metod för percentilmodellen

Relativa citeringsindikatorer har stort informationsvärde, men det är lätt att de misstolkas. Framförallt bör de hanteras med försiktighet eftersom de är känsliga för hur stor del av forskarna som satsats ut på publiceringsmarknaden. Ju fler nya forskare som ingår desto fler med svagt genomslag, och relativa indikatorer gör att en grupp med några bra publiceringar blir successivt sämre när fler ur gruppen börjar publicera artiklar som inte uppmärksammas, dvs. inte erhåller citeringar.

Med syftet att etablera ett alternativt sätt presenteras här percentilmodellen som istället lägger ihop, adderar, prestationer.

Databasen består av 55 850 svenska forskare, varav ca.47 000 med anknytning till universitet och högskolor, och täcker publiceringar under perioden 2012–2015. Närmare 100 000 artiklar och 304 000 författarandelar har identifierats tillhöra svenska organisationer. I några få fall har artiklar från personer som haft verksamhet såväl i Sverige som i något annat nordiskt land hållits samman och artiklar således tillgodoräknats även om de tillkommit utanför Sverige (särskiljningen av namn utförs på nordisk nivå). Vanligt förekommande efternamn och förnamnsinitial ingår inte i underlaget. Detta bortfall påverkar inte resultatet (särskild undersökning har genomförts).

Varje forskares genomslag har beräknats på artiklar publicerade 2012–2015 med citeringar under perioden 2012–17. Rapporten räknar bråkdelar (fraktioner) av artiklar – om fem forskare skrivit en artikel får varje forskare 0,2 andelar av artikeln.

Varje artikel rangordnas, baserat på dess citeringar, inom respektive vetenskapsfält, vilket definieras utifrån de ämnesklasser (cirka 260 i antal) som används i *Web of Science*, och delas in i percentilgrupper (de 1 procent, 5 procent, 10 procent högst rankade och så vidare). Mått baserade på percentiler har fördelen att de inte påverkas av skevheter i citeringsfördelningar. Inom vissa vetenskapsområden kan några

publikationer med ett mycket stort antal citeringar annars dra upp medelvärdet, vilket kan leda till att 70 procent av artiklarna inom området ligger under medelvärdet.

Percentilindikatorn har "översatts" till ett poängtal för varje artikel, där en artikel som tillhör den mest citerade procenten av artiklarna tilldelas 100 poäng, topp-5-procent ger 20 poäng, och så vidare (se tabell 1). En artikel som hör till de 50 procent minst citerade ger 1 poäng, vilket gör att en forskare aldrig kan förlora på att publicera en artikel. Metoden bygger på sannolikheten att få en artikel citerad så att den uppfyller villkoren för respektive percentilgrupp.

Tabell 1. Poängfördelning till percentilklasser

Percentilgrupp	Poäng
TOP1 %	100
TOP5 %	20
TOP10 %	10
TOP25 %	4
TOP50 %	2
TOP100 %	1

Anm: Baserad på Sandström & Wold (2015)

De poäng som varje artikel sålunda erhåller justeras genom den viktning via Field Adjusted Production (FAP), metoden för detta beskrivs i Sandström & Sandström *Resurser för citeringar* (2009). Poängen med denna metod är att kompensera för olikheterna i vetenskaplig produktionsvolym mellan forskningsområden. Samtliga tidskrifter i *Web of Science* har kategoriserats till ett av sex områden (applied sciences, natural sciences, astronomy & high energy physics, health sciences, economic & social sciences samt arts & humanities). På basis av detta har waring-metoden gett underlag för en FAP-faktor. Måttet är således ett kompositmått som i ett enda värde uttrycker produktivitet (antal artiklar) och citeringsnivå (kvalitet). Fördelen, jämfört andra liknande mått, exempelvis h-index, är att detta mått är konstruerat för att användas över och mellan samtliga vetenskapsområden. Detta gör att vi kan jämföra mellan universitet och även mellan grupper av universitet. Det kan h-index inte göra.

I jämförelse med den tidigare undersökningen från 2015 har ett forskningsområde tillkommit "Astronomy / High Energy Physics". Anledningen är att det under perioden framkommit ett stort antal artiklar som samförfattats av stora forskarkonstellationer ofta med 3 000 – 5 000 författare. Även om en enskild författare har cirka 350 artiklar under perioden 2012-2015 uppgår den sammanlagda fraktionen för en enskild författare inte till mer än 0,2 artikelfraktioner under perioden. Högenergiforskarna skulle således behöva skriva 1200-1500 artiklar tillsammans för att åstadkomma förväntat värde om 1 artikel under perioden. Av detta skäl har det bedömts nödvändigt att sortera ut detta

