



Stockholms
universitet

Statistik analys av tidstrender i arbetslöshet

Yuqi Ma

Kandidatuppsats 2017:3
Matematisk statistik
Juni 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Statistisk analys av tidstrender i arbetslöshet

Yuqi Ma*

Juni 2017

Sammanfattning

Denna studie använder sig av arbetslöshetsdata under åren 2006-2015 i Sverige. Syftet med denna kandidatarbete är att undersöka trender inom arbetslöshet i olika utbildningsgrupper mellan åren 2006-2015 samt trender inom arbetslöshet för män och kvinnor i dessa utbildningsgrupper. Vi har kommit fram att arbetslösheten bland personer med förgymnasial utbildning har en ökande trend mellan åren 2006-2015, medan arbetslösheten för gymnasialt samt eftergymnasialt utbildade visar en annorlunda trend under dessa år. För att undersöka tidstrender inom arbetslöshet för män respektive kvinnor i olika utbildningsgrupper, plottar vi en B-spline för män och kvinnor i respektive utbildningsgrupp. Vi har kommit fram till att män och kvinnor i samtliga utbildningsgrupp har en kraftig ökande trend mellan åren 2008-2009 som berodde på finanskrisen. För män och kvinnor med förgymnasial utbildning har arbetslöshet fortsatt att öka efter år 2009 med olika fart i trend, medan för de med gymnasial och eftergymnasial utbildning har visat skillnader i trenderna.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: angelama94@hotmail.com. Handledare: Martin Sköld.

Förord

Detta arbete är ett examensarbete i matematisk statistik som motsvara 15 högskolepoäng vid Stockholms Universitet. Jag vill ge ett stort tack till min handledare Martin Sköld för all hjälp under arbetets gång.

Innehåll

1	Inledning	4
2	Teori	5
2.1	Generaliserade linjära modeller	5
2.2	Oddsquot	6
2.3	Logistisk regression	6
2.4	Spline	7
2.4.1	B-spline	7
2.5	AIC	8
2.6	Likelihood ratio test	8
2.7	Konfidensintervall	9
3	Data	9
3.1	Mål och syfte	9
3.2	Beskrivning av data	9
3.3	Mer om data	10
4	Modellering	11
4.1	Antagande om GLM	12
4.2	Analys av arbetslöshet efter utbildningsnivå	13
4.3	Analys av könets inverkan	17
4.4	Prediktion	22
5	Diskussion	24
6	Referenser	27
7	Appendix	28

1 Inledning

Arbetslöshet är ett problem både för individen och för samhället, den ser olika ut i olika utbildningsgrupper. Enligt siffror från Statistiska Centralbyrån(SCB), idag har 12% inte utbildat sig vidare efter grundskola, medan 26% är högutbildade, jämfört med år 1990 hade 33% var lågutbildade och bara 11% hade en utbildning efter gymnasiet. Utbildningsnivån i Sverige har alltså ökat mycket de senaste årtiondena, samtidigt har arbetslöshet inom olika utbildningsgrupper förändrats. Arbetslösheten i Sverige ligger nu på omkring 6.8%. Arbetslösheten för personer som saknar gymnasial utbildning ligger på cirka 20.5%, bland personer med eftergymnasial utbildning ligger på under 5%.

I en rapport publicerad av Spector(2015) diskuterade hur olika faktorer påverka arbetslöshet, de ansåg att ”Även om arbetslöshet generellt är låg i Sverige så är det stora skillnader mellan olika utbildningsgrupper. Arbetslösheten för högutbildade har minskat varje år sedan 2005. För personer med kort utbildning ökade snabbt när finanskrisen slog till och har sedan dess fortsatt att öka, medan för de med gymnasieutbildning minskade arbetslösheten något efter finanskrisen.”

I en studie publicerad Wolbers(2000) visade ” i generellt sett har kvinnor svårare att få jobb än män, men det motsatta gällande bland universitetsstudenter.”

I denna studie vill vi med hjälp av arbetslöshetsdata från Arbetsförmedlingen, undersöka trender i arbetslösheten för olika utbildningsgrupper samt arbetslösheten för män och kvinnor i respektive utbildningsgrupper. Har olika utbildningsgrupper samma trender inom arbetslöshet? Hur ser trender i arbetslöshet ut för män respektive kvinnor i dessa grupper? I denna studie vill vi besvara dessa frågor med hjälp av statistiska metoder.

Vi börjar med att gå igenom teorin som vi kommer att behöva använda, sedan gör vi en kort beskrivning av data. Vidare går vi över till modelleringen, där kommer vi först att kolla på hur trender i arbetslöshet för olika utbildningsgrupper ser ut. Därefter jobba vi med datamaterialet där kön ingår för att analysera trender inom arbetslöshet för män respektive kvinnor i olika utbildningsgrupper. Vi kommer även att skapa konfidensintervall för att undersöka osäkerheten i resultatet. Till sist predikterar vi en framtida arbetslöshet för olika utbildningsgrupper samt en framtida arbetslöshet för män respektive kvinnor i de grupperna.

2 Teori

Teorin kommer i huvudsak från Agresti(2002). Teorin om spline och B-spline är hämtad från Racine(2014)

2.1 Generaliserade linjära modeller

För att kunna använda linjär regression, finns vissa antaganden som måste vara uppfyllda, till exempel responsvariabeln måste vara normalfördelade. Om antaganden inte är uppfyllda, kan istället generaliserade linjära modeller användas.

Generaliserade linjära modeller är generalisering av den vanliga linjära modellen vars responsvariabel måste komma från en fördelning som tillhör den naturliga exponentialfamiljen.

En generaliserade linjär modell består av tre komponenter:

1. En slumpmässig komponent
2. En systematisk komponent
3. En länkfunktion

Slumpmässig komponent: Den slumpmässiga komponenten i en generaliserade linjär modell består av en responsvariabel med oberoende observationer (y_1, \dots, y_n) , där responsvariabeln komma från en fördelning som tillhör den naturliga exponentialfamiljen, vars täthetsfunktion kan skrivas på formen

$$f(y_i; \theta_i) = a(\theta_i)b(y_i) \exp\{y_i Q(\theta_i)\},$$

där $Q(\theta_i)$ kallas den naturliga parametern.

Systematisk komponent: Den systematiska komponenten är en linjär prediktor η som beskriver sambandet mellan de förklaringsvariablerna och den slumpmässig komponenten genom

$$\eta_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, n, \quad j = 1, 2, \dots, p$$

där x_{ij} är värdet av prediktor j , p är antalet förklaringsvariabler, n är antalet observationer.

Länkfunktion: En länk funktion i den generaliserade linjära modeller kopplar ihop den slumpmässiga komponenten och den systematiska komponenten, det vill säga $E(Y_i) = \mu_i$ koppla ihop η_i och μ_i genom funktionen g , så att $\eta_i = g(\mu_i)$. g är en länkfunktion som är monoton kontinuerlig differentierbar. Länkfunktionen g transformera $E(Y_i)$ till förklaringsvariablerna genom formeln

$$g(\mu_i) = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, n.$$

2.2 Oddskvot

Oddset definieras som

$$\Omega = \frac{\pi}{(1 - \pi)},$$

där π är sannolikhet att en händelse inträffar och $1 - \pi$ är sannolikhet att en händelse inte inträffar. Ett odds är alltså sannolikheten att en händelse ska inträffa i förhållande till sannolikheten för att den inte ska inträffa. Eftersom $0 \leq \pi \leq 1$, då har vi att Ω är icke-negativt och kan variera mellan 0 till oändligheten.

