

## **Prognosmodell för medlemstal i Svenska kyrkan.**

**Av Thomas Holgersson**

Det framtida medlemsantalet i svenska kyrkan tycks vara intressant för många, då det regelbundet diskuteras i olika sammanhang. Att kyrkans medlemsantal minskar är väl känt, och det har utförts studier där man försökt analysera varför det förhåller sig så. En annan typ av frågeställning rör det framtida medlemsantalet. Kausala samband är ofta utomordentligt svåra att analysera med statistiska modeller. Däremot brukar det vara desto mer fruktsamt att göra ickekausala prognosmodeller, d.v.s. sådana modeller som inte analyserar ”varför”, utan ”hur många”. Frågan är då om man skulle kunna göra prognoser av svenska kyrkans framtida medlemsantal, med hygglig precision, och därmed få bättre underlag för ekonomiska beslut. På svenska kyrkans hemsida (<http://www.svenskakyrkan.se/statistik/>) finns data över medlemsantalet, dels som procent av befolkningen, dels som medlemsantal direkt. Data presenteras både i tabellform och i grafisk form. Där avslöjas genast att *medlemsantal* är ganska skakigt över tiden. Kanske har detta med invandring att göra, eller också finns annan orsak. Hur som helst är *medlemsandelen*, i procent av befolkningen, betydligt stabilare (se Figur 1 nedan). Frågan är då om man kan anpassa en prognosmodell till dessa data.

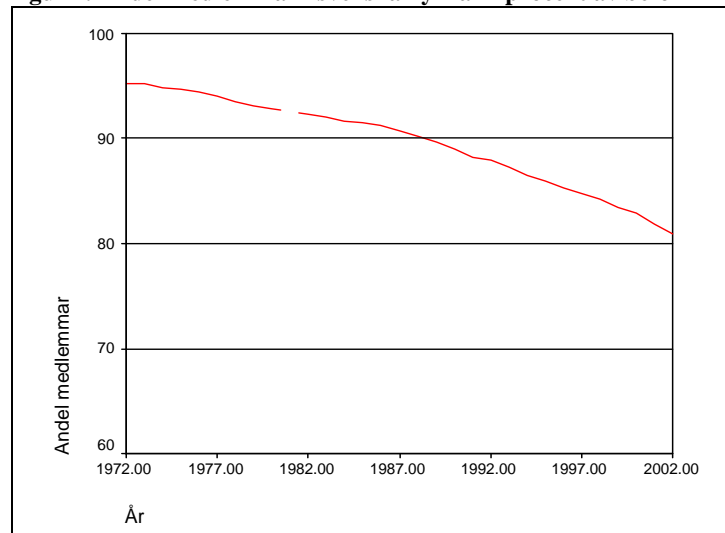
En visuell granskning avslöjar att linjen är förvånansvärt lik ett andragradspolynom. Sådana modeller brukar som regel ha autokorrelerade resttermer. Med dessa två kriterier har följande prognosmodell valts,

$$\widehat{pct}_\tau = \hat{\beta}_0 + \hat{\beta}_1(\tau - \bar{t}) + \hat{\beta}_2(\tau - \bar{t})^2 + \hat{\phi}^h(\widehat{pct}_T - pct_T) \quad (1)$$

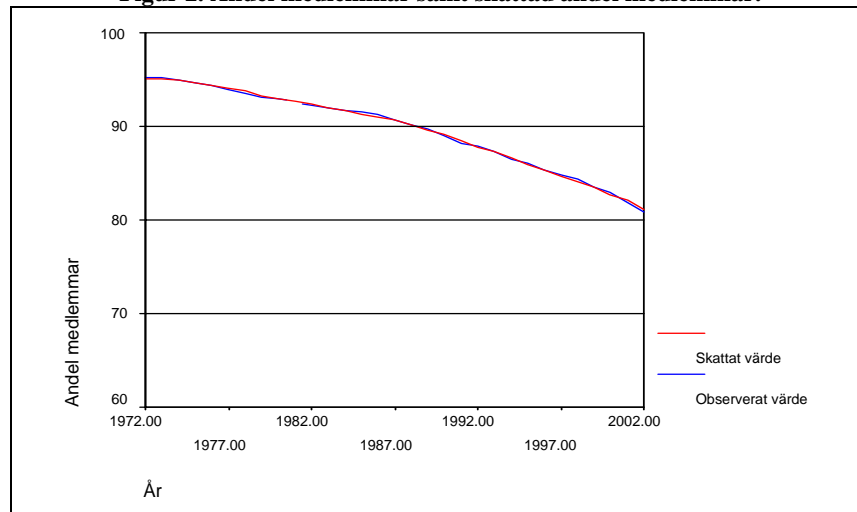
där  $\bar{t} = \left( \frac{1}{(T-1972)} \right) \sum_{t=1972}^T t$ ,  $\tau = T+h$ ,  $pct_\tau$  och  $\widehat{pct}_\tau$  är observerade resp.