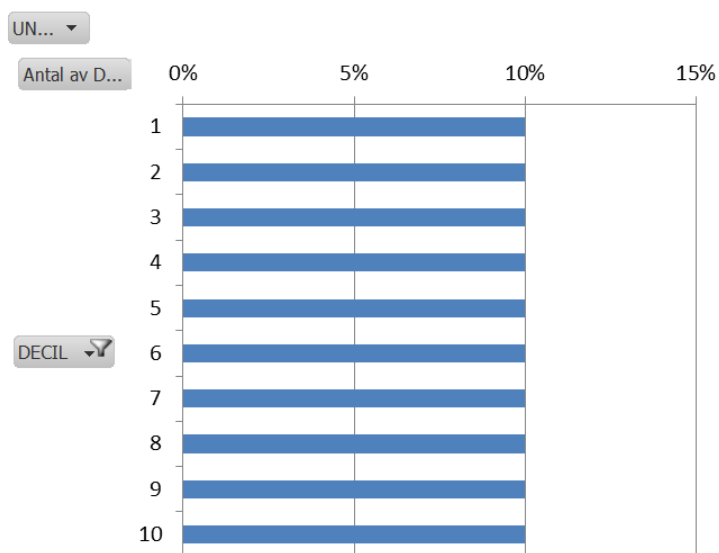
ämnesområde och en särskild beräkning på hela materialet med A-F som särskild kategori har utförts.

En forskare som identifieras enligt metodiken ovan får en poängsumma baserad på artikelandelar och dessas citeringsbaserade poäng. På basis av detta går det att rangordna samtliga svenska forskare. Resultaten på Sverigenivå blir högst rimliga i så måtto att av de hundra högst rankade forskarna är 40 % inom *Life Science*, 40 % inom *Natural and Engineering Sciences* och 20 % inom *Humanities and the Social Sciences*. Samma fördelning finns på andra aggregeringsnivåer.

Decentilprofiler för att analysera prestationer

En högskola eller ett universitet som har en prestationsprofil som överensstämmer med den svenska kommer att ha 10 % av sin personal i respektive decentil (tiondel), se figur 1.

Figur 1: Ideal fördelning över decentiler



Verkligheten är förstås en annan än idealmodellen. Vissa universitet har en överrepresentation i de övre decentilerna, andra har lägre representation bland dessa topprestationer. Vad vi nu intresserar oss för är de högre decentilerna, dvs decentilerna 1-3. Hur ser det ut för olika grupperingar av lärosäten? Har toppgrupperna koncentrerats till de stora lärosätena eller finns de jämnt utspridda över samtliga grupper?

Det bör klart och tydligt sägas att definitionen av att vara forskare vid ett lärosäte helt enkelt är att man har publicerat minst en artikel under den aktuella perioden. Nämnas kan att vid sidan av lärosätena är det drygt 3 300 är forskare vid företag (Aktiebolag) och drygt 2 200 är forskare i institutssektorn (RISE-instituten och andra statliga institut och institutsliknande verksamheter t ex Naturhistoriska museet). Knappa 1 000-talet har sin

huvudsakliga anknytning till sjukhus vilka inte har fördelats till universiteten såsom del av deras verksamhet (s.k. universitetssjukhus).

Av företagsforskarna har mer än 1 100 med anknytning till Astra Zeneca. Av dessa har en mycket stor andel mindre än 0,33 artikelfraktioner under perioden, 350 har färre än 0,2 artikelfraktioner. Detta illustrerar ett problem som uppkommer i anslutning till frågan om hur många personer är det som är forskare? Många av de som arbetar på Astra Zeneca är inte i första hand publicerande forskare utan tekniker som arbetar för att ta fram nya produkter åt företaget. Publicering ses som en biprodukt och något som görs då man uppfattar det som produktivt att släppa ut viss information, dels för att attrahera forskare i närliggande områden, dels för att effektivisera den teknologiska avspaningen. Något egentligt svar på frågan hur många publicerande forskare det finns vid det aktuella företaget ges förstås inte, utan diskussionen får vara svar på frågan.

En kontroll har möjliggjorts med tillgång till uppgifter om namn på publicerande forskare vid Uppsala universitet (UU) under perioden 2008-2013 kan vara värd att rekapitulera även här. Intressant med den listan är att den upptar namn som institutionerna acceptar som affilierade på ett eller annat sätt utan att alla nödvändigtvis har en formell anställning. Dessa har matchats mot percentildatabasens namn och resultatet är att den förra består av 5 000 forskare (publicerande personer) och den senare likaledes av 5 000 namn. Här finns således ett antal publicerande läkare och sjuksköterskor vid Akademiska sjukhuset, informella gästforskare, honorary doctors, och icke registrerade stipendiater. Den goda överensstämmelsen med UU:s egen databas tyder på att disambigueringen (disambiguering=eliminera osäkerhet) och analysen är någorlunda korrekt utförd.