Om vi vill jämföra oddsen mellan två händelse, då har vi en oddskvot som

$$\Theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)}.$$

När $\Theta > 1$, betyder att det mer troligt att händelse 1 inträffar än att den inte inträffar. När $\Theta < 1$, betyder att det troligare att händelse 1 inte inträffar än att den inträffar.

2.3 Logistisk regression

Om vi har en binär responsvariabel Y , kan logistisk regression vara lämpligt att användas för att modellera. Vi har då två utfall, $Y = 0$ samt $Y = 1$. Sannolikheten att en händelse ska inträffa givet en viss förklaringsvariabel X kan formuleras som

$$\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x).$$

Eftersom vår responsvariabeln är binär, vi behöver då hitta en lämplig transformation för att kunna genomföra regressionen. En naturlig transformation för $\pi(x)$ är logit transformation som är modellens länkfunktion. Detta kan vi skrivas på följande sätt

$$\text{logit}(\pi(x)) = \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x,$$

där $0 \leq \pi \leq 1 \implies -\infty < \text{logit}(\pi(x)) < \infty$.

Den logistisk regressions modell ges av

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

$\pi(x)$ minskar om $\beta < 0$ och $\pi(x)$ ökar om $\beta > 0$, då $x \rightarrow \infty$.

Fallet då man har flera förklaringsvariabler, byts βx mot $(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$.

2.4 Spline

Om vi har en generaliserad linjär modell, och datapunkter inte kan förklaras av en rät linje, då kan vi använda en polynomfunktion som kan beskriva datapunkter av högre ordning. Fallet då vi har fler än tre grads polynom, kan en vanlig polynomfunktion vara svårt att jobba med, eftersom kurva kommer ha många upp- och nedgångar. Ett annat sätt som kan beskriva data av högre grads polynom är spline, där den dela in funktionen i mindre intervall, i varje intervall tilldelas en polynom.

En spline funktion defineras

$$F(x) = b_0 P_0(x) \mathbb{1}\{t_0 \leq x \leq t_1\} + b_1 P_1(x) \mathbb{1}\{t_1 \leq x \leq t_2\} + \dots + b_k P_k(x) \mathbb{1}\{t_k \leq x \leq t_{k+1}\},$$

där b_i är konstant koefficienter, $P_i(x)$ är polynom funktioner som kallas för basfunktion och t_i är knutar. Som vi ser från formen, en spline funktion sätter ihop ett antal basfunktioner, dessa basfunktioner är polynomen multiplicerade med indikatorvariabler. Indikatorvariabeln är 1 i det intervallet där vi vill använda denna funktion och 0 annars.

En nackdel med dess spline är när vi flytta en knut, så kommer hela kurvan formas om. Detta innebär att om en koefficient ändras, så kommer basfunktionen också ändras.

2.4.1 B-spline

I denna studie kommer vi att använda oss av B-spline som är en annan typ av splinefunktion. B-spline är en linjär kombination av kontrollpunkter och basfunktioner. Till skillnad från en vanlig spline, en B-spline kurva går inte genom alla kontrollpunkter, utan den ligger på närhet av dessa punkter, vilket gör att det blir lättare att göra förändringar i kurvan.

En B-spline funktion ges av

$$S(x) = \sum_{i=0}^{N+p} \beta_i B_{i,p}(x),$$

där $B_{i,p}(x)$ är basfunktioner av grad p och β_i är koefficienter som också kallas för kontrollpunkter. Vi har totala N inre knutar och två yttre knutar som är ordnade i stigande ordning

$$T = (t_0 < t_1 < \dots < t_{N+1}),$$

där t_0 och t_{N+1} är en undre respektive en övre knut.

Basfunktioner $B_{i,p}(x)$ definieras

$$B_{i,0}(x) = \begin{cases} 1 & \text{om } t_i \leq x < t_{i+1} \\ 0 & \text{annars} \end{cases}$$
$$B_{i,p}(x) = \frac{x - t_i}{t_{i+p} - t_i} B_{i,p-1}(x) + \frac{t_{i+p+1} - x}{t_{i+p+1} - t_{i+1}} B_{i+1,p-1}(x)$$

Vidare kan vi skriva en regressionsmodell med term av B-spline funktioner som

$$F(x) = b_0 B_0(x) + b_1 B_1(x) + \dots + b_k B_k(x) + \epsilon,$$

där b_i är koefficienterna, B_i är bas funktioner och ϵ är en felterm.

2.5 AIC

AIC står för Akaike's information criterion, vi använda det för att jämföra olika modeller, där de modellerna är baserade på samma data. Det betyder att AIC säger ingenting om hur bra en modell är, utan den gör en jämförelse med hur bra en modell är jämför med en annan modell. Ju lägre AIC-värde, desto bättre anpassning.

AIC definieras som

$$AIC = -2(\log(L) - p),$$

där p är antal parametrar i modellen och L är den maximerad likelihood funktionen för modellen.

2.6 Likelihood ratio test

Likelihood Ratio-test används för att jämföra en modell mot en alternativ modell. Låt Modell 1 vara en komplett modell och Modell 0 är en special fall av Modell 1, då har vi :

$$H_0 : \text{Modell 0 håller.}$$

$$H_1 : \text{Modell 1 håller, men inte Modell 0.}$$

Likelihood ratio test har följande formen

$$-2\log(\Lambda) = -2\log(L_0/L_1) = -2(l_0 - l_1),$$

där $\Lambda = L_0/L_1$ och $-2\log(\Lambda)$ är asymptotiskt χ^2 - fördelad under nollhypotesen med antal frihetsgrader lika med differensen i antal parametrar mellan Modell

1 och Modell 0. L_0 och L_1 är den maximerade likelihood funktionen för nollhypotes respektive alternativ hypotes. Med hjälp av likelihood ratio test kan vi bestämma om vi ska välja den kompletta modellen eller den enklare modellen.

2.7 Konfidensintervall

Genom att skapa ett konfidensintervall kan man undersöka osäkerhet i en modell. För att kunna skapa ett konfidensintervall för parametrarna θ , använda vi Maximum likelihood metoden för att skatta parametrar och dess skattningarnas varians, vilket följer en asymptotisk normalfördelning.

Konfidensintervall för θ definieras

$$\hat{\theta} \pm z_{(1+\gamma)/2} \cdot se(\hat{\theta})$$

där $\hat{\theta}$ är predikterade värdet, $se(\hat{\theta})$ är skattade standardavvikelse för $\hat{\theta}$ som bestäms med hjälp av inversen av Fisher informationsmatrisen och $z_{(1+\gamma)/2}$ är normalfördelningens kvantilen.

Vi kommer få två kurvor som stänger in de predikterade värdena, bredden på konfidensintervallet speglar osäkerheten i skattningarna, ju bredare intervallet är desto osäkrare är skattningarna.

3 Data

3.1 Mål och syfte

Syftet med denna uppsats är att undersöka tidstrender inom arbetslöshet i olika utbildningsgrupper mellan åren 2006-2015 samt trender inom arbetslöshet för män respektive kvinnor i dessa utbildningsgrupper.

3.2 Beskrivning av data

Datamaterialet som vi arbetar med är hämtad från Arbetsförmedlingen. Data består av antalet personer i arbetskraften och antalet arbetslösa i åldern 16-64 år i Sverige mellan åren 2006-2015 som är fördelad efter utbildningsnivå. Data innehåller även information om antalet män respektive kvinnor i arbetskraften samt antalet arbetslösa män och kvinnor under dessa år.

