

Vem blir långtidsarbetslös?

av

Helge Benmmarker, Kenneth Carling och Anders Forslund

2006-12-20

Sammanfattning

I ett antal länder har man på senare tid utvecklat statistiska modeller som ett stöd för förmedlarna att förutsäga vilka arbetslösa som riskerar att bli långtidsarbetslösa, s k profilning. I den här rapporten redovisar vi resultaten av att skatta hasardmodeller för att förutsäga vem som riskerar att bli långtidsarbetslös bland dem som registrerar sig som arbetslösa arbetssökande vid arbetsförmedlingen. Av resultaten framgår det att den skattade modellen lyckas relativt väl med att förutsäga vem som riskerar att bli arbetslös i minst sex månader – träffsäkerheten i prognoser som görs utanför modellens skattningsperiod är nästan 70 %.

Innehållsförteckning

1.	Inledning.....	3
2.	Modellen.....	5
2.1	Deltagande i arbetsmarknadspolitiska program.....	7
2.2	Arbetslöshetsperiod	7
3.	Data.....	8
4.	Resultat	10
5.	Referenser	14

1. Inledning

En arbetsförmedlare vid den offentliga arbetsförmedlingen har ofta ansvaret för hundratal arbetslösa arbetssökande. Många av dessa arbetssökande kan förväntas hitta ett jobb relativt snabbt och utan alltför omfattande insatser från förmedlaren. Vissa arbetssökande riskerar emellertid att bli långvarigt arbetslösa. Ett sätt att motverka långtidsarbetslöshet är att erbjuda de arbetssökande omfattande stöd redan tidigt under en arbetslöshetsperiod. Den stora mängden arbetssökande per förmedlare innebär emellertid att alla arbetssökande inte kan få sådan hjälp. Det skulle också bli mycket kostsamt att förstärka förmedlingarnas resurser så att alla arbetssökande kunde ges ett tidigt och omfattande stöd. Detta problem kan lösas på olika sätt. Ett sätt är att helt vänta med insatser under en inledande del av en arbetslöshetsperiod. En del av de arbetssökande kommer då att hitta jobb utan förmedlingens hjälp och antalet potentiella långtidsarbetslösa reduceras. Väntar man tillräckligt länge med insatserna kommer uppenbarligen alla kvarvarande arbetssökande att vara långtidsarbetslösa, men då har ingen av dessa kunnat få tidig hjälp. Görs perioden istället kort, så kommer antalet kvarvarande arbetssökande fortfarande att vara stort och inte så mycket är vunnet. Ett annat sätt att lösa problemet är att överlåta beslutet om vilka som ska få tidiga insatser till förmedlingen/förmedlaren. Denna strategi förutsätter att förmedlaren kan göra en god prognos om vem som riskerar att bli långtidsarbetslös så att förmedlingens resurser faktiskt kommer att användas för att ge stöd åt dem som har störst nytta och behov av det. Redan idag har förmedlaren tillgång till den information som den arbetssökande matat in i förmedlingens administrativa system vid registreringen hos arbetsförmedlingen samt, naturligtvis, den information som mötet med den arbetslöse ger upphov till.

I ett antal länder har man på senare tid utvecklat statistiska modeller som ett stöd för förmedlarna att förutsäga vilka arbetslösa som riskerar att bli långtidsarbetslösa, s k profilering.¹ Profileringmodeller används alltså för att identifiera vem som riskerar långtidsarbetslöshet. En relaterad, men ambitiösare uppgift är att dessutom använda statistiska modeller för att bedöma vilka som ska få ta del av olika arbetsmarknadspolitiska insatser och när de i så fall ska sättas in, s k targeting. I den här rapporten skattar vi profileringmodeller för dem som söker jobb vid arbetsförmedlingen. Den mycket mer ambitiösa uppgiften att använda statistiska metoder för att placera rätt person i rätt program vid rätt tidpunkt ligger långt bortom horisonten för föreliggande rapporten.

