Bayesian Models for Earnings Dynamics

Guy Davis   Elisa Keller   Michael Nielsen

Department of Statistics and Actuarial Science
The University of Iowa

22S:138 Project Presentation
December 9, 2009

Earnings

- The study of earnings inequality across individuals is one of the main fields of economic research
- How does the labor market reward productive attributes like schooling and work experience?
- How much of wage inequality is determined by unobservable individual-specific characteristics?
- To what extent is inequality spreading over time?

Our approach

- Investigated the extent to which observable individual-specific characteristics shape life-cycle earnings
- These characteristics will make up our covariates and include: Years of experience, Gender, and Years of Education
- Used Bayesian regression methods to estimate a standard earning function
- After checking model diagnostics, investigated fitting subsets of full model to find 'best' model

Simplifying Data

- The PSID data set is "a nationally representative longitudinal study of nearly 9,000 U.S. families. Following the same families and individuals since 1968, the PSID collects data on economic, health, and social behavior."
  - http://psidonline.isr.umich.edu/ and J. Geweke and M. Keane (2000)
- Narrowed variables from dozens to the ones of interest (Gender, Age, Education, and Wages)
- Created an 'experience' variable to replace age; still called "age" (Age - education)
- Only took a cross section of years (1975,1980,1985,1990,1995,2001)
- Changed wage over these years to all be in 1975 log(dollars)

Model's Core

- Started with a main effects model
- Had different coefficients for each year
  - Beta[1,year[i]] for ith year's intercept
  - Beta[3,year[i]] for ith year's gender coefficient
  - etc
- Added a random effects term for each individual
  - theta[subj[i]]
- Centered variables around their means
- Needed to address correlation due to nature of panel data
- car.normal prior on betas was the answer

car.normal

- car.normal also can be used for time-spatial data, not just geographically spatial
  - adj[] is a vector listing of adjacent time points- accounting for the prior and subsequent years
  - weights[] are set to 1 for each
  - num[] is a set of time points set to 1 if either 1975 or 2001 and 2 otherwise
- This follows the form of a random walk of order 1
- This modification calls for our initial betas to become (gamma + beta) for each year-betas became a zero mean random error and gamma will be our flat overall mean
  - (gamma[1] + beta[1,year[i]]) + (gamma[2] + beta[2,year[i]]) * (age[i] - age.bar) + ...
- Thus we created b[1,i]<-gamma[1]+beta[1,i] for each coefficient to monitor

Prior Specifics and Initial Values

- So the car.normal was the prior assigned to beta[] and dflat[] to gamma
- The beta[]s have a vague precision with their car.normal
  - One of our models has a precision on intercept betas that are informative gammas based on our(Elisa's) prior knowledge
- Random effect for each subject has Normal prior centered at 0 with vague precision
- Ran three chains of initial values
  - One started with estimated values of the b[]s based on a frequentist regression approach
  - Others were an uninformed, wider range of values

Model Selection

- We ran two models: informative and vague precision priors on intercept beta
- We monitored DIC but took these results lightly. We based our model selection on the credible intervals of the b[]s
  - Ran 2000 iterations, then monitored DIC over 10000 more iterations.
- Ran 12000 iterations total (DIC based on last 10000)
- Burn-in of 2000 iterations
- The two models...

Vague Intercept Priors

	1975	1980	1985	1990	1995	2001
intercept	1.623	1.558	1.555	1.530	1.545	1.648
age	0.013	0.012	0.012	0.010	0.008	0.006
gender	-0.459	-0.440	-0.401	-0.341	-0.340	-0.309
educ	0.081	0.075	0.095	0.101	0.115	0.127
DIC	26268.9					

Informative Intercept Priors

	1975	1980	1985	1990	1995	2001
intercept	1.626	1.556	1.554	1.528	1.541	1.653
age	0.013	0.012	0.012	0.010	0.008	0.006
gender	-0.458	-0.439	-0.401	-0.343	-0.340	-0.310
educ	0.081	0.075	0.095	0.101	0.115	0.126
tau.beta1	6.358	6.282	5.922	5.854	5.739	5.648
DIC	26262.2					

Conclusions

- b[]s make sense
- tau.beta1 decreases over time

Problems and Complications

- Very slow interation steps due to complicated model with car.normal
- Convergence problems was a recurring theme with most models fitted
- New concepts - car.normal