Arbetskraft: Arbetskraften består av personer som är sysselsatta och arbetslösa .

Arbetslös: Personer räknas som arbetslösa om de är inskrivna på Arbetsförmedlingen.

Utbildningsnivå: En klassificering av utbildningsnivå som gjord enligt Svensk Utbildningsnomenklatur (SUN), där anger den högsta utbildningsnivån som en person har för närvarande. Utbildningsgruppen består av följande: Förgymnasial utbildning, Gymnasial utbildning, Eftergymnasial utbildning.

Kön: Kvinna eller Man.

År: Mellan åren 2006 och 2015.

3.3 Mer om data

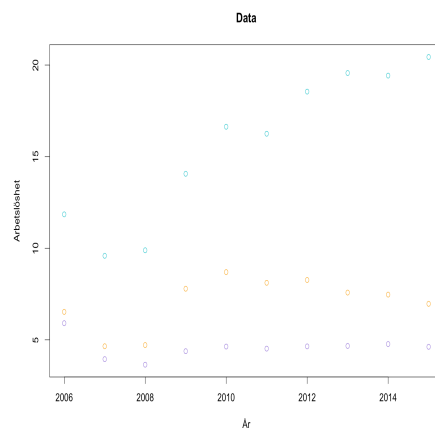


Figure 1: Arbetslöshet för personer med förgymnasial utbildning(grön), gymnasial utbildning(orange), eftergymnasial utbildning(lilla) under åren 2006-2015.

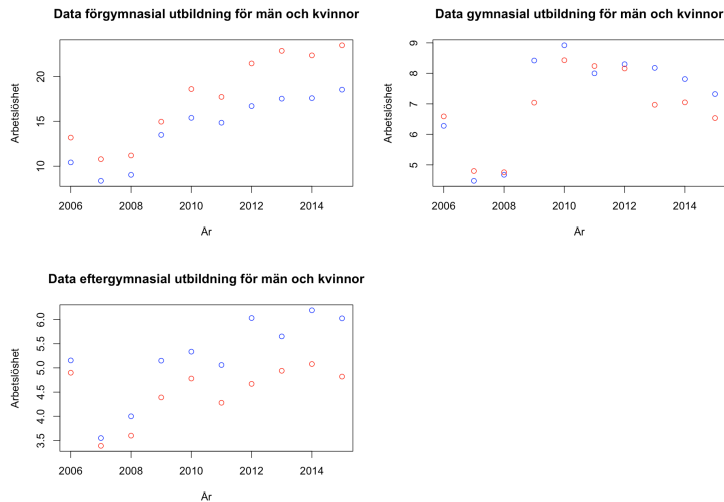


Figure 2: Arbetslöshet för män(blå) respektive kvinnor(röd) med förgymnasial utbildning, gymnasial utbildning och eftergymnasial utbildning under åren 2006-2015.

Figur 1 beskriver hur arbetslösheten ser ut mellan de olika utbildningsgrupperna i Sverige under åren 2006-2015. Vi har totala 10 punkter, en för varje år. Vi ser att arbetslösheten bland personer med högst förgymnasial utbildning ökade kraftigt under åren 2008-2015, speciellt mellan år 2008 och i slutet av år 2009.

För personer med gymnasial utbildning och eftergymnasial utbildning ökade inte arbetslösheten kontinuerligt, men mellan år 2008 och i slutet av år 2009 skedde en kraftig ökning av arbetslösheten. Orsaker till snabbt ökande arbetslöshet i samtliga utbildningsgrupper under dessa år kan vara mellan år 2008 och år 2009 påverkade svensk ekonomi negativt av finanskrisen, vilket leder till att behovet av arbetskraft minskade, antalet arbetslösa ökade kraftigt under tiden. Vi ser även att arbetslösheten i eftergymnasial utbildningsgrupp är bara ungefär en tredjedel av förgymnasial utbildningsgrupp.

I Figur 2 ovan presenteras arbetslöshet för män och kvinnor i olika utbildningsgrupper mellan åren 2006-2015, där vi har totala 20 punkter, varje kön har 10 punkter, en för varje år. Vi ser även här att arbetslösheten drabbades hårt av finanskrisen mellan år 2008 och i slutet av år 2009 för både män och kvinnor i samtliga utbildningsgrupper.

4 Modellering

Arbetslöshet är inte slumpmässig, vissa individer har högre risk att bli arbetslös än andra. I denna studie vill vi undersöka trender inom arbetslöshet i olika ut-

bildningsgrupper i Sverige mellan åren 2006-2015 samt trender i arbetslöshet för respektive kön i de olika utbildningsgrupperna, vidare kommer vi även prediktera arbetslösheten i Sverige under år 2016.

Då syftet är att undersöka tidstrender inom arbetslöshet för personer med olika utbildningsnivåer, därför börjar vi med att ställa upp tre lämpliga modeller, en för respektive utbildningsgrupper. För att undersöka tidstrender för arbetslösheten i de olika utbildningsgrupperna under åren 2006-2015 skatta vi en B-spline för de tre grupperna separat, vidare kommer vi skapa ett 95% konfidensintervall runt kurvan för att se osäkerheten i skattningar. Därefter vill vi även få en bild av hur arbetslöshet i olika utbildningsgrupper varierar beroende på kön, då behöver vi konstruera en regression med en splinekurva för män respektive kvinnor separat, sedan skapa vi ett 95% konfidensintervall för dem. Sist vill vi även prediktera arbetslösheten för respektive utbildningsgrupper under år 2016, detta görs genom att lägga en övre knut vid 2016.

4.1 Antagande om GLM

Modellen för multipel linjär regression bygger på några förutsättningar som måste vara uppfyllda. Dessa förutsättningar är:

1. Residualerna i modellen måste vara normalfördelade.
2. Observationerna måste vara oberoende.
3. Variansen i residualerna måste vara konstant.

Antagandet är uppfyllt om residualerna i residual plot fördelas jämnt runt noll. Residual plottarna för olika utbildningsgrupper visar att residualerna inte är oberoende, det följer ett systematiskt mönster vilket tyder på att antagande är inte uppfyllt, vi kan alltså inte använda oss av en multipel linjär regression.

Eftersom varje person i arbetskraften är antingen arbetslös eller sysselsatt, därför har vi en respons variabel Y som antar värdet 1 en person är arbetslös eller 0 en person är sysselsatt, det vill säga $Y \sim Be(\pi)$ och π är sannolikheten att en person är arbetslös. Summan av alla oberoende bernoullivariabler är binomialfördelad, vilket ger $Z = \sum_{i=1}^n Y_i \sim Bin(n, \pi)$, där n är antalet personer i arbetskraften. Vi kan alltså använda oss av en generaliserad linjär modell, där sannolikhetsfördelningen kan skrivas på formen

$$f(y_i; \theta_i) = a(\theta_i)b(y_i) \exp\{y_i Q(\theta_i)\}$$

Parametern θ_i kan variera för $i=1, \dots, N$ och $Q(\theta)$ kallas den naturliga parametern.

