Bilaga 5.
Teoretiskt ramverk

RiR 2019:21

Resurseffektivitet och produktivitet vid Sveriges lärosäten i nordisk jämförelse

Bilaga 5 Teoretiskt ramverk

Utgångspunkten för att studera teknisk effektivitet är reala resurser och real produktion. De vanligaste utgångspunkterna för att studera teknisk effektivitet inom nationalekonomisk produktionsteori är resursminimering alternativt produktionsmaximering. Resursminimering utgår från att produktionsenheterna, i detta fall lärosäten, minimerar resursåtgången (input) för att producera ett på förhand fastställt produktionsmål (output). Detta sätt benämns inputorienterad teknisk effektivitet. Det andra sättet att analysera effektivitet benämns outputorienterad teknisk effektivitet och har sin utgångspunkt i produktionsmaximering. I det fallet antas att ett effektivt lärosäte producerar en maximal kvantitet givet den mängd resurser som är tillgängliga. I denna granskning har Riksrevisionen, i likhet med huvuddelen av den internationella litteraturen som studerat effektivitet för lärosäten, valt en outputorienterad modell.

Beräkning av teknisk effektivitet

Teknisk effektivitet kan studeras från både input- och outputsidan beroende på vilka övergripande mål som finns med produktionen. En produktionsenhet, i detta fall ett lärosäte, kommer att vara tekniskt effektiv om den inte kan producera mer tjänster (output) utan att resurser tillförs. Alternativt om det sker en reduktion av resurser (input) så klarar ett effektivt lärosäte inte längre att producera sin tidigare volym.

För att formellt definiera effektivitet och modell behövs ett antal beteckningar. Låt K vara antalet lärosäten. Dessa lärosäten använder N olika typer av resurser (input) för att producera M olika dimensioner av output. $\boldsymbol{x}_j$ är en vektor av dessa N inputs för lärosäte j och $\boldsymbol{y}_j$ är en vektor av output för samma lärosäte.

Produktionsteknologin definieras då som

$$T = \{\mathbf{y}_j \text{ kan produceras av } \boldsymbol{x}_j\} \tag{1}$$

Produktionsfronten kan sedan definieras som

$$TE = \{(\boldsymbol{x}_j, \boldsymbol{y}_j) \in T, (\boldsymbol{x}_j, \lambda \boldsymbol{y}_j) \notin T \text{ för } \lambda > 1\} \tag{2}$$

I ekvation (2) är λ ett tal och tolkningen av den tekniska effektiviteten om produktionen ska kunna öka ($\lambda > 1$) blir att det inte är möjligt ($\notin$) om inte resurserna ökar. Således kommer ett lärosäte att vara ineffektivt om det går att finna ett $\lambda > 1$ sådan att $(\boldsymbol{x}_j, \lambda \boldsymbol{y}_j) \in T$. Om så är fallet är λ, graden av effektivitet. Om exempelvis $\lambda = 1{,}10$ innebär det att det aktuella lärosätet (j) skulle kunna öka sin produktion med 1,10, det vill säga lärosätet är 10 procent ineffektivt.

Det första steget när DEA-metoden används är att identifiera de lärosäten som är effektiva och som tillsammans utgör produktionsmöjlighetsfronten. För att konstruera produktionsmöjlighetsfronten binds effektiva lärosäten samman av konvexa kombinationer. Detta innebär att om två effektiva lärosäten med en produktion på 10 respektive 20 enheter kan observeras så går det alltid att konstruera ett hypotetiskt effektivt lärosäte som producerar 15 enheter. Detta görs genom att kombinera de två förstnämnda. Eftersom DEA använder en linjärprogrammeringsansats för att konstruera produktionsfronten benämns teknologin ibland som styckvis linjär.

Linjärprogrammeringsproblemet som ska lösas är följande:

$$\max \lambda_j \tag{3}$$

under bivillkoren

$$\sum_{k=1}^{K} z_k x_{k,n} \leq x_{j,n}, n = 1,..,N \tag{4}$$

$$\sum_{k=1}^{K} z_k y_{k,m} \geq \lambda_j y_{j,m}\,, m = 1, \ldots, M \tag{5}$$

$$z_k \geq 0 \quad (CRS) \tag{6}$$

Ekvation (3) definierar effektivitetstalet som ska beräknas. Vad som söks är det minsta avståndet från fronten. Ekvation (4) definierar inputvariablerna. z-variablerna är de vikter som används för att konstruera de konvexa kombinationerna av effektiva lärosäten och benämns ibland intensitetsvariabel. Om ett lärosäte är ineffektivt kommer den inte att användas i konstruktionen av produktionsfronten varför värdet på z kommer att vara 0 för dessa enheter.[1] På högersidan i ekvation (5) återfinns λ -effektivitetstalet. Om det inte är möjligt att öka produktionen kommer denna att vara 1 och i annat fall ett tal över 1. Slutligen anges i ekvation (6) att intensitetsvariablerna ska vara positiva.[2]