Den skattade modellen har flera möjliga användningar. Den mest uppenbara användningen är att modellens prognoser kan vara ett stöd för förmedlarna när

¹Se exempelvis Rosholm m fl (2004), Frölich m fl (2003) eller Behnke m fl (2006).

de gör bedömningar av vilka arbetssökande som riskerar att bli långtidsarbetslösa. Förutsatt att modellens prediktioner är rimligt precisa kan man även tänka sig andra användningar. För det första kan modellen användas för att fördela resurser mellan arbetsförmedlingarna. Tanken är rättfram: modellen förutsäger hur lång tid inflödet av arbetslösa vid olika förmedlingar kan förväntas vara arbetslösa vid en oförändrad politik (den politik som använts under den period som modellen skattats på). Resurserna kan då styras över till förmedlingar som möter arbetssökande som kan förväntas ha större problem att hitta ett arbete från förmedlingar med en lättare situation. Man kan också tänka sig att utforma bonussystem för förmedlingarna där man belönar insatser som leder till arbetstider som är kortare än de enligt modellen förväntade. För det andra kan modellen användas för att precisera innebörden i den ”likabehandling” som de arbetssökande har rätt att förvänta sig från den offentliga arbetsförmedlingen – ”likar” kan definieras i termer av förväntade arbetslöshetstider.

Den modell vi skattar är en varaktighetsmodell för tid i arbetslöshet enligt Ams register.² I den här versionen är modellen skattad på ett 5-procentigt slumpmässigt urval av nyinflödet av (arbetslösa) arbetssökande vid arbetsförmedlingarna mellan oktober 2002 och oktober 2004. Vi skattar en modell för hela denna population istället för att skatta separata modeller med stratifiering efter (exempelvis) ålder, region, tillgång till a-kassa eller kön. Valet att skatta en modell för samtliga arbetssökande snarare än separata modeller för olika grupper baseras på flera överväganden. För det första finns det en risk att modellen överparametriseras om man väljer att skatta olika modeller för ett stort antal grupper. En möjlig konsekvens skulle vara att goda egenskaper i det sample som modellen skattas på köps till priset av instabilitet som skulle kunna manifesteras i dåliga prognosegenskaper, särskilt om betingelserna i prognosperioden avviker från betingelserna i skattningsperioden (exempelvis beroende på politikomläggningar). För det andra kan modellen inte användas som en norm för likabehandling eller för resursfördelning om den skattas separat för ett stort antal regioner.

I skattningarna utnyttjar vi ingen information som på ett uppenbart sätt grundar sig på en bedömning som gjorts av en förmedlare – den skattade modellen ska ju vara ett underlag för förmedlarens bedömningar.

Den skattade modellen kan, som vi redan påpekat, användas på olika sätt. Hittills har vi utnyttjat skattningarna för att prognostisera om en arbetssökande kan förväntas vara arbetslös sex månader efter det första mötet med en arbetsförmedlare (som i sin tur har antagits inträffa efter fyra veckors arbetslöshet).

Vi har huvudsakligen utvärderat modellen genom att prognostisera sannolikheten att bli långtidsarbetslös i en population (prognospopulationen) utanför

²Vi utnyttjar informationen i Ams Händel-databas. Vi definierar senare exakt vad vi menar med tid i arbetslöshet och vad som innebär att man lämnar arbetslöshet.

det urval modellen skattats för (skattningspopulationen). Ett sätt att beskriva modellens prognosegenskaper är följande: i prognospopulationen kommer ungefär varannan nyinskriven arbetslös att bli långtidsarbetslös. En enkel och naiv prognosmetod som enbart utnyttjar detta faktum skulle vara att singla slant för var och en av individerna; krona blir långtidsarbetslös, klave blir det inte. Denna prognosmetod skulle ge en korrekt prediktion i ungefär 50 % av fallen. Vår skattade modell ger väsentligt bättre prognoser, vi kan prognostisera ungefär 70 % av fallen rätt, en 40-procentig förbättring av den naiva slantsinglingsprognosen.

2. Modellen

Våra skattningar syftar till att förutsäga vem som riskerar att bli långtidsarbetslös. Ett rättframt sätt att skatta detta skulle vara att använda en diskret valmodell (exempelvis logit eller probit) där den beroende variabeln är om individen är eller inte är arbetslös vid en viss tid efter nyregistreringen som arbetsökande arbetslös. Nackdelen med denna ansats är att inte all variation i den tid det tar att komma till ett jobb för de arbetslösa utnyttjas i skattningarna. Vi har därför valt att istället skatta en varaktighetsmodell för tid utan arbete. Den skattade varaktighetsmodellen kan sedan användas för att göra prognoser för att beräkna sannolikheten att en individ kommer att vara arbetslös vid en given, framtida tidpunkt.