Då syfte är att analysera tidstrender inom arbetslöshet i tre olika utbildningsgrupper, därför i vårt fall har vi tre binomialfördelad responsvariabler, sannolikhetsfunktionen har formen

$$P(Z = y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad y = 1, 2, \dots, n$$

där n och y anger antalet personer i arbetskraften samt antalet arbetslösa. Sedan har vi $\theta = \frac{\text{Antalet arbetslösa}}{\text{Antalet personer i arbetskraften}}$ och $0 \leq \theta \leq 1$. Vi kan vidare skriva om sannolikhetsfunktionen som

$$\begin{aligned} & \binom{n}{y} \exp\{y \ln(\theta) + (n - y) \ln(1 - \theta)\} = \\ & = \binom{n}{y} \exp\left\{y \ln\left(\frac{\theta}{1 - \theta}\right)\right\} \exp\{n \ln(1 - \theta)\} \end{aligned}$$

Vi har alltså sannolikhetsfördelning som tillför den naturliga exponential familjen med

$$a(\theta) = \exp\{n \ln(1 - \theta)\}$$

$$b(y) = \binom{n}{y}$$

$$Q(\theta) = \ln\left(\frac{\theta}{1 - \theta}\right) = \text{logit}(\theta)$$

där den naturliga parametern $\ln\left(\frac{\theta}{1 - \theta}\right)$ är en logit-länkfunktion.

4.2 Analys av arbetslöshet efter utbildningsnivå

Först är vi intresserade av att undersöka trender inom arbetslöshet i olika utbildningsgrupper under åren 2006-2015. Vi delar upp datasetet för olika utbildningsgrupper, då har vi tre generaliserad linjär modeller med responsvariabeln antalet arbetslösa som är binomialfördelad och logit funktionen som länkfunktion, alltså har vi tre logistisk regressionsmodeller som kan skrivas på formen

$$\text{logit}(\pi(x_i)) = \alpha + \beta x_i, \quad i = 1, 2, \dots, 10$$

där $\pi(x_i) = P(Y = 1|X = x_i)$ och x_i anger ett visst kalenderår.

För att få en bild av trender i arbetslösheten mellan åren 2006-2015, skattar vi en linje för vår logistisk regressionsmodell.

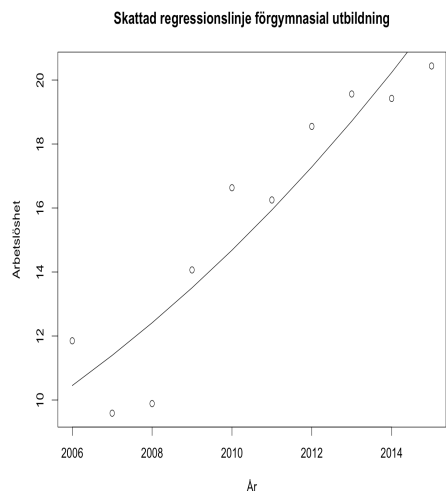


Figure 3: Datapunkter samt skattad regressionslinje för personer med gymnasial utbildning

Av att titta på figuren ovan kan vi se att den skattade linjen inte passar våra datapunkter särskilt bra. Linjen har en positiv lutning bland personer med gymnasial utbildning, men datapunkterna har inte ett perfekt linjärt samband som är positivt, utan den vänder flera gånger mellan åren 2006 och 2015, det vill säga att man behöver beskriva förändringen i arbetslöshet med en högre grads polynom, vi väljer därför att använda oss av en B-spline som kan kombinera flera polynom. Samma gäller arbetslöshet för personer med gymnasial utbildning och eftergymnasial utbildning. Plottar för gymnasial och eftergymnasial utbildningsgrupp ses i Appendix.

Det finns flera sätt att skapa en B-spline, i denna studie används vi den standardinställningen i R som använder polynom av grad 3 och vi använder kvartilerna av X (årtalen) för knutarnas placering. Om vi till exempel har tre inre knutar, då används 25%, 50%, 75% kvartilerna av årtalen som knutar. Genom att analysera deras AIC-värden fördrar vi modellen som har lägst AIC.

Eftersom vi ska undersöka tidstrender i arbetslöshet mellan åren 2006-2015, då kommer vi att ha tio punkter, för att undvika överanpassning väljer vi att bara titta på modellen med 2, 3, 4, 5 knutar. Vi börjar med att bara ha två yttre knutar, detta ger ett AIC-värde på 7534.9, vidare placerar vi en inre knut vid 50% kvantilen, vi får då ett AIC på 3422.5. Därefter testar vi två inre knutar vid 33% kvantilen och 67% kvantilen. Vi fortsätter att addera antalet knutar tills vi har fem knutar, sedan undersöker vi AIC på de modellerna, då vi använder AIC som ett mått för att hitta den lämpligaste modellen. Tabellen nedan visar att vi får bäst anpassning för gymnasial utbildningsgrupp när vi har fyra knutar

vid (2006, 2009, 2012, 2015).

Antal knutar	AIC
2	7534.9
3	3422.5
4	1617.6
5	2114.5

Table 1: Olika antal knutar och dess AIC-värden för föregymnasial utbildningsgrupp

Vi genomför samma process för gymnasial samt eftergymnasial utbildningsgrupp. Genom att köra samma procedur får vi bäst anpassning för de grupperna med fyra knutar, vilket är samma som förgymnasial utbildningsgrupp.

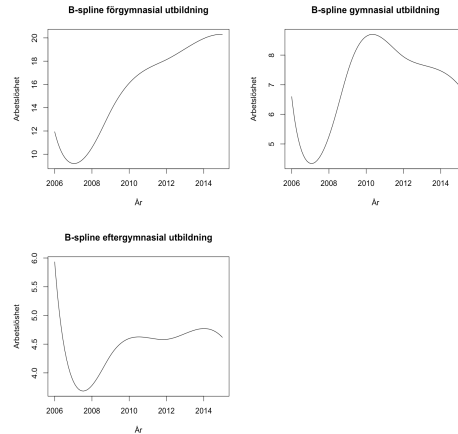


Figure 4: B-spline för olika utbildningsgrupper

I Figur 4 presenteras B-spline för olika utbildningsgrupper. Vi ser att vår B-spline passa datapunkterna bättre än de linjerna vi hade tidigare. Om vi jämför Figur 4 med Figur 1, upptäcker vi att den kraftigt ökning av arbetslösheten som ska finnas mellan år 2008 och år 2009 har flyttat när vi anpassa B-spline för de grupperna. Ökningen har flyttat till år 2007 och år 2009, vilket inte stämmer med våra datapunkter. För att få en bättre B-spline som kan beskriva trender inom arbetslöshet i olika utbildningsgrupper, lägger vi en knut vid 2008. Tabellen nedan visar AIC för modellen med knutar vid 2008 och utan.

Utbildningsnivå	Knutar	AIC
Förgymnasial	(2006,2009,2012,2015)	1617.6
	(2006,2008,2009,2012,2015)	1377.5
Gymnasial	(2006,2009,2012,2015)	3271.9
	(2006,2008,2009,2012,2015)	2280.5
Eftergymnasial	(2006,2009,2012,2015)	329.0
	(2006,2008,2009,2012,2015)	255.4

Table 2: Olika placering av knutar samt dess AIC-värden för olika utbildningsgrupper

Det visar sig att AIC minskar i alla tre utbildningsgrupper när vi placera knuten vid 2008, vilket betyder att modellen med fem knutar vid (2006, 2008, 2009, 2012, 2015) ger bättre anpassning för data. Plotten för B-spline med ny placering av knutar visas i Appendix.