Tyvärr löser detta inte vårt egentliga problem som har att göra med definitionen av en "forskare". Universiteten drar till sig ett relativt stort antal personer som sannolikt har en pågående publiceringsverksamhet vid andra lärosäten *utanför* Sverige. De har också ett relativt stort antal mastersstudenter och stipendiater som blir tillfälliga gäster. Hur skall detta problem hanteras? Om vi inte tar bort dessa personer kommer det oförskyllt att se ut som att universiteten har en omfattande verksamhet med förhållandevis lågt genomslag. Det är sannolikt att detta problem drabbar Karolinska Institutet med större kraft än andra lärosäten. Se resultatdelen för närmare uppgifter om detta.

Syftet med vår analys är att försöka ringa in verksamhet vid landets lärosäten utan att för den skull dra in de personer som aldrig varit särskilt ägnade åt att försöka sig på en att vara svenska *forskare*. Svårigheten är att det samtidigt riskerar att ta bort just det som vi samtidigt är ute efter i detta projekt.

Resultatredovisning

Bibliometriska material har en del egenskaper som vi tenderar att vara ovana vid och som många kan ha svårt att greppa. Snedheten i fördelningen, man kan kalla det ojämlikheten, mellan de som får många citeringar jämfört med flertalet som erhåller få citeringar, är ett sådant förhållande. Och det kan vara svårt att ta till sig detta problem. Det är häpnadsväckande stora skillnader mellan forskare men de statistiska regelbundenheterna bakom detta är synnerligen välkända och belagda i andra material. Således kan vi säga att det är helt i sin ordning.

En illustration är att balanspunkten i materialet ligger vid den 95:e percentilen, dvs. de fem översta procenten forskare producerar lika många citerings- och produktionspoäng som de övriga 95 procenten. Det var 94:e percentilen i förra rapporten. Koncentrationen till de få ser ut att vara trendmässig. Vidare känner vi igen Paretos lag i detta material: 20 procent av de mest produktiva gör 80 procent av artikelfraktionerna. De forskare som producerar många papers har större chans att få åtminstone en artikel högt citerad. Med få artiklar minskar chanserna och detta är tunga statistiska regelbundenheter som lägger grunden för de fördelningar vi här har att göra med.

Redan det faktum att de stora universiteten är så mycket större borde således ge ökade möjligheter att ha högciterade publikationer (se vidare Sandström & van den Besselaar i PLoS ONE 2016). Chansen att träffa rätt (i detta fall, att ha produktiv personal) blir så mycket större när urvalet är stort. Vi borde således kunna förvänta oss att de större universiteten i grupp 1 och 2 i tabell 2 (nedan) har en fördelning med en förhållandevis större tyngd i toppen medan grupp 3 och 4 har större sannolikhet att vara överrepresenterade i bottenskikten, dvs. en större andel forskare med svagt genomslag och liten produktion. Utöver detta kan också anföras en rad andra skäl till varför nya universitet och högskolor, som rimligen befinner sig i en uppbyggnadsfas av sin forskning, inte presterar på samma nivå som de reguljära universiteten (se ovan).

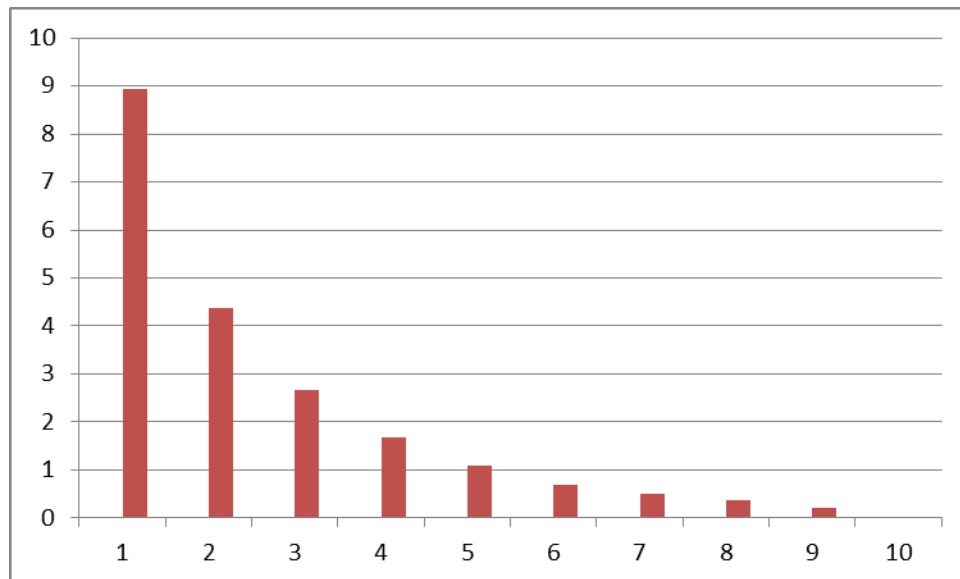
Tabell 2: Statistik för decentiler 2012-2015