prognoserade andelen medlemmar (i procent av totalbefolkningen) i svenska kyrkan vid år  $\tau$ . Orsaken till att modellen är uttryckt som en funktion av  $(\tau - \bar{t})$ , snarare än av  $\tau$  direkt, är att polynommodeller får instabila skattningar om de inte är centrerade. Se Neter [1] för vidare detaljer om polynomregression. När detta skrivs finns endast officiella data t.o.m. år 2002, dvs  $T = 2002$ . Indexet  $h$  representerar antalet år in i framtiden att prognosera. Om man t.ex. vill prognosera andelen medlemmar år 2004 blir således  $h = 2$  och  $\tau = 2002 + 2 = 2004$ . Parametrarna i (1) har estimerats med maximum likelihood metoden, se Appendix för detaljer. Skattningar av andelen medlemmar har gjorts för åren 1972-2002 (av olika skäl brukar estimering av historiska utfall kallas ”skattningar” medan estimering av framtida utfall kallas ”prognoser”). Dessa presenteras tillsammans med observerade medlemsandelar 1972-2002 i Figur 2.

**Figur 1. Andel medlemmar i svenska kyrkan i procent av befolkning.**

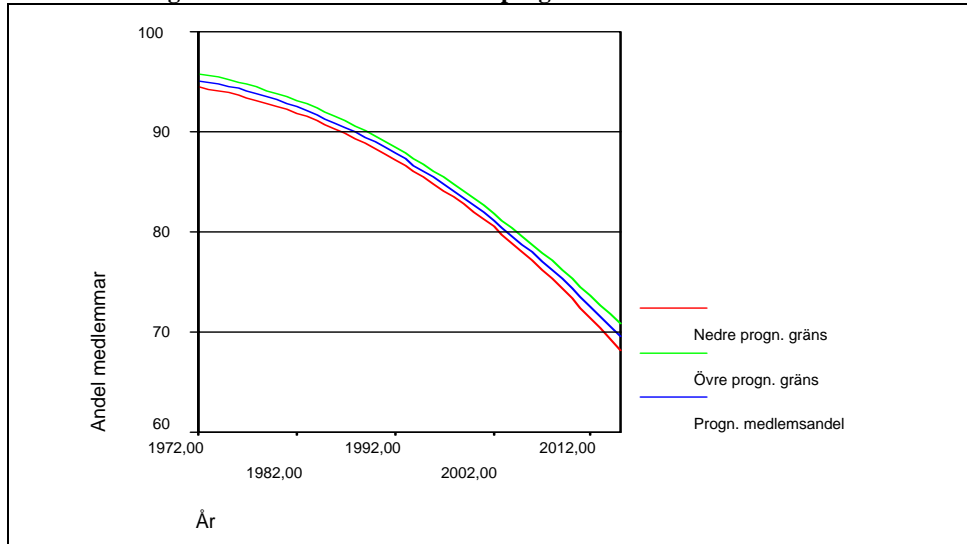


**Figur 2. Andel medlemmar samt skattad andel medlemmar.**



Anpassningen av modellen till data är förvånansvärt god för åren 1972-2002. Vad det gäller de prognoserade framtida värdena kan man naturligtvis inte veta exakt. Därför presenteras de prognoserade värdena tillsammans med 95%-iga prognosintervall i Tabell 3. Lite grovt kan man säga att prognosintervall är felmarginaler, d.v.s. inom dessa gränser kan man förvänta sig att sanningen ligger. Nu skall man tänka på att detta är en statistisk modell, och den bygger på att framtiden ser ut ungefär som idag. Stora händelser i världen, landet eller i kyrkan kommer naturligtvis att förändra situationen. Men om allt rullar på ungefär som vanligt bör modellen vara riktigt hygglig. Graferna visar tydligt trenden; raskt nedåt. De senaste 10 åren har minskningen i andelen procent av befolkningen varit 0.5-1% årligen. Några separata prognoser för 2003, 2005, 2010 och 2015 presenteras i Tabell 1. Man bör vara försiktig med att göra prognoser ännu längre fram, redan tolv år är en lång prognoshorisont, men man kan med fördel uppdatera modellen för varje nytt år (och därmed en ny observation) och fortlöpande göra prognoser fem- eller tio år framåt. Rimligtvis bör linjen så småningom plana ut, och antingen lägga sig någorlunda konstant på en fix nivå, vilket är mest rimligt, eller också få formen av ett bakvänt ”s” och fortsätta avta. Den som lever får se.