Vi ser i Figur 9(se Appendix) att det är betydligt högre arbetslöshet i förgymnasial utbildningsgrupp än andra grupper. För denna gruppen har en väldigt anorlunda trenden, mellan år 2006 och 2007 minskar arbetslösheten med en nedåtgående trend, från och med år 2008 så ökar arbetslösheten med en snabb stigande trend, därefter fortsätter arbetslösheten med uppåtgående trend i en långsam takt.

Från år 2006 till 2007 visar en nedåtgående trend på arbetslösheten för både gymnasial och eftergymnasial utbildningsgrupp, där är det lägsta nivå på arbetslösheten på tio år. Från och med år 2008 vänder arbetslösheten med en stigande trend fram till slutet av år 2009 för både utbildningsgrupper. Därefter sjunker trenden långsamt för personer med gymnasial utbildning medan för personer med eftergymnasial utbildning så ligger arbetslösheten kvar på ungefär samma nivå som år 2009.

Om vi jämför de tre B-spline så ser vi att gymnasial utbildningsgrupp har högre arbetslöshet än eftergymnasial utbildningsgrupp, arbetslöshet bland personer med eftergymnasial utbildning är bara ungefär en tredjedel av personer med förgymnasial utbildning, vilket stämmer överens med verkligheten. Vi ser också att personer med förgymnasial och gymnasial utbildning påverkar mest av finanskrisen, mellan år 2008 och 2009 uppgick arbetslösheten kraftigare för de två grupperna än eftergymnasial utbildningsgrupp.

Nästa steg är att undersöka hur säker är dessa skattningar, i denna studie väljer vi att skapa ett 95%-igt konfidensintervall. För att skapa ett konfidensintervall för arbetslöshet, börjar vi med att ha ett konfidensintervall för $\text{logit}(\pi(x_i))$, sedan transformera detta.

Ett Wald-konfidensintervall för $\text{logit}(\pi(x_i))$ ser ut som följande

$$KI(\text{logit}(\pi(x_i))) : \hat{\alpha} + \hat{\beta}x_i \pm 1.96 [(1 \ x_i) I^{-1} \begin{pmatrix} 1 \\ x_i \end{pmatrix}]^{1/2} = (\gamma_1, \gamma_2), \quad i =$$

1, 2, ..., 10

där I^{-1} är den inversa Fisher informationsmatrisen.

Efter transformationen får vi ett konfidensintervall för $\pi(x_i)$

$$KI(\pi(x_i)) = \left(\frac{\exp\{\gamma_1\}}{1 + \exp\{\gamma_1\}}, \frac{\exp\{\gamma_2\}}{1 + \exp\{\gamma_2\}} \right)$$

I Figur 5 visar B-spline med 95%-igt konfidensintervall för respektive utbildningsgrupper.

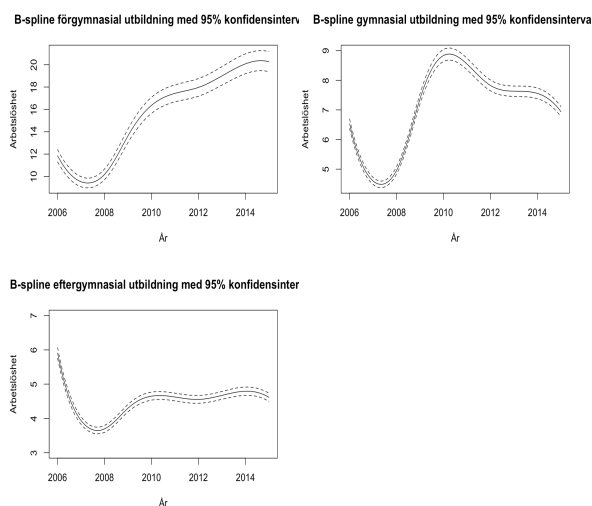


Figure 5: B-spline med 95%-igt konfidensintervall för olika utbildningsgrupper

Bredden på konfidensintervallet speglar osäkerheten i skattningarna. För alla tre utbildningsgrupper verkar konfidensintervall följer våra B-spline bra, det är bredare från och med år 2009 för samtliga utbildningsgrupper. Om vi jämför alla tre konfidensintervall, så ser vi att konfidensbanden är bredare för förgymnasial och gymnasial utbildningsgrupp mellan åren 2009-2015 jämfört med eftergymnasial utbildningsgrupp, vilket tyder på att det är mer osäkert på hur arbetslösheten har förändrat under dessa år för förgymnasial samt gymnasial utbildningsgrupp.

4.3 Analys av könets inverkan

Nu undersöka vi hur arbetslösheten för olika utbildningsgrupper varierar beroende på kön. För att undersöka detta, delar vi upp data för män och kvinnor. På samma sätt som för datamaterialet utan kön, vi plottar arbetslöshet mot år för respektive kön, sedan skatta vi en varsin logistisk regressionslinje för män och kvinnor. Det visar sig att de skattade linjerna inte passa datapunkter särskilt

bra för både könen i samtliga utbildningsgrupper. För att få en bättre anpassning behöver vi skatta en B-spline för män och en B-spline för kvinnor i respektive utbildningsgrupper. Men nu när vi har två faktorer kön och år, kan vara intressant att testa om det föreligger samspelseffekt mellan kön och splinefunktion för år. För att kunna göra det slå vi först ihop datamaterialet och introducerar en samspelsterm som beskriver relationen mellan kön och B-spline för år.

Först ska vi bestämma B-spline funktion för både modeller. Vi använda samma metod som tidigare, där vi använda kvantilerna av årtalen för knutarnas placering, genom att analysera deras AIC-värden fördrar vi modellen med lägst AIC. Då vi ska jämföra modellerna, därför vill vi helst ha samma antal knutar för båda modeller.

Som vi nämnde tidigare, för att undvika överanpassning, titta vi endast på modellen med 2, 3, 4, 5 knutar. Tabellen nedan visar AIC för modellen med och utan samspelsterm för förgymnasial utbildningsgrupp, tabellerna för övriga utbildningsgrupper presenteras i Appendix.

Antal knutar	AIC(Modellen med samspelsterm)	AIC(Modellen utan samspelsterm)
2	8625.0	8171.8
3	4793.7	4306.3
4	2649.2	2065.3
5	3341.8	2754.4

Table 3: AIC för modellen med och utan samspelsterm(förgymnasial utbildning)

Som vi ser i Tabell 3, Tabell 7 och Tabell 8 (se Appendix), bäst anpassning för samtliga utbildningsgrupper fås då vi har fyra knutar för både med samspelsterm och utan samspelsterms modell.

Nu har vi hittat B-spline för både modeller, nästa steg är att testa om det föreligger samspelseffekt mellan kön och splinefunktion för år. Vår modell har nu utseendet

$$\text{logit}(\pi(x_i)) = \alpha + \beta \mathbb{1}\{M\}_i + S_1(x_i) + S_2(x_i) \mathbb{1}\{M\}_i, \quad i = 1, 2, \dots, 10$$

där $\mathbb{1}\{M\}$ är en indikatorvariabel som antar värdet 1 för män, 0 för kvinnor. S_1 och S_2 är splinefunktioner.

De splinefunktioner har följande formen

$$S_1(x_i) = \beta_{11}B_1(x_i) + \beta_{12}B_2(x_i) + \beta_{13}B_3(x_i) + \dots + \beta_{1n}B_n(x_i)$$

och

$$S_2(x_i) = \beta_{21}B_1(x_i) + \beta_{22}B_2(x_i) + \beta_{23}B_3(x_i) + \dots + \beta_{2n}B_n(x_i)$$

där B_k är basfunktioner för $k=1,2,\dots,5$.