GRUPP	Totalt i grp	Norm / decentil
Grupp 1 (gamla universitet)	26 450	2 645
Grupp 2 (specialinriktade universitet)	17 452	1 745
Grupp 3 (nya universitet)	2 052	205
Grupp 4 (regionala högskolor)	1 540	154

Denna förväntade fördelning kan dock motverkas av det faktum att universiteten dels attraherar tillfälliga gäster, dels organiserar forskning på ett sätt som i inte så liten utsträckning innebär att enskilda personer som kanske inte har anställning vid universitetet ändå kommer att uppge universitetets adress. Utöver detta är det också fler stipendiater och masterstudenter som ansluter till universiteten. Efter en inledande

redovisning som innefattar samtliga forskare skall vi studera effekten av att i en första omgång ta bort de som har färre än 0,2 artikelfraktioner och i en andra omgång 0,33 artikelfraktioner. Innan vi går till detta skall vi förklara förutsättningarna för att tillhöra olika decentiler (tiondelar).

Figur 2. Poängnivåer för decentiler



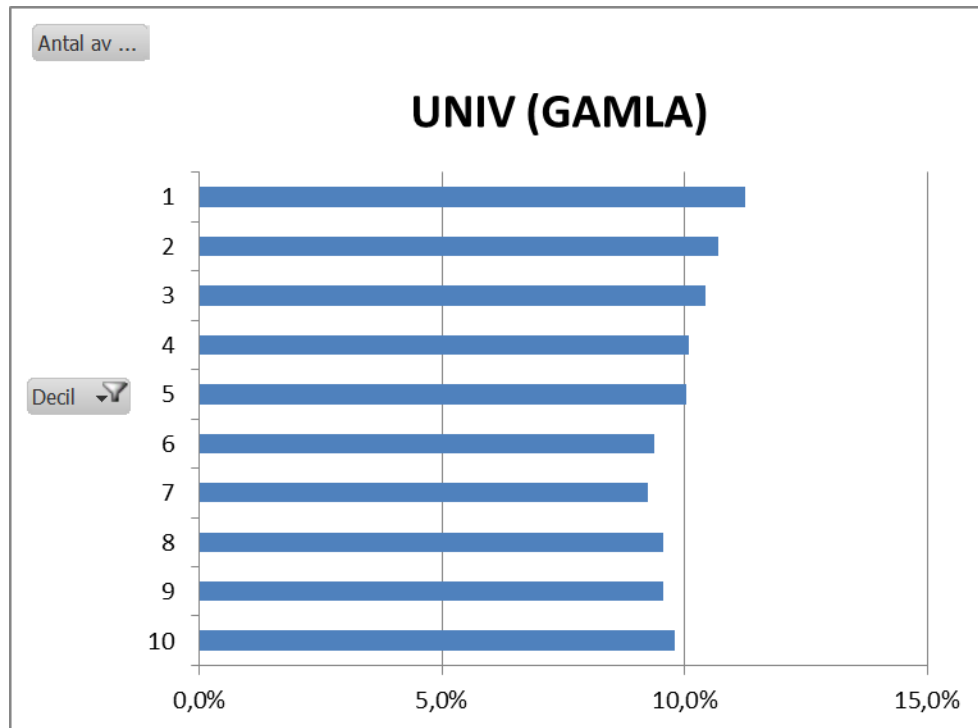
För att kvalificera till den översta, bästa, decentilen ska man ha producerat närmare nio poäng (figur 2). Det motsvarar lite mindre än att ha en artikel, som ensamförfattare, som blir citerad i topp 10 % av respektive tidskriftsklass. För att komma upp till decentil 2 krävs att forskaren ensam publicerat fyra artiklar plus ett litet antal citeringar (poäng ges även för ej citerade bidrag eftersom att publicera artiklar kräver en arbetsinsats).

Låt oss se närmare på figurerna 3-7. Där ges en profil för varje grupp av högskolor. I dessa framkommer att universiteten har prestationsprofiler som överensstämmer väl med förväntningarna och till och med något bättre. Redan genom sin storlek har de förutsättningar att ha produktiv personal, snittet för de gamla och breda universiteten ligger vid 2500 personer och för högskolorna är den inte större än 100 personer per högskola. **Toppen för de nya universiteten är förhållandevis stark, de översta två decentilerna är 22,5 procent och för högskolorna är denna topp drygt 21 procent. Motsvarande för universiteten och specialhögskolorna är 21,9 och 21,3. Typ av högskola ser inte ut att någon betydelse.**