Praktiskt går vi till väga så att vi skattar hasardfunktioner och använder dessa för att prognostisera tid i arbetslöshet. Låt T , $T \in (0, \infty)$ beteckna arbetslöshetens varaktighet (tid till arbete). Då ges hasarden, sannolikheten att en individ med egenskaperna X som är arbetslös vid tidpunkten T hittar ett arbete i tidsintervallet $t + dt$, av

$$h(t | x_t) = \lim_{dt \rightarrow 0} \frac{P(t < T \leq t + dt | T > t, x_t)}{dt} = \frac{f(t | \{x_s\}_0^t)}{S(t | \{x_s\}_0^t)} \quad (1)$$

där $f(\cdot)$ är täthetsfunktionen, $S(\cdot)$ överlevnadsfunktionen och $\{x_s\}_0^t$ är utvecklingen över tiden för alla kovariater fram till tidpunkten t . En överlevnadsfunktion visar sannolikheten att inte ha fått ett arbete vid en viss tidpunkt. När

man har skattat en hasardfunktion ($h(t)$) kan överlevnadsfunktionen beräknas som³

$$S(t) = \exp\left(\int_0^t h(s)ds\right) \quad (2)$$

Om vi vill göra en prognos om sannolikheten att en arbetssökande kommer att vara arbetslös 6 månader (26 veckor) efter sitt möte med en arbetsförmedlare (som vi antar sker efter τ veckor) kan vi därför finna denna sannolikhet som

$$\Pr(T > \tau + 26 | T > \tau) = \exp\left(\int_{\tau}^{\tau+26} h(s)ds\right) \quad (3)$$

Vi specificerar hasardfunktionen som en proportionell hasardmodell, så att hasarden är produkten av, för det första, en baselinehasard som fångar tidsberoende och, för det andra, en funktion som fångar beroendet av konstanta och tidsvarierande faktorer (egenskaper) för individerna, x_t

$$h(t | x_t) = \lambda(t) \cdot \varphi(x_t) \quad (4)$$

där $\lambda(t)$ är baselinehasarden och $\varphi(x_t)$, som vi specificerar som $\exp(x_t \beta)$, ”skalar om” baselinehasarden beroende på de individuella egenskaperna (x_t). Baselinehasarden specificeras som en trappstegsfunktion som antas vara konstant inom de tidsintervall vi använder och skiftar mellan dessa intervall. Under de tre första månaderna är intervallen veckor, därefter månader. Vi får alltså $\lambda(\tau) = \lambda_k$ för $\tau = k$, $\tau = 1, \dots$

Med de valda funktionsformerna kan vi skatta sannolikheten att en individ kommer att vara arbetslös sex månader efter sitt möte med en arbetsförmedlare som

$$\hat{\Pr}(T > \tau + 26 | T > \tau) = \exp\left(-\exp(x_{\tau} \hat{\beta}) \sum_{k=\tau}^{\tau+26} \hat{\lambda}_k\right) \quad (5)$$

³ Vi utelämnar beroendet av kovariaternas tidsbanor (x_t) för att göra notationen enklare.

om kovariaterna x hålls konstanta; eftersom modellen ska användas för att förutsäga om en individ riskerar att bli långtidsarbetslös kan vi inte använda tidsvarierande kovariater i prediktionerna. Det naturliga valet av värden för x är de värden variablerna antar vid tidpunkten för mötet med arbetsförmedlaren, x_t .