Genom att använda likelihood ratio test kan vi testa hypotesen att samspelstermen inte har någon signifikant effekt. Vår nollhypotes och alternativhypotes är

$$H_0 : \beta_{21} = \beta_{22} = \dots = \beta_{25} = 0$$

$$H_1 : \text{minst en av } \beta_{2k} \neq 0$$

Om samspelseffekt mellan kön och B-spline för år har en signifikant effekt, då väljer vi den modell med samspelsterm, annars väljer vi den enklare modellen, där män och kvinnor har samma splinefunktion.

Tabellen nedan visar P-värdet vid likelihood ratio test för olika utbildningsgrupper. Vi väljer signifikansnivå på 5%, om P-värdet är lägre än 5% så förkastas nollhypotesen.

Utbildningsnivå	P-värde
Förgymnasial	2.2e-16
Gymnasial	2.2e-16
Eftergymnasial	2.2e-16

Table 4: Likelihood ratio test mellan kön och B-spline för år för olika utbildningsgrupper

Det visar sig att P-värde är väldigt litet i samtliga utbildningsgrupper, vilket betyder att vi ska förkasta vår nollhypotes på alla signifikansnivå, samspellet mellan kön och B-spline för år har signifikant effekt. Detta innebär att män och kvinnor har olika splinefunktion och kan ha olika antal knutar.

Då syftet är att undersöka tidstrender inom arbetslöshet för män respektive kvinnor i olika utbildningsgrupper, därför väljer vi att skatta en B-spline för män och kvinnor i respektive utbildningsgrupp. Vi kör på samma process som modell utan kön, det visar sig att vi få lägst AIC med fyra knutar vid (2006, 2009, 2012, 2015) för män respektive kvinnor i samtliga utbildningsgrupper. Sedan uppstår samma problem som tidigare, spline kurva som vi har med knutar vid (2006, 2009, 2012, 2015) förklara inte den kraftigt ökning som skulle finnas mellan år 2008 och år 2009, därför väljer vi att lägga en knut vid 2008, då har vi fem knutar som är vid (2006, 2008, 2009, 2012, 2015), vilket är samma som modellen utan kön. Vidare undersöker vi AIC på de nya spline funktioner, det visar sig att de nya spline funktioner ger även lägre AIC-värde än när vi hade knutar vid (2006, 2009, 2012, 2015), vilket betyder att modellen med knutar vid (2006, 2008, 2009, 2012, 2015) ger bättre anpassning för data. Figuren nedan

visar B-spline för män och kvinnor i respektive utbildningsgrupper med knutar vid (2006, 2008, 2009, 2012, 2015).

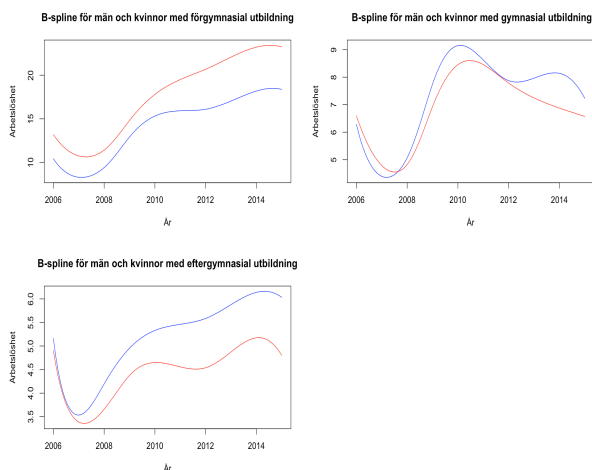


Figure 6: B-spline för män(blå) respektive kvinnor(röd) i olika utbildningsgrupper

Vi ser från Figur 6 att arbetslöshet för män i förgymnasial utbildningsgrupp har nästan samma form som för modell utan kön. Mellan år 2006 och 2009, tenderer inom arbetslöshet ser snarlik ut för både kön i förgymnasial utbildningsgrupp, där arbetslöshet har en sjunkande trend mellan åren 2006-2007, sedan har en snabbt ökande trend mellan åren 2008-2009. Efter år 2009 ökar arbetslöshet hos män mycket långsamt medan för kvinnor har en kraftigare ökande trend i arbetslöshet. Vi ser även att kvinnornas arbetslöshet i denna grupp ligger alltid på högre nivå än männens arbetslöshet.

Det gäller samma för män respektive kvinnor i gymnasial och eftergymnasial utbildningsgrupp, där mellan åren 2006-2009 har mäns och kvinnors arbetslöshet liknande form, från och med år 2010 börjar trender inom arbetslöshet skiljer sig för respektive kön.

Mellan åren 2006-2007 har arbetslöshet en nedgående trend för både kön i gymnasial och eftergymnasial utbildningsgrupp, därefter har arbetslöshet en kraftigt ökande trend mellan åren 2008-2009.

För män med gymnasial utbildning har en mer upp och nedgång mellan åren 2010-2015, medan för kvinnor med samma utbildning har arbetslöshet en sjunkande trend under dessa år.

För män med eftergymnasial utbildning ökar arbetslöshet långsamt under åren

2010-2015, medan för kvinnor med samma utbildning har en mer upp och nedgång i trend under dessa år. Vi ser även att kvinnornas arbetslöshet i denna grupp ligger i stort sett under männens arbetslöshet.

Vidare vill vi även skapa ett 95%-igt konfidensintervall för B-spline för män och kvinnor i respektive utbildningsgrupper för att undersöka hur säker är våra skattningar.

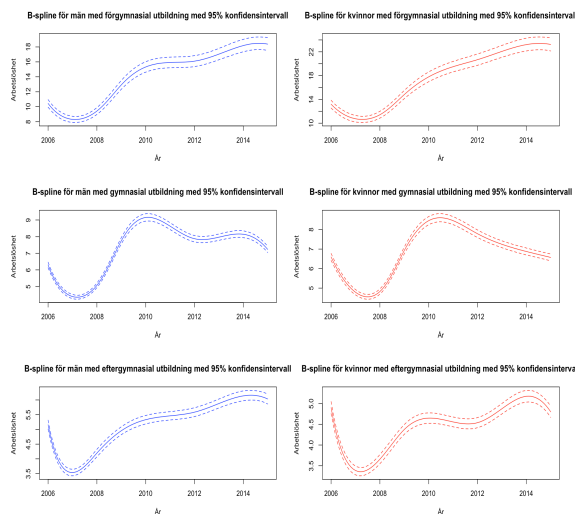


Figure 7: B-spline män respektive kvinnor med 95%-igt konfidensintervall för olika utbildningsgrupper

Vi ser från Figuren att konfidensbanden i stort sett följer vår B-spline bra för både män och kvinnor i samtliga utbildningsgrupper, bredden på konfidensintervallet är väl lik för båda könen i respektive utbildningsgrupp.

För män och kvinnor med förgymnasial utbildning är konfidensintervallet smalt mellan åren 2006-2008, men från och med år 2009 börjar konfidensintervall blir bredare för både könen. Konfidensintervallet för män och kvinnor med gymnasial utbildning ser generellt bra ut, det är bredare mellan åren 2010-2011, vilket betyder att vi är mer osäker på hur arbetslöshet för män respektive kvinnor i denna grupp har förändrat under dessa år. Men i stort sett verkar B-spline på ett bra sätt beskriver tidstrender inom arbetslöshet för män och kvinnor med gymnasial utbildning.