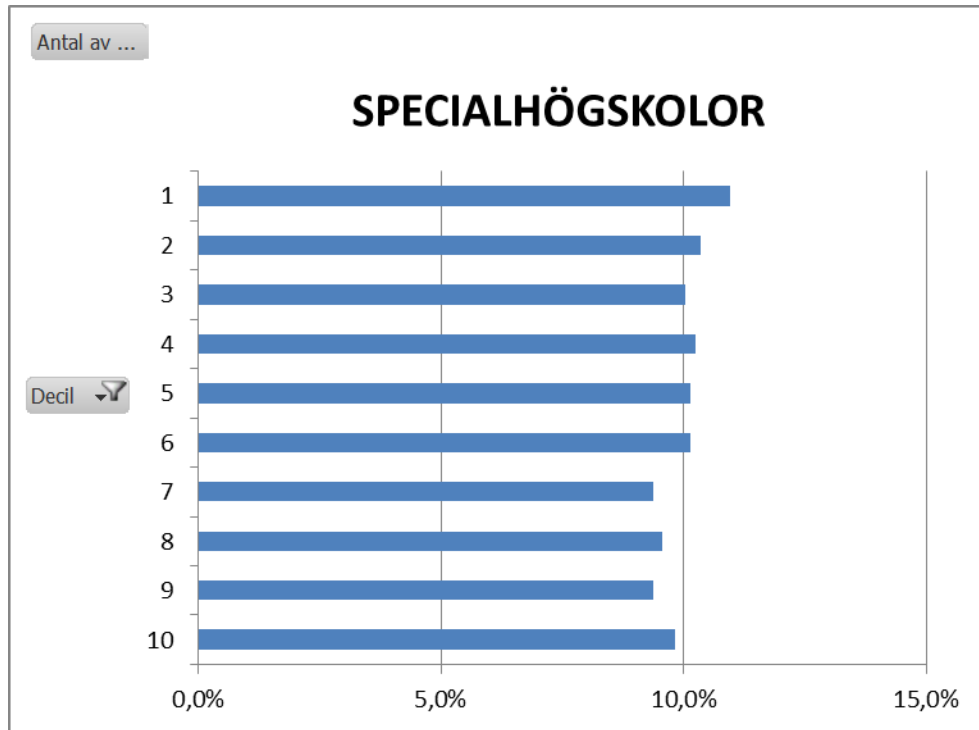
De lägsta decentilerna är påtagligt underrepresenterade för nya universitet och högskolor vilket illustrerar en viktig sida av sannolikhetsläran: att med en stor personal

(universitetens belägenhet) undkomma personer med svagt genomslag är närmast omöjligt.

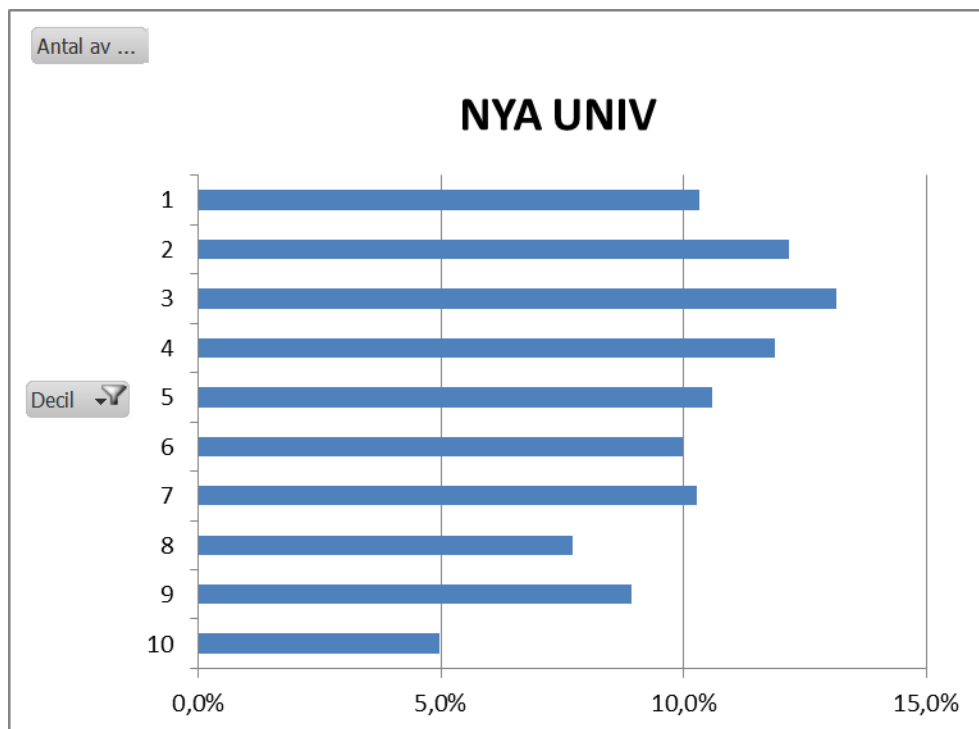
Figur 3: GU, LU, UU, SU, LiU, UmU (gamla)



Figur 4: KTH, CTH, KI, SLU, LTU, HHS (spec)