2.1 Deltagande i arbetsmarknadspolitiska program

En del arbets sökande kommer att gå in i arbetsmarknadspolitiska program innan de har varit arbetslösa i sex månader. Programdeltagandet påverkar potentiellt sannolikheten att gå till ett arbete både före, under och efter programdeltagandet.⁴ Det är inte självklart hur detta ska hanteras i vår analys. Ett möjligt tillvägagångssätt skulle vara att censurera alla arbetslöshetsperioder när de övergår till deltagande i något arbetsmarknadspolitiskt program.⁵ Våra prognoser skulle då i praktiken avse dem som inte går in i program tidigt under en arbetslöshetsperiod. Detta skulle uppenbarligen skapa särskilda problem när vi vill göra prognoser för unga arbetslösa, som ju i genomsnitt går in i program tidigt i sina arbetslöshetsperioder (Forsslund & Nordström Skans, 2006). Vi har därför valt att hantera problemet genom att i skattningarna inkludera programdeltagandet som tidsvarierande kovariater.⁶ Det vi uppnår med detta är att vi ”rensar” de skattade tiderna till arbete för eventuella effekter av programdeltagandet.⁷ När vi prognostiserar arbetslöshetstider sätter vi naturligtvis allt programdeltagande lika med noll.

2.2 Arbetslöshetsperiod

Det är inte självklart vad man ska mena med ”långtidsarbetslöshet”. För det första måste man bestämma sig för hur länge en arbetssökande ska ha varit arbetslös innan arbetslösheten övergår i långtidsarbetslöshet. Vi har hittills valt att definiera långtidsarbetslöshet som sex månader från det första mötet mellan

⁴ Det finns evidens för att programdeltagare under pågående program ägnar mindre resurser åt att hitta ett arbete så att flödet till jobb minskar (se exempelvis översikten i Calmfors, Forsslund & Hemström, 2002); man talar ofta om en inläsningseffekt. Programmen kan också ha en behandlingseffekt så att flödet till arbete påverkas (positivt eller negativt). Om individen förväntar sig att i en framtida tidpunkt gå in i ett program kan detta också påverka flödet till arbete redan innan programplaceringen faktiskt ägt rum (Hägglund, 2006; Forsslund & Nordström Skans, 2006).

⁵ Detta innebär att informationen från dessa perioder används för skattningarna så länge som perioderna inte innebär att man går in i ett program.

⁶ Vi grupperar programmen i tre större kategorier.

⁷ Vi är ju intresserade av att prognostisera den förväntade arbetslöshetstiden i frånvaro av insatser för att kunna styra insatserna till dem som kan förväntas bli långtidsarbetslösa.

en arbetslös och en arbetsförmedlare. För det andra är det inte självklart vad som ska anses bryta en arbetslöshetsperiod. Vi har valt att i första hand använda en snäv definition av vad som bryter arbetslöshet, nämligen reguljärt arbete⁸. Vi utesluter därmed både övergångar till subventionerade anställningar och övergångar till reguljär utbildning. Vårt huvudargument för detta är att förmedlingarnas huvuduppgift är att förmedla arbeten, inte andra typer av utfall. Ett annat skäl för den snäva definitionen är att personer som lämnar arbetslösheten för studier sannolikt skiljer sig från dem som går till arbete såväl i observerade som i icke-observerade avseenden. Skillnaderna kan vara så stora att de skulle motivera att vi skulle skatta separata modeller för utflöden till studier och arbete.

3. Data

De data vi använder i skattningarna kommer från Ams Händelsedatabas (Händel). Detta innebär att data är tillgängliga praktiskt taget i realtid vid förmedlingskontoren så att indata för prognoserna är tillgängliga vid behov (dvs. när förmedlaren förbereder det första mötet med en arbetslös). Dessutom är de data som används i skattningarna antingen icke tidsvarierande eller daterade en period tidigare (exempelvis är lokal arbetslöshet och lokala lediga platser daterade en månad bakåt i tiden). Även detta bidrar till att den skattade modellen på ett rättfram sätt kan användas för prognoser ”i realtid”.

I ett senare skede kommer vi, som ett test av hur robust den skattade modellen är, att använda information från andra register för att se hur mycket denna extra information tillför modellens prognosförmåga.