För eftergymnasial utbildningsgrupp ser vi att konfidensintervallet är smalt i början, sedan börjar bredda ut för både könen, vilket betyder att B-spline med mindre osäkerhet förklara tidstrender inom arbetslöshet i denna utbildningsgrupp.

4.4 Prediktion

När vi har hittat lämpliga modeller som kan beskriva trender i arbetslöshet för de olika utbildningsgrupperna samt för män respektive kvinnor i olika utbildningsgrupper, kan vara intressant att prediktera en framtida arbetslöshet för de olika grupperna. I denna studie väljer vi att prediktera arbetslöshet under år 2016.

Vi börjar med att prediktera en framtida arbetslöshet för de olika utbildningsgrupper utan att ta hänsyn till kön, vi vill påminna här att den slutgiltiga modellen som vi valde för respektive utbildningsgrupperna är en logistisk regressionsmodell med B-spline funktionen som har knutar vid (2006, 2008, 2009, 2012, 2015). Vid prediktion av arbetslöshet för år 2016, behöver vi sätta en yttre knut vid 2016, då är frågan om vi ska lägga en knut vid år 2015. Det visar sig att AIC kommer vara oförändrat om vi utesluta 2015, detta tyder på att knuten 2015 är överflödig. I detta fall väljer vi att använda tre inre knutar vid (2008, 2009, 2012) och två yttre knutar vid (2006, 2016).

Utbildningsnivå	Arbetsförmedlingen(%)	Prediktion(%)
Förgymnasial	20.5	18.7
Gymnasial	6.2	5.38
Eftergymnasial	4.4	3.9

Table 5: Arbetslöshet för olika utbildningsgrupper under år 2016

I Tabell 5 ovan visar en jämförelse mellan vår predikterade värdena och arbetslöshet som vi har hämtat från Arbetsförmedlingen. Som vi ser, vår prediktion stämmer bra med de sanna värdena.

När vi har predikterat en framtida arbetslöshet under år 2016 för de tre utbildningsgrupperna, vill vi även göra en prediktion för män och kvinnor i respektive utbildningsgrupper. Vi kör då på samma process som ovan, i Tabell 6 nedan visar en jämförelse mellan de sanna arbetslösheterna som vi har hämtat från Arbetsförmedlingen och de predikterade värdena för män och kvinnor i respektive utbildningsgrupper under år 2016.

Utbildningsnivå	Kön	Arbetsförmedlingen(%)	Prediktion(%)
Förgymnasial	Män	18.1	16.5
	Kvinnor	24.1	21.4
Gymnasial	Män	6.9	4.8
	Kvinnor	6.8	6.3
Eftergymnasial	Män	5.2	5.3
	Kvinnor	4.2	3.4

Table 6: Arbetslöshet för män respektive kvinnor i olika utbildningsgrupper under år 2016

Vi ser i Tabell 6 att vår predikterade värdena för män och kvinnor med gymnasial samt eftergymnasial utbildning väl stämmer med de värdena från Arbetsförmedlingen, arbetslöshet för män och kvinnor med förgymnasial utbildning avviker lite ifrån de sanna värdena.

5 Diskussion

Vi skattade en spline kurva för respektive utbildningsgrupper, det visade sig att arbetslösheten bland förgymnasialt utbildade hade en kontinuerlig ökande trend mellan åren 2008-2015. Trender inom arbetslöshets i samtliga utbildningsgrupp ser snarlik ut mellan åren 2006-2009, där arbetslöshet minskade mellan åren 2006-2007 och ökade mellan åren 2008-2009. Anledning till den snabbt ökande arbetslösheten som uppstår för samtliga utbildningsgrupper berodde på att finanskrisen slog hårt mot den svenska arbetsmarknaden, speciellt för de med lägre utbildning, arbetslösheten för respektive utbildningsgrupper under finanskrisen ökade på omkring 29%, 40% och 16%. Efter finanskrisen fortsatt arbetslösheten öka för personer med förgymnasialt utbildning, medan bland personer med gymnasial utbildning minskade arbetslösheten och för de med eftergymnasial utbildning låg arbetslösheten kvar på ungefär samma nivå.

Anledning till att trender inom arbetslöshet såg så annorlunda ut för respektive utbildningsgrupper kan vara att yrken som endast kräver förgymnasial utbildning minskar, detta innebär att det blir svårare för de med lågutbildning att få jobb i Sverige, alltså arbetskraften har blivit mer högkvalificerad. Vi få dock komma ihåg att det kan också hända att personer med högre utbildning arbeta i yrken som kräver lägre kvalifikationen, vilket medför att jobb som inte kräver särskilda utbildning kommer att ha stora konkurrens, arbetsgivare ersätter då personer med lägre utbildning mot högre utbildade personer.

I (Spector. 2015) kom fram till att "Till skillnad från sysselsättningskrisen på 90-talet som drabbade alla grupper tycks finanskrisen 2008 ha slagit främst mot personer med kortare utbildning. Arbetslösheten för personer med kort utbildning ökade snabbt när finanskrisen slog till och har sedan dess fortsatt att öka. Gruppen har också ändrat karaktär under perioden, allt fler i gruppen är utrikes födda. År 2009 minskade antalet jobb inom den lägst kvalifikationsnivån med fem procent och inom den näst lägsta kvalifikationsnivån med fyra procent." Detta skulle kunna vara en förklaring till varför trenden i arbetslöshet för personer med förgymnasial utbildning såg annorlunda ut jämfört med andra utbildningsgrupper, samt varför arbetslösheten ökade kraftigare bland förgymnasial och gymnasieutbildade under finanskrisen.

För att se hur säkert är vår skattningar, skapade vi ett 95%-igt konfidensintervall för B-spline för respektive utbildningsgrupper. Konfidensbanden är bredare för samtliga utbildningsgrupp mellan åren 2009-2015. Vi vet att arbetslösheten har en ökande trend mellan åren 2009-2015 i förgymnasial utbildningsgrupp, men vi kan inte säkerställa hur mycket arbetslösheten bland förgymnasial utbildade ökade mellan dessa åren. Samma gäller för gymnasial utbildningsgrupp, vi vet att arbetslöshet minskade mellan 2009-2015, men vi kan inte säkerställa hur mycket arbetslöshet minskade under dessa åren. För personer med eftergymnasial utbildning såg konfidensintervall bättre ut än andra grupper. Konfidensintervallet är lite bredare mellan åren 2009-2015, men i stort sett förklarar B-spline tidstrender inom arbetslöshet för personer som är eftergymnasial ut-

bildade bra.

Vidare undersökte vi hur arbetslösheten i olika utbildningsgrupper varierar beroende på kön. Genom att använda likelihood ratio test kom vi fram till att det finns samspel mellan kön och B-spline för år, vilket betyder att män och kvinnor har olika spline funktion.

För att undersöka tidstrender inom arbetslöshet för män respektive kvinnor i olika utbildningsgrupper, skattade vi en spline kurva för män och kvinnor i respektive utbildningsgrupp. Vi såg att arbetslösheten för samtliga utbildningsgrupper har liknande form för respektive kön mellan åren 2006-2009, från och med år 2010 har männens arbetslöshet i förgymnasial utbildningsgrupp en långsamt ökande trend medan för kvinnor med samma utbildning hade en kraftigare ökande trend, däremot låg kvinnornas arbetslöshet alltid på högre nivå än männens arbetslöshet mellan åren 2006-2015.

