AK Replication a la Bayes

Rachel Anderson

September 23, 2018

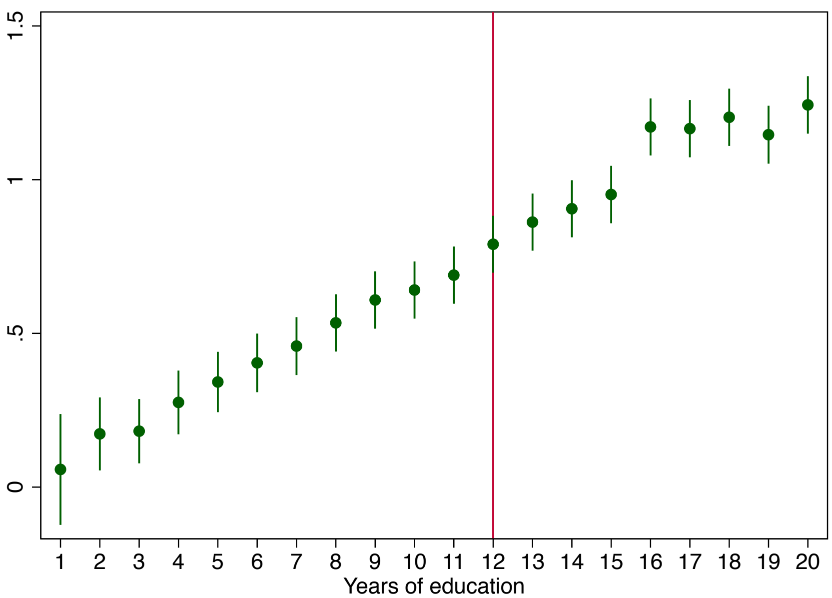

Figure 1: Plot of OLS coefficients for education variables

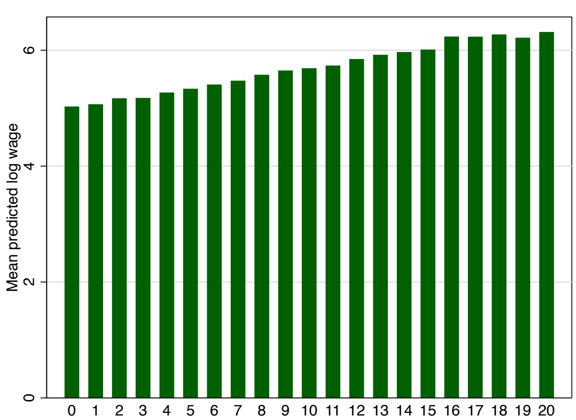

Figure 2: Predicted log-wage by education level

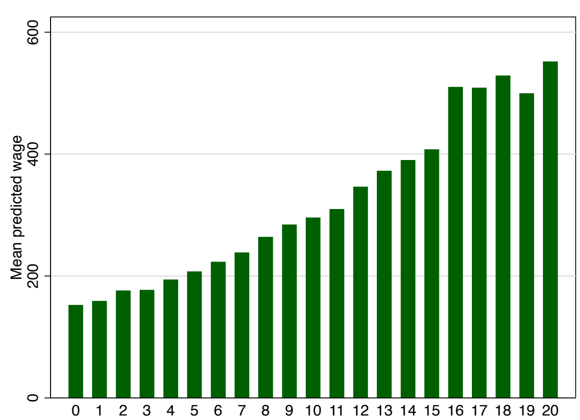

Figure 3: Predicted wage by education level

Table 1: OLS results using year/place of birth fixed effects and education dummies

	logwage
educ1	0.0574
	(0.0920)
educ2	0.173
	(0.0606)
educ3	0.182
	(0.0533)
educ4	0.275
	(0.0529)
educ5	0.342
	(0.0501)
educ6	0.404
	(0.0486)
educ7	0.459
	(0.0481)
educ8	0.534
	(0.0475)
educ9	0.609
	(0.0476)
educ10	0.641
	(0.0475)
educ11	0.689
	(0.0475)
educ12	0.790
	(0.0472)
educ13	0.862
	(0.0474)
educ14	0.905
	(0.0474)
educ15	0.952
	(0.0476)
educ16	1.172
	(0.0473)
educ17	1.166
	(0.0475)
educ18	1.203
	(0.0475)
educ19	1.146
	(0.0481)
educ20	1.243
	(0.0477)
_cons	4.962
	(0.0478)
N	329509

Table 2: IV resultst

	(1) iv1	(2) iv2	(3) iv3	(4) iv4
educ	0.103	0.0677	0.108	0.0446
	(0.0195)	(0.0706)	(0.0196)	(0.0773)
highSchool		0.245		0.439
		(0.477)		(0.520)
_cons	4.590	4.846	4.492	4.944
	(0.249)	(0.557)	(0.222)	(0.579)
N	329509	329509	329509	329509