Skattningarna baseras på nyinskrivna vid arbetsförmedlingen, vilka heller inte varit inskrivna under en föregående tremånaders period, fr o m vecka 40 2002 t.o.m. vecka 44 2004. Som beskrivits ovan bortser vi dock från inskrivningsperioder som är kortare än fyra veckor. Detta innebär att vi i skattningarna de facto använder oss av information om övergångar från arbetslöshet till arbete från vecka 44 år 2002 till vecka 44 år 2004. För att utvärdera hur bra den skattade modellen kan förutse långtidsarbetslöshet tillämpas den sedan på inflödet av arbetslösa under en annan tidsperiod, det som kallas ”out-of-sample prediktioner”. Denna utvärderingsperiod sträcker sig fr o m vecka 45 år 2004

⁸ Reguljärt arbete definieras här av avaktualiseringsorsakerna 1-3, samt sökandekategorierna 21, 22, 31 och 41. Den innefattar således deltidsarbete, timanställningar och tillfälliga arbeten. Vidare krävs att den arbetslösa kan hänföras till någon av dessa kategorier under i minst 10 dagar.

t o m vecka 43 år 2005. Eftersom vi predicerar sannolikheten att fortfarande vara arbetslös efter sex månader, så kommer dock bara individer som skrivs in minst sex månader innan vecka 43 2005 att vara användbara i utvärderingen. Sammanfattningsvis skattas alltså modellen för inflödet under två år och utvärderas under det därpå följande året, medelst inflödet under den första halvan av detta år.

För att skatta den statistiska modellen med så god precision som möjligt har vissa begränsningar gjorts i data. En sådan avser hur lång en inskrivningsperiod maximalt kan vara, d v s hur länge vi observerar individer. Om individen inte har övergått i reguljärt arbete efter 57 veckor så stängs observationsfönstret, observationen censureras⁹. Detta innebär att våra skattningar av vilka faktorer som bestämmer arbetslöshetstiderna begränsas till de första 57 veckorna av varje inskrivningsperiod. Vår definition av långtidsarbetslöshet –sex månader– ligger alltså väl inom detta intervall.

De karaktäristika som vi har använt i våra skattningar har uteslutande hämtats från databasen Händel¹⁰. Dessa karaktäristika kan vara av olika typer. Det mest uppenbara är uppgifter om individens karaktäristika, detta inkluderar bl a uppgifter som ålder, yrke, utbildning och invandrarbakgrund. En annan kategori är individens arbetslöshetshistoria. Båda dessa kategorier är också i princip oföränderliga under en given inskrivningsperiod.

Våra modeller inkluderar dock även uppgifter som förändras över tiden. Målet är ju, som klargjorts ovan, att skatta varaktighetsmodeller för hur lång tid det tar innan individer får arbete. Rent praktiskt innebär detta att vi organiserar uppgifterna om individerna samt övriga betingelser veckovis. Modellerna skattas sedan på veckodata och räknas slutligen om till arbetslöshetstider. På detta sätt kan vi alltså också ta hänsyn till omständigheter som vecka för vecka förändras under en inskrivningsperiod. En ytterligare grupp av karaktäristika avser därför uppgifter om hur den för individen relevanta arbetsmarknaden ser ut och förändras. I denna kategori ingår aggregerade uppgifter om hur stor arbetslösheten är lokalt, samt hur den förändras månad för månad. Även information om

⁹ Observationer censureras också av alla de avaktualiseringsorsaker som inte definieras som reguljärt arbete, exempelvis om individen gått in i utbildning (dock ej AMU).

¹⁰ Tabellerna Inesper, Sokatper och Soyrc i Händel har nyttjats. Däremot har inte information från individernas handlingsplaner använts. Vidare har inte heller information från HFSSO använts. Detta innebär att vi inte tagit hänsyn till det faktiskt kan ske förändringar av individens karaktäristika under en inskrivningsperiod, exempelvis avseende medborgarskap. Om ett profilingsystem skulle komma att introduceras skulle man troligen vilja nyttja även denna typ av information i skattningarna.

var individen bor inkluderas här. Den näst sista gruppen av uppgifter avser tiden. Modellerna innehåller kontroller för när under året individen är arbetslös. Detta inkluderar både tidpunkt för inskrivning och uppgifter som hur säsongerna skiftar under inskrivningsperioden. Slutligen innefattar modellen uppgifter om varaktighetsberoende och programdeltagande, d v s om hur arbetslöshetens naturalförlopp ser ut – hur sannolikheten att övergå i arbete förändras över inskrivningsperioden – samt huruvida individen deltar i arbetsmarknadspolitiska program eller inte. Effekten av att gå in i program används självklart inte i våra prognoser. Informationen används till emellertid till att renodla effekten av övriga karaktäristika.