Återsampling en metod för att öka resultatens tillförlitlighet

DEA-metoden är deterministisk[3] och i sin enklaste form underskattar den produktionsfrontens ”sanna” placering.[4] Det beror på att produktionsfronten baseras på beräkningar för de enheter som för varje år kategoriseras som effektiva. Därmed

1 z-variablerna motsvarar de aktivitetsparametrar eller intensitetsvariabler som introducerades av von Neuman (1938).

2 I granskningen antas konstant skalavkastning (CRS). Genom att variera restriktionen i ekvation (4) kan olika typer av skalavkastning modelleras.

3 Deterministisk innebär att det inte finns något slumpmoment i metoden.

4 Badunenko m.fl. (2008).

beräknas egentligen en fiktiv nedre gräns för den faktiska produktionsfronten. Det medför att måttet på effektivitet blir större än den verkliga effektiviteten, vilket innebär att DEA-resultatet underskattar ineffektiviteten. Måttet på ineffektivitet är avståndet mellan de ineffektiva enheter som inte ligger på fronten och de effektiva enheter som utgör fronten. För att hantera problemet har tekniker utvecklats för att justera den skattade produktionsfronten, så kallad återsampling exempelvis genom bootstrappingmetoden, vilken är den metod som har utnyttjats i denna granskning.[5] Kortfattat innebär det att ett urval av enheter dras från de data för vilka effektivitetsberäkningar görs. Dessa beräkningar återupprepas sedan 2 000 gånger. Genom att upprepa processen skapas en produktionsfront som är mindre skev och därmed mer sann utifrån de resurser som tilldelats och de prestationer som produceras. En ytterligare fördel med bootstrapping är att metoden gör det möjligt att skapa konfidensintervall kring varje enhets beräknade effektivitet, vilket möjliggör inferensanalys. Det är alltså möjligt att testa om ineffektiviteten för enskilda lärosäten är signifikant skild från noll och det är därmed också möjligt att skilja effektiva lärosäten från ineffektiva.[6]

Beräkning av produktivitet

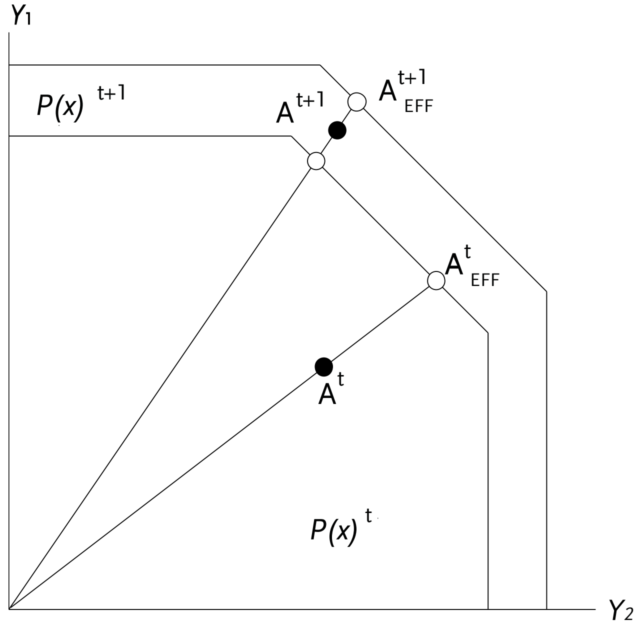
Den ansats för beräkning av produktivitet som används i granskningen bygger på en kombination av effektivitetstal som mäter årliga förändringar. Detta sätt att beräkna produktivitet brukar benämnas Malmquists produktivitetsindex.[7] I figur B1 nedan illustreras produktivitetsutvecklingen med samma fiktiva lärosäte A som fanns med i exemplet i avsnitt 2.1. $P(x)^t$ betecknar produktionsmöjlighetsområdet för år 1 och $P(x)^{t+1}$ betecknar motsvarande område för år 2. Mellan år 1 och år 2 har teknologin förändrats så att produktionsmöjlighetsfronten har förskjutits utåt. Det innebär att år 2 kan lärosätena producera mer output med samma mängd resurser.