Bayesian Parts

The model is

$$\begin{aligned} \underset{n\times 1}{y} &= \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{\epsilon} \qquad (1)\\ \underset{n\times k}{X} &= \underset{n\times L}{Z}\underset{L\times k}{\gamma} + \underset{n\times k}{\nu} \\ \begin{pmatrix}\epsilon\\ \nu\end{pmatrix} &\sim \mathcal{N}(0,\Sigma_1) \end{aligned}$$

which implies

$$\begin{aligned} y = \underset{n\times L}{Z}\underset{L\times k}{\gamma}\underset{k\times 1}{\beta} + \underset{n\times k}{\nu}\underset{k\times 1}{\beta} + \underset{n\times 1}{\epsilon} \\ \begin{pmatrix}\nu\beta+\epsilon\\ \nu\end{pmatrix} \sim \mathcal{N}(0,\Sigma_2) \end{aligned}$$

In the most simple case, with $k = 1$ (one endogenous regressor), we can reparameterize the model:

$$\begin{aligned} \underset{n\times 2}{\mathcal{Y}} &= \underset{n\times L}{Z}\underset{L\times 2}{\Pi} + \underset{n\times 2}{\eta} \qquad (2)\\ \iff \underset{n\times 2}{\begin{bmatrix} y & x\end{bmatrix}} &= \underset{n\times L}{Z}\underset{L\times 2}{\begin{bmatrix}\gamma\beta & \gamma\end{bmatrix}} + \underset{n\times 2}{\begin{bmatrix}\nu\beta+\epsilon & \nu\end{bmatrix}} \\ \eta &\sim \mathcal{N}(0,\Omega) \qquad (3) \end{aligned}$$

If we impose the restriction that rank(Π) = 1, the model in (2) traces out the same set of probability models for the data as in (1). (This imposes that the elements of Π are linear combinations of each other, since $\gamma\beta$ is a linear combination of γ when ($L \geq k$).)
In the initial parameterization (Model 1), the likelihood is not integrable for a flat prior on β, and so it is useful to describe an uninformative prior over Π in (2)[1]. In fact, a flat prior on Π with no rank restrictions produces an integrable posterior if the sample size is not too small. Hence, we can construct a "flat" prior for (1) by transforming a flat prior over the reduced-rank submanifold of Π into (β, γ) space. The improper

[1]See the argument in Sims (2007)

prior on (β, γ) that emerges from this approach is (assuming $L = k = 1$):

$$\begin{aligned}
\left| \frac{\partial \Pi}{\partial(\beta,\gamma)} \left(\frac{\partial \Pi}{\partial(\beta,\gamma)} \right)' \right|^{\frac{1}{2}} &= \left| \begin{bmatrix} \gamma & 0 \\ \beta & 1 \end{bmatrix} \begin{bmatrix} \gamma' & \beta' \\ 0' & 1' \end{bmatrix} \right|^{\frac{1}{2}} \\
&= \begin{vmatrix} \gamma\gamma' & \gamma\beta' + \gamma 1' \\ \beta\gamma' & \beta\beta' + 11' \end{vmatrix}^{\frac{1}{2}} \\
&= (\gamma\gamma')(\beta\beta' + 11') - (\gamma\beta' + \gamma 1')(\beta\gamma') \\
&= ||\gamma||(1+\beta^2)^{\frac{1}{2}}
\end{aligned}$$

LIML

Question 3 requires estimating model (2) with $k = 1$, $L = 3$ with LIML. The likelihood function of the data is:

$$\ell(\beta, \gamma, \Omega) \propto \sum_{i=1}^{N} \left(-\frac{1}{2} \ln |\Omega| - \frac{1}{2} \begin{pmatrix} y_i - \beta x_i & x_i - Z_i\gamma \end{pmatrix} \Omega^{-1} \begin{pmatrix} y_i - \beta x_i \\ x_i - Z_i\gamma \end{pmatrix} \right)$$

Since y_i, x_i are scalar, the likelihood can be written as:

$$-\frac{N}{2} \ln |\Omega| - \frac{1}{2|\Omega|} \sum_{i=1}^{N} \Big((y_i - \beta x_i)^2 \omega_{22} - 2(y_i - \beta x_i)(x_i - Z_i\gamma)\omega_{12} + (x_i - Z_i\gamma)^2 \omega_{11} \Big)$$

$$-\frac{N}{2} \ln |\Omega| - \frac{1}{2|\Omega|} \left(\omega_{22} \sum_{i=1}^{N} (y_i - \beta x_i)^2 - 2\omega_{12} \sum_{i=1}^{N} (y_i - \beta x_i)(x_i - Z_i\gamma) + \omega_{11} \sum_{i=1}^{N} (x_i - Z_i\gamma)^2 \right)$$

where $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$.

In higher dimensions, this inner part corresponds with the cross product/second moment of the data, so that we can optimize code by passing the second moment as an argument.

The likelihood is combined with an improper flat prior:

$$\pi(\beta, \gamma) \propto \sum_{i=1}^{k} \log(\text{diag}(d)) \equiv f(\gamma)$$

where d is defined according to the singular value decomposition of $\gamma = udv'$. When $L = 1$, as in part 3, this is equal to $||\gamma||$ same as we derived in the section above.