Utöver de ovan beskrivna uppgifterna har vi laborerat med hur dessa olika uppgifter samverkar, interagerar. Vi har således även inkluderat ett mindre antal interaktionseffekter i modellen.

Av tekniska skäl har skattningarna gjorts på ett slumpmässigt urval av individer. I aktuell version av Händel fanns uppgifter om 3 676 257 individer. Ett 5%-igt urval gav 183 812 individer. Av dessa hade 56 522 individer minst en inskrivningsperiod under skattnings- eller utvärderingsperiod. Antalet inskrivningsperioder uppgick till 73 270. Av dessa var dock 18 012 inskrivningsperioder kortare än fyra veckor. Av de återstående inskrivningsperioderna inföll 36 861 under skattningsperioden och 18 397 under utvärderingsperioden.

4. Resultat

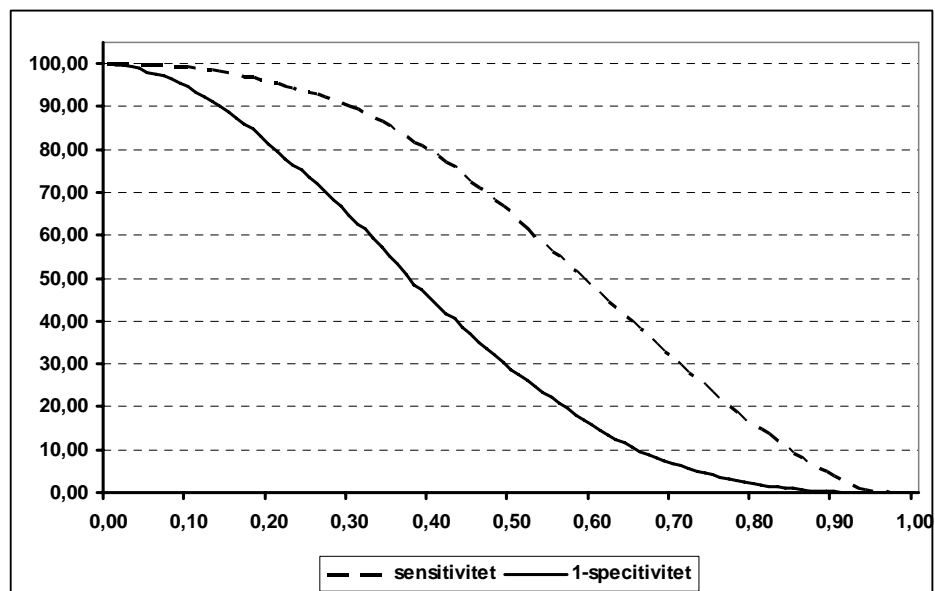
I ett profilingsystem kan man tänka sig att för varje individ presentera ett sannolikt utfall: *att bli långtidsarbetslös* eller *att inte bli långtidsarbetslös*. Detta är dock naturligtvis inte det enda sättet att presentera en prediktion för arbetsförmedlaren, ett annat sätt kan exempelvis vara att presentera individens förväntade arbetslöshetstid. I denna rapport väljer vi – av pedagogiska skäl – att uteslutande kategorisera prediktionerna på det ovan beskrivna sättet, det ger nämligen möjlighet att direkt jämföra med faktiskt utfall. Utifrån skattade modeller predicerar vi således individer som antingen långtidsarbetslösa, eller inte långtidsarbetslösa, varefter vi jämför detta med det faktiska utfallet i utvärderingsperioden (out of sample). Detta medför att vi predicerar fel av två olika skäl. Vi kan dels predicera en individ som *långtidsarbetslös* som de facto inte blir det, dels kan vi predicera en individ som *inte långtidsarbetslös* som sedan ändå visar sig bli det. Vi inför därför två begrepp:

- Sensitivitet=Sannolikheten att en individ prognostiseras som långtidsarbetslös givet att han/hon sedan också blir långtidsarbetslös.
- Specificitet= Sannolikheten att en individ *inte* prognostiseras som långtidsarbetslös givet att han/hon sedan inte heller blir långtidsarbetslös.