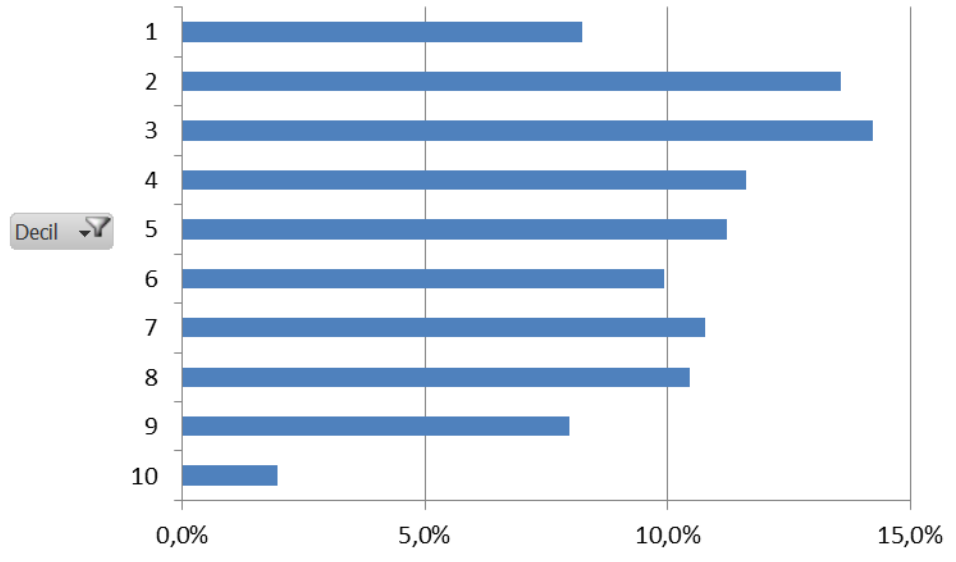
Figur 5: KAU, LNU, MIUN, ORU (nya univ)



Figur 6: HÖGSKOLOR

Antal av ...

HÖGSKOLOR



Decl ▾

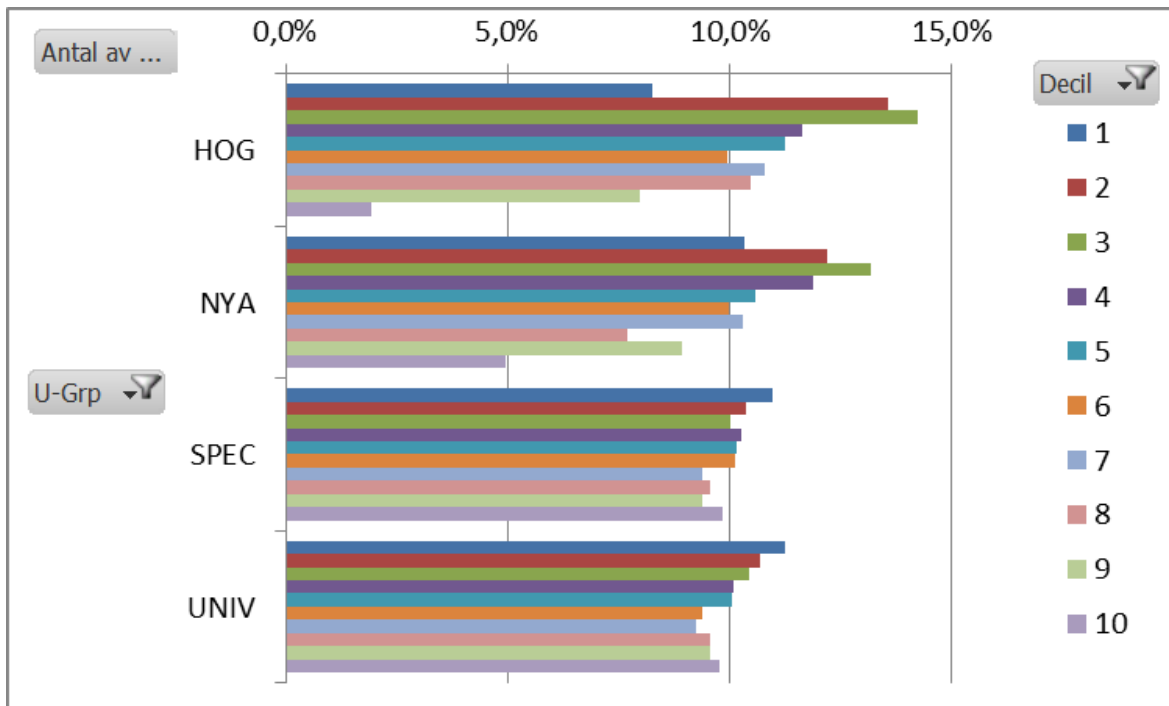
Trender och framtider

Det kan vara intressant att jämföra vad som har förändrats mellan de båda perioderna 2008-2011 och 2012-2015. Underlag för en sådan jämförelse framgår av figur 7 och 8 nedan. Figuren lägger högskoletyperna bredvid varandra med högskolorna överst och den bästa decentilen överst i respektive grupp.