För gymnasial utbildningsgruppen såg vi att B-spline för män respektive kvinnor i denna grupp till stor del följer samma form som för modell utan kön. Arbetslösheten hos männen ökade snabbare än kvinnorna mellan åren 2008-2009. Vi såg också att arbetslösheten bland männen med gymnasial utbildningsnivå hade fler upp och nedgång i trend mellan åren 2010-2015, medan arbetslösheten hos kvinnorna hade en kraftig sjunkande trend under dessa år.

För män respektive kvinnor i eftergymnasial utbildningsgrupp såg vi att arbetslöshet låg på ungefär samma nivå under år 2006, sedan började andelen arbetslösa män öka och låg på högre nivå än andelen arbetslösa kvinnor. Alltså i största del låg männens arbetslöshet högre än kvinnornas i eftergymnasial utbildningsgrupp.

En förklaring till varför ökade lågutbildade kvinnors arbetslöshet snabbare från och med år 2010 kan vara att ” Låg utbildade kvinnors arbetsmarknad har försämrats mest under de senaste åren, detta skulle kunna bero på att det är systerställningen inom yrken som administration och kundtjänst som har minskat under senare år. Dessutom har det skett en uppkvalificering av utbildningsnivå i jobben. Tidigare arbetade många kvinnor med förgymnasial utbildning inom dessa yrken, men av de som finns kvar har en allt större andel högre utbildning” (Spector)

För att undersöka hur säkert vår skattning är, skapade vi konfidensintervall. Vi såg att konfidensintervallet för män och kvinnor i förgymnasial utbildningsgrupp är bredare mellan åren 2010-2015 vilket betyder vi kan säkerställa på 95% nivå att arbetslösheten hade en positiv lutning mellan åren 2010-2015, men det är svårt att avgöra hur mycket de lutar. Konfidensintervallet för män och kvinnor med gymnasial utbildning är smallare än andra grupper, det verkar som att B-spline passar bra för att beskriva tidstrender inom arbetslöshet i denna grupp. För män respektive kvinnor i eftergymnasial utbildningsgrupp såg vi att konfidensintervallet är smalt i början, och är bredare från och med år 2007, vilket betyder att vi är mindre säkert på hur arbetslöshet har förändrats i denna grupp.

Wolbers(2000) visade att i generellt sett har kvinnor svårare att få jobb än män, men att motsatta gällande bland universitetsstuderande. Detta överensstämmer med våra B-spline. Då det är högre arbetslöshet bland männen med förgymnasial utbildning än kvinnorna under åren 2006-2015, men motsatta gällande bland eftergymnasial utbildade, där till stor del ligger männens arbetslöshet över kvinnornas.

Efter vi hittade lämpliga modeller för arbetslöshet för olika utbildningsgrupper samt arbetslöshet för män och kvinnor i dessa grupper, predikterade vi arbetslöshet under år 2016. Det visade sig att vår prediktioner skiljer sig inte mycket åt de sanna arbetslösheterna som vi hämtade från Arbetsförmedlingen.

För en bättre analys behöver vi ha mer data, alltså arbetslöshet i Sverige under lägre period. Dessutom behöver vi ha fler förklaringsvariabler för att få en bättre framtida prediktion i arbetslöshet.

6 Referenser

Agresti, Alan. 2002. *Categorical Data Analysis*. 2 uppl. John Wiley & Sons.

Held, Leonhard och Sabaner Bove, Daniel. 2014. *Applied Statistical Inference :Likelihood and Bayes*. Springer-verlag Berline Heidelberg.

Sven Erik Alm och Tom Britton. 2008. *Stokastik*. 1 uppl. Liber AB, Stockholm.

Racine, Jeffrey S. 2014. *A primer on regression splines*. http://mirrors.xmu.edu.cn/CRAN/web/packages/crs/vignettes/spline_primer.pdf (2017-05-25).

S.Sinha. *A very short note on B-splines*. Department of Statistics, Texas A&M University. <https://www.stat.tamu.edu/~sinha/research/notes1.pdf> (2017-05-25).

Wolbers, M.H.J. 2000. *The effects of level of education on mobility between employment and unemployment in the Netherlands*. European Sociological Review. vol.16. no.2. pp.185-200.

Spector, Susanne(Svenskt näringsliv). *Hög sysselsättning och låg arbetslöshet* https://www.svensktnaringsliv.se/migration_catalog/Rapporter_och_opinionsmaterial/Rapporter/hog-sysselsattningpdf_620758.html/BINARY/Hg%20sysselsttning.pdf (2017-05-25).

Statistiska Centralbyrån(SCB). *Var fjärde i Sverige är högutbildad* <http://www.scb.se/hitta-statistik/sverige-i-siffror/utbildning-jobb-och-pengar/befolkningens-utbildning/#806a2fdb-04dc-4704-bcab-27d02c16e603> (2017-05-25).

7 Appendix

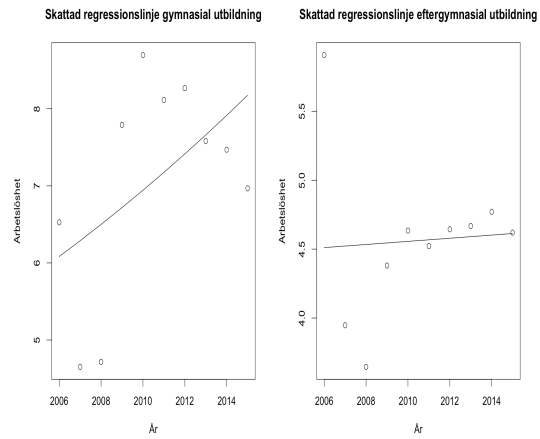


Figure 8: Skattad regressionslinje för gymnasial och eftergymnasial utbildningsgrupp samt deras datapunkter.

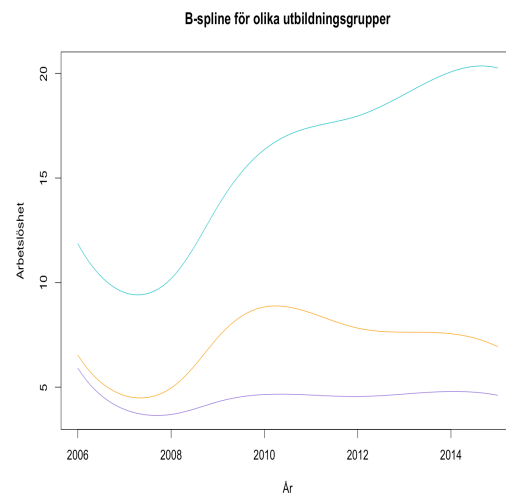


Figure 9: B-spline för olika utbildningsgrupper med inre knutar vid (2008, 2009, 2012)

Antal knutar	AIC(Modell med samspelsterm)	AIC(Modell utan samspelsterm)
2	32604.4	31405.4
3	15046.5	13800.6
4	6348.1	4311.7
5	8815.5	6565.7

Table 7: AIC för modellen med och utan samspelsterm(Gymnasial utbildning)

Antal knutar	AIC(Modellen med samspelsterm)	AIC(Modell utan samspelsterm)
2	6792.2	6183.9
3	5001.3	4385.8
4	1998.9	1361.4
5	2008.1	1368.9

Table 8: AIC för modellen med och utan samspelsterm(Eftergymnasial utbildning)