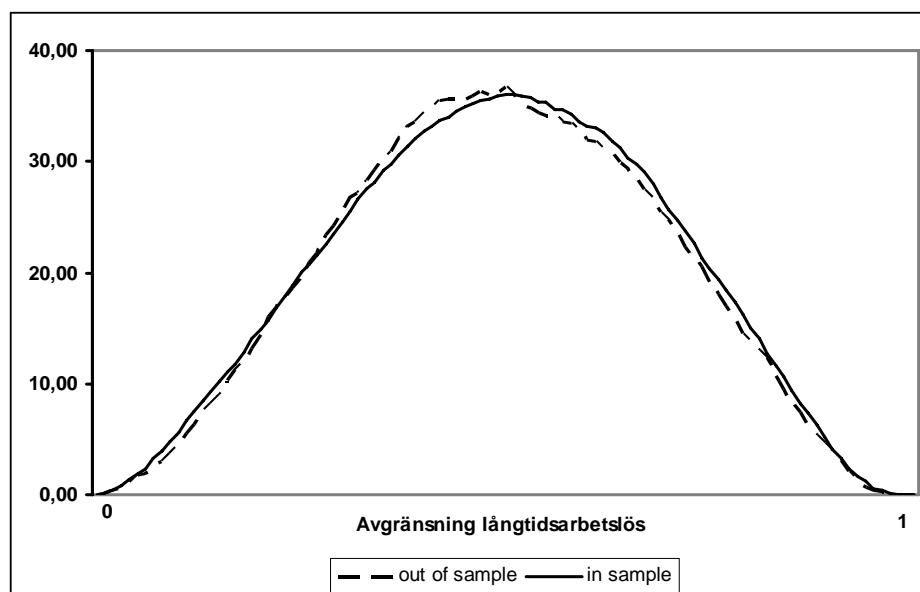
Dessa begrepp fungerar alltså som hjälpmedel för att beskriva hur väl vi lyckas med våra prediktioner. I Figur 1 presenteras resultatet av våra skattningar i termer av sensitivitet (som vi vill ska vara så hög som möjligt) och ”1-specificitet” (som vi vill ska vara så låg som möjligt). En hög sensitivitet i kombination med låg ”1-specificitet”, d v s att kurvorna är väl separerade, innebär att vi kan identifiera individer som kommer att bli långtidsarbetslösa, utan att samtidigt felaktigt prognostisera individer till långtidsarbetslöshet som sedan inte blir det.

X-axeln i tabellen nedan avser vilken gräns man väljer att sätta för klassificeringen av prediktionerna. Värdet 0.6, exempelvis, innebär att individer som enligt den skattade modellen har större sannolikhet än 60 % att bli långtidsarbetslös också prognosticeras som detta, medan individer med lägre sannolikhet ej prognosticeras som långtidsarbetslösa. Beroende på prioriteringar och annat kan man välja en gräns som gör att man prioriterar antingen sensitivitet (lågt värde) eller specificitet (høgt värde).

Som framgår av *Figur 1* lyckas modellen relativt väl prognostisera sannolikheten att bli långtidsarbetslös. I *Figur 2* sammanfattas informationen från *Figur 1* i den kurva som betecknas ”Out of sample”. Denna visar skillnaden mellan de två kurvorna i *Figur 1*. Eftersom kurvan når upp till drygt 35 % så är alltså skillnaden mellan sensitivitet och ”1-specificitet” maximalt ca 35%. *Figur 2* fungerar också som en utvärdering av hur övergången mellan in-sample och out-of-sample fungerar. De två kurvorna ligger mycket nära varandra, vilket innebär att prognoserna ”out of sample” är ungefärligen lika bra som ”förklaringsgraden” i skattningsperioden. En tolkning av detta är att de individ-, arbetsmarknads- och övriga karaktäristika som predicerar långtidsarbetslöshet i skattningsperioden (in-sample) också gör det i utvärderingsperioden (out-of-sample), samt att de också predicerar på ungefär samma sätt i utvärderingsperioden. Detta ger en bild av att ett profilingsystem baserat på en modell av detta slag även har förutsättningar för att – i detta avseende – kunna fungera i praktisk tillämpning.