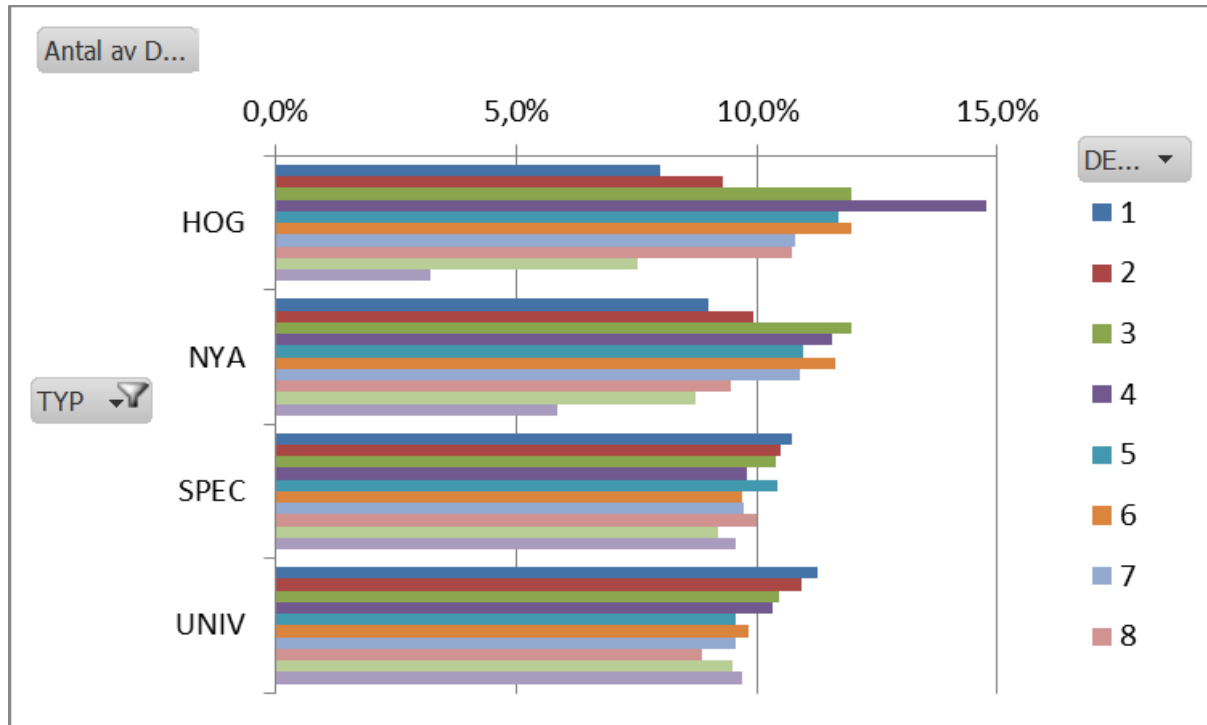
De trendmässiga förändringar som framträder är att högskolorna och de nya universiteten blir starkare i de översta decentilerna. Högskolorna har en ökning på 23-26 % beroende på om det är de två eller de tre översta decentilerna som räknas. Nya universitet har 16-19 % ökning mellan perioderna räknat på samma sätt. Detta medan universitet och specialhögskolor inte uppvisar några större förändringar, annat än att specialhögskolorna tar ytterligare någon liten andel från universiteten.

Figur 7-8: Jämförelse mellan två perioder

2008-2011



2012-2015



Tabell 3. Underlag 2008-2011

Grupp	1	2	3	4	5	6	7	8	9	10	Total
HOG	8,0%	9,3%	12,0%	14,8%	11,7%	12,0%	10,8%	10,7%	7,5%	3,2%	100%
NYA	9,0%	9,9%	12,0%	11,6%	11,0%	11,6%	10,9%	9,4%	8,7%	5,9%	100%
SPEC	10,7%	10,5%	10,4%	9,8%	10,4%	9,7%	9,7%	10,0%	9,2%	9,6%	100%
UNIV	11,3%	10,9%	10,5%	10,3%	9,6%	9,8%	9,6%	8,8%	9,5%	9,7%	100%
Total	10,9%	10,7%	10,5%	10,3%	10,0%	9,9%	9,7%	9,4%	9,3%	9,3%	100%

Tabell 4. Underlag 2012-2015

Grupp	1	2	3	4	5	6	7	8	9	10	Total
HÖG	8,2%	13,6%	14,2%	11,6%	11,2%	9,9%	10,8%	10,5%	8,0%	1,9%	100%
NYA	10,3%	12,2%	13,2%	11,9%	10,6%	10,0%	10,3%	7,7%	8,9%	5,0%	100%
SPEC	11,0%	10,4%	10,0%	10,2%	10,1%	10,1%	9,4%	9,6%	9,4%	9,8%	100%
UNIV	11,2%	10,7%	10,4%	10,1%	10,0%	9,4%	9,2%	9,6%	9,6%	9,8%	100%
Total	11,0%	10,7%	10,5%	10,3%	10,1%	9,7%	9,4%	9,5%	9,4%	9,3%	100%

Övriga underlag finns i bilaga.