**Figur 3. Andel medlemmar samt prognoserad andel medlemmar.**



**Tabell 1. Punktprognoser med tillhörande prognosintervall.**

År	Nedre konfidensgräns	Punktprognos	Övre konfidensgräns
<b>2003</b>	79.8	<b>80.2</b>	80.5
<b>2005</b>	78.0	<b>78.6</b>	79.2
<b>2010</b>	73.1	<b>74.1</b>	75.1
<b>2015</b>	67.7	<b>69.1</b>	70.6

### Referenslista.

- [1] Neter, Kutner, Nachtsheim and Wasserman, *Applied Linear Regression Models*, ISBN 0-256-08601-X.
- [2] Greene, W. H. *Econometric Analysis*, fifth ed. ISBN 0-13-110849-2.

**Appendix.** Parametrarna i model (1) har preliminärt skattats med vanlig minstakvadratskattning. Ett klassiskt Box-Ljung test (Greene [2] s. 622) visar att resstermen är autokorrelerad. *P*-värdet för Box-Ljung testet för de fem första laggarna är väldigt små (Tabell A1). Därför har rettermen i (1) tillåtits följa en AR(1) process för att få en mer dynamisk modell (minstakvadratskattningarna är konsistenta och väntevärdesriktiga, men hypotestest och prediktionsintervall kan bli missvisande). Parametrarna har därför skattats med maximum likelihood istället (Greene [2] s. 273). Dessa finns presenterade i Tabell A2. Enligt tabell A3 är residualerna från den utökade modellen inte autokorrelerade, vilket ger förutsättningar för att beräkna prognosintervall.

**Tabell A1. Box-Ljung test av residualer från den preliminära modellen.**

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.602	.171				.		*****	*****			12.433	.000
2	.249	.168				.		*****	.			14.640	.001
3	.054	.165				.		*	.			14.747	.002
4	-.203	.161				.	****		.			16.329	.003
5	-.298	.158				.	*****		.			19.871	.001

Plot Symbols: Autocorrelations \* Two Standard Error Limits .

**Tabell A2. Maximum likelihoodestimater samt ANOVA av parametrar i prognosmodellen.**

Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	26	.60187801	.02230572	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	.711018	.14204083	5.00573	.00003310
CT	-.468683	.00833944	-56.20078	.00000000
CT2	-.010448	.00090839	-11.50221	.00000000
CONSTANT	90.451295	.12528845	721.94440	.00000000

**Tabell A3. Box-Ljung test av residualer från den sekundära modellen.**

Autocorrelations: ERR_1 Error for PCT from AREG, MOD_6													
Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.215	.171				.		****	.			1.581	.209
2	-.006	.168				.		*	.			1.582	.453
3	.120	.165				.		**	.			2.110	.550
4	-.267	.161				.	*****		.			4.847	.303
5	-.157	.158				.	***		.			5.831	.323

Plot Symbols: Autocorrelations \* Two Standard Error Limits .