5 Det finns även andra återsamplingsmetoder som kan användas, exempelvis *jack knife*. För ytterligare information se exempelvis Efron (1979).

6 Se exempelvis Simar och Wilson (1998, 1999, 2000, 2008).

7 Detta index utvecklades av Sten Malmquist (Malmquist, 1953) och var från början en tillämpning på priser. Anpassningen för produktivitetsmätningar presenterades av Caves m.fl. (1982) och den första tillämpningen gjordes i Färe m.fl. (1989, 1992).

Figur B1 Illustration av Malmquists produktivitetsindex



Som figur B1 visar, ligger lärosäte A:s produktion (A^t och A^{t+1}) innanför produktionsmöjlighetsfronten båda åren. Graden av effektivitet år 1 anges av kvoten $(0 – A^t)/(0 – A^t_{EFF})$. Nästföljande år (*t*+1) producerar lärosäte A i punkten A^{t+1} och graden av effektivitet för år 2 ges av kvoten $(0 – A^{t+1})/(0 – A^{t+1}_{EFF})$. Lärosäte A är visserligen fortfarande ineffektivt, men eftersom lärosäte A år 2 befinner sig närmare produktionsmöjlighetsfronten är ineffektiviteten inte lika stor.[8]

Utvecklingen över tid påverkas samtidigt av pågående förändringar i teknologi och organisation, vilket leder till att det kan bli möjligt att producera mer med tillgängliga resurser.

I beräkningarna formuleras fyra linjära programmeringsproblem för varje lärosäte, som resulterar i fyra olika effektivitetstal. Två av dessa bygger på tvärsnittsberäkningar (en för respektive tidsperiod) där det enskilda lärosätet jämförs med övriga lärosäten. De två andra är korskopplade beräkningar så att lärosätets produktion under period 1 jämförs med övriga lärosätens produktion under period 2 och så att lärosätets produktion under period 2 jämförs med övriga lärosätens produktion under period 1.

8 Mer precist kan Malmquists produktivitetsindex definieras som effektivitetsförändring • teknologisk förändring. Se exempelvis Färe m.fl. (1992).

Malmquists produktivitetsindex beräknas med hjälp av de fyra effektivitetstalen som ingående komponenter. Indexet visar den sammanlagda förändringen i produktivitet mellan perioderna. Dessutom kan indexet delas upp i två olika komponenter.

En del av förändringen i produktivitet beror på att produktionsmöjlighetsfronten har förändrats. En teknologisk utveckling har alltså skett. Förutsatt att förändringen är positiv så är det möjligt att producera mer prestationer med oförändrad mängd resurser. Fronten har därmed skiftats utåt så att produktionsmöjlighetsområdet blivit större.

En andra del av förändringen i produktivitet beror på att lärosätets avstånd till produktionsmöjlighetsfronten har förändrats. Om denna förändring är positiv, innebär det att lärosätet blivit mer effektivt i sitt resursutnyttjande mellan perioderna. För att definiera produktivitet måste vi ta hänsyn till både period *t* och period *t*+1. Detta går att göra genom att använda distansfunktioner som introducerades av Shephard (1953) respektive Shephard (1970). Distansfunktionen och den tekniska effektiviteten är direkt relaterade enligt följande:

$$\frac{1}{D_o(x,y)} = \lambda \tag{7}$$

där $D_o(x,y)$ i ekvation (7) är distansfunktionen från outputsidan. Malmquists produktivitetsindex kan med hjälp av dessa distansfunktioner definieras och delas upp enligt ovan enligt följande:

$$M_0(x^{t+1}, y^{t+1}, x^t, y^t) = \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \times \left(\frac{D_0^t(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D_o^t(x^t, y^t)}{D_o^{t+1}(x^t, y^t)} \right)^{\frac{1}{2}} \tag{8}$$

Kvoterna under rottecknet i ekvation (8) avser den första delkomponenten enligt ovan. De mäter avståndet mellan produktionsmöjlighetsfronterna och lärosätets produktion för respektive period. Genom att multiplicera dessa kvoter och dra kvadratroten av detta erhålls ett genomsnittligt avstånd mellan de två periodernas produktionsmöjlighetsfronter.

Kvoten utanför rottecknet avser den andra komponenten. Den mäter förändringen i grad av effektivitet, det vill säga ett lärosätes förändrade avstånd till fronten. Ett värde större än 1 innebär att avståndet minskat och att lärosätet därmed blivit mer effektivt i sin resursanvändning.