Slutsats

Slutsatsen av denna undersökning baserad på publiceringsdata är att det inte finns belägg för myten att vissa lärosäten skulle producera sämre forskning. Snarare tyder resultaten på att högskolorna under den här aktuella perioden har kommit ifatt de gamla universiteten och specialhögskolorna. Vidare kan konstateras att de nya universiteten har befest sin ställning i forskningssystemet.

Resultatet av denna undersökning antyder att högskolorna ser ut att ha löst sina problem med att rekrytera produktiv personal och har lyckats att i allt större utsträckning organisera forskningen i produktiva former.

Den iakttagna förändringen från en nivå till en annan mellan de båda perioderna 2008-2011 och 2012-2015 antyder att högskolornas forskare borde likställas med universitetens och betraktas som likvärdiga såväl i egenskap av sökande till som anslagsmottagare från forskningsråden.

Referenser:

- U Sandström & van den Besselaar Quantity and/or Quality? The Importance of Publishing Many Papers. *PLoS ONE*(2015).
- U Sandström & E Sandström: *Resurser för citeringar* (2009), Höskoleverket Rapport 2008:18R
- U Sandström & E Sandström: Bibliometrisk rapport för Linköpings universitet (2010).
- U Sandström, A Wold, m fl :*Hans Excellens: om miljardsatsningarna på starka forskningsmiljöer*. (2010).
- U Sandström & A Wold: *Excellenssatsningarna – belöning för kön eller för toppforskning?* Årsbok RJ (2015).
- U Sandström: *Är forskning med svagt genomslag koncentrerad till vissa lärosäten?* Rapport till KK-stiftelsen 16 mars 2015.

Docent Ulf Sandström har utfört detta uppdrag från KK-stiftelsen inom ramen för konsultföretaget Forskningspolitik Sverige AB. Sandström är knuten till KTH och Örebro universitet som forskare i *Science & Technology Studies* med inriktning på forskningspolitik och publiceringsanalys.

Bilaga. Fördelning över decentiler per lärosäte (2012-2015)

Högskola	1	2	3	4	5	6	7	8	9	10	Total
BTH	14	14	17	17	24	12	27	16	17	4	162
DU	5	14	20	11	16	11	8	12	8	5	110
HB	19	21	26	12	27	8	15	25	16	1	170
HH	7	8	8	12	14	13	12	17	7	1	99
HIG	11	29	20	24	16	13	13	12	15		153
HIS	1	16	15	13	18	19	19	9	10	4	124
HJ	26	32	32	33	16	29	27	16	14	6	231
HKR	7	12	9	9	5		11	6	4	6	69
HV	13	14	14	9	8	8	9	13	6		94
MAH	40	40	42	35	32	31	43	33	34	12	342
MDH	14	24	25	25	15	29	21	25	21	1	200
SH	10	25	33	14	14	11	4	10	5	2	128
Nya univ	1	2	3	4	5	6	7	8	9	10	Total
KAU	38	49	43	30	31	32	44	7	14	8	296
LNU	47	48	61	57	53	40	32	26	18	23	405
MIUN	27	43	55	52	28	32	34	29	32	10	342
ORU	60	70	69	70	73	70	58	63	85	49	667
Spec	1	2	3	4	5	6	7	8	9	10	Total
KI	784	663	698	776	854	915	844	981	1171	1370	9056
CTH	296	321	314	233	233	218	204	195	112	78	2204
HHS	29	32	27	9	16	6	8	5	1	2	135
KTH	432	417	378	398	352	304	317	239	166	172	3175
LTU	82	86	97	95	90	87	73	71	41	11	733
Univ	1	2	3	4	5	6	7	8	9	10	Total
GU	486	477	470	476	526	478	468	481	529	509	4900
LIU	332	294	297	292	281	272	268	252	255	243	2786
LU	785	699	716	700	663	656	606	722	668	700	6915
SLU	293	288	229	277	228	234	191	184	150	87	2161
SU	432	427	398	358	304	245	280	250	196	264	3154
UMU	256	286	295	278	287	260	265	246	265	217	2655
UU	681	643	586	567	593	567	556	576	616	655	6040
Total	5232	5102	5009	4892	4823	4613	4465	4525	4479	4444	47584

Källa: Web of Science