Figur 1. Sensitivitet och specificitet vid out-of-sample prediktion.



Figur 2. Skillnad mellan sensitivitet och specificitet (%) som funktion av avgränsningen för predicerad långtidsarbetslöshet.

Totalt sett innebär de ovan presenterade resultaten att förklaringsgraden kan illustreras med följande räkneexempel. I skattningsperioden blir drygt 45 % långtidsarbetslösa. Om man bestämmer sig för prognostisera individer till långtidsarbetslöshet som har mer 50 % sannolikhet att bli just långtidsarbetslös, så uppnår man 68 % korrekta prediktioner i skattningsperioden (både sensitivitet och specificitet är ca 68 %). En något högre andel korrekta prediktioner hade för övrigt kunnat fås om vi valt en annan definition än 50 %. I utvärderingsperioden blir ca 48 % långtidsarbetslösa. Om den skattade modellen tillämpas i utvärderingsperioden uppnås också då 68 % korrekta prediktioner (sensitiviteten är då ca 65% och specificiteten ca 71 %). Denna förklaringsgrad är betydligt högre än den man skulle uppnå genom att singla slant – i vår period skulle då andelen korrekta prediktioner bli 48 %.

Vilka faktorer bidrar då mest till prognosen? I Tabell 1 har vi delat in de använda förklaringsvariablerna de grupper som vi diskuterade i Avsnitt 3 och redovisar hur stor del av den förklarade variationen som kan tillskrivas var och en av dessa grupper. De två grupper av variabler som förklarar särklassigt mest, knappt 40 procent vardera enligt denna analys, är individkaraktäristika och programdeltagande samt varaktighetsberoende. Individens historia, kalendertid (säsong) och förhållanden på den lokala arbetsmarknaden förklarar betydligt mindre, omkring 10 procent vardera.

Tabell 1. Modellens prediktionsförmåga uppdelad på några grupper av förklaringsvariabler.

Förklarande variabler, grupperade	Förklaringsgrad (andel)
Arbetsmarknad	0.07
Programdeltagande och varaktighetsberoende	0.36
Individens historia	0.10
Individens karaktäristika	0.37
Kalendertid	0.10

5. Sammanfattande diskussion

I den här rapporten har vi redovisat resultaten av att skatta hasardmodeller för att förutsäga vem som riskerar att bli långtidsarbetslös bland dem som registrerar sig som arbetslösa arbetssökande vid arbetsförmedlingen. Av resultaten framgår det att den skattade modellen lyckas relativt väl med att förutsäga vem

som riskerar att bli arbetslös i minst sex månader – träffsäkerheten i prognoser som görs utanför modellens skattningsperiod är nästan 70 %. Vi tycker att resultatet är så lovande att det är motiverat att på prov implementera ett profilingsystem som ett stöd för förmedlarnas arbete. Om detta görs, är det naturligtvis en fördel om införandet görs på ett sätt som underlättar en senare utvärdering. Ett sätt att uppnå detta skulle kunna vara att införa profilering på försök vid ett antal utvalda förmedlingskontor eller för ett antal förmedlare vid olika kontor.

6. Referenser

- Behnke, S, M Frölich & M Lechner (2006), Statistical assistance for programme selection – For a better targeting of active labour market policies in Switzerland, CESifo DICE Report 2/2006, s. 61-68.
- Calmfors, L, A Forslund & M Hemström (2002), Vad vet vi om den svenska arbetsmarknadspolitikens sysselsättningseffekter?, Rapport 2002:8, IFAU, Uppsala.
- Forslund, A & O Nordström Skans (2006), (Hur) hjälps ungdomar av arbetsmarknadspolitiska program för unga? Rapport 2006:5, IFAU Uppsala.
- Frölich, M, M Lechner & H Steiger (2003), Statistically assisted programme selection – International experience and potential benefits for Switzerland, *Swiss Journal of Economics and Statistics* 139, s. 311-31.
- Hägglund, P (2006), Are there pre-programme effects of Swedish active labour market policies? Evidence from three randomised experiments, Working Paper 2006:2, IFAU, Uppsala.
- Rosholm, M, M Svarer & B Hammer (2004), A Danish profiling system, Working Paper 2004-13, Department of Economics, University of Aarhus.