Bayesian Linear Regression Applied to Wage Data

Senan Hogan-H.

MATH 153: Bayesian Statistics

May 2018

1 Mincer Wage Equation

The distribution of income, and thus inequality of income, is regularly explained in labour economics by a frequentistic linear regression of the following form:

$$Y_i = Y_0 + \rho s_i + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \tag{1}$$

Y_i represents the log of a measure of income for an individual , s_i years of education, and Y_0 the standard intercept. ρ, β_1, β_2 are coefficients to be estimated with residual ε_i. Potential experience, x_{it}, is defined as age minus years of education minus 6, i.e. $x_{it} = Age_{it} - s_{it} - 6$. The equation may also include a dummy variable for race, gender, and other variables to control for these differences.

This regression approach has been names the Mincer wage equation or earnings function, which dates back to some of the first studies that focus on wage inequality [1, 2]. This model is extremely influential in labour economics to describe and predict inequality in wages in the US population. Its influence comes in part from its theoretical foundations and simplicity in interpretation, yet is documented as being only accurate in predicting wages[1] for the 1950s, and less so after. The approach is classic in the frequentist, econometrics paradiagm and is still (often egregiously) used in economic research today for predictive purposes. Perhaps there is an alternative approach, that can better explain observed inequality in the income distribution and its relationship to income and labour market experience.

[1] Where the equation may estimated independently for different years.

2 Choice of Prior

3 Derivation: Better prior

4 Application with Data, March CPS

5 Appendix: R code

```
Read 0.0% of 2185520 rows
Read 10.1% of 2185520 rows
Read 20.6% of 2185520 rows
Read 30.2% of 2185520 rows
Read 39.3% of 2185520 rows
Read 48.5% of 2185520 rows
Read 57.7% of 2185520 rows
Read 66.8% of 2185520 rows
Read 75.5% of 2185520 rows
Read 84.6% of 2185520 rows
Read 93.8% of 2185520 rows
Read 2185520 rows and 22 (of 22) columns from 0.204 GB file in 00:00:18
```

References

[1] Jacob Mincer. Investment in human capital and personal income distribution. *Journal of political economy*, 66(4):281–302, 1958.

[2] Jacob A Mincer. Schooling and earnings. In *Schooling, experience, and earnings*, pages 41–63. NBER, 1974.

Table 1: Mincer Equation Results

	Dependent variable:	
	1980-1985	
	(1)	(2)
Years education	0.090^{***}	0.114^{***}
	(0.0004)	(0.0004)
Potential experience	0.055^{***}	0.054^{***}
	(0.0003)	(0.0003)
(Potential experience)^2	-0.001^{***}	-0.001^{***}
	(0.00001)	(0.00001)
Gender	-0.543^{***}	-0.373^{***}
	(0.002)	(0.002)
Race – black	-0.097^{***}	-0.124^{***}
	(0.004)	(0.003)
Race – Hispanic	-0.080^{***}	-0.075^{***}
	(0.004)	(0.003)
Constant	8.953^{***}	8.585^{***}
	(0.007)	(0.006)
Observations	329,754	474,776
R^2	0.348	0.301
Adjusted R^2	0.348	0.301
Note:	*p<0.1; **p<0.05; ***p<0.01	