Data Analysis Project

Lukas

2021-01-11

Abstract

In this report, I replicate the results of a paper analysing the effect of receiving a job training grant on the scrap rate in manufacturing firms, using a Bayesian hierarchical model instead of a frequentist approach. Choosing noninformative priors, I get similar results when looking at the posterior mean of this effect. However, treating the parameters of the model as random variables allows me to make probabilistic statements about the size of the benefit of the job training grant on worker productivity.

1 Introduction

In this report, I analyse the effect of receiving a job training grant for manufacturing firms in Michigan during 1988 and 1989 on worker productivity. The analysis is based on a paper written by Holzer et al. (1993), who originally collected the data for and analysed this panel, and on examples used in the textbook "Introductory Econometrics: A Modern Approach" by Wooldridge (2013), who extended the analysis with more sophisticated methods.

Different from the frequentist approaches taken by them, I am using a Bayesian model to determine whether I get the same conclusion that receiving such a job training grant reduces the scrap rate as a measure of worker productivity. In particular, I'm setting up a Bayesian hierarchical model, allowing for firm specific effects and a time trend independent of the grant, which I then compare to the frequentist fixed effects model of Wooldridge (2013, 497)

2 Data

The data for this analysis come from a survey conducted by Holzer et al. (1993), which they mailed to the 498 firms that had applied for a job training grant during 1988 and 1989. In total, they received answers from 157 firms, among which 66 had received a grant. The data can be found in the `wooldridge` package using `data("jtrain")`. In my analysis, I will use the six variables shown in Table 1 below.

year	fcode	scrap	grant	sales	employ
1987	410523	0.06	0	8,000,000	70
1988	410523	0.05	0	11,000,000	85
1989	410523	0.05	0	15,000,000	98
1987	418011	10.00	0	4,200,000	70
1988	418011	3.00	1	3,740,000	50
1989	418011	2.00	0	2,040,000	30

Table 1: Data

This table illustrates the structure of the data set based on two firms. The variable *year* can take on values 1987 to 1989, *fcode* is a unique firm code number, *scrap* is number of defective items out of every 100 produced, *grant* takes on value 1 if the firm received a job training grant in that year, *sales* shows the annual sales in USD, and *employ* is the number of employees of the firm in that year. As shown, not all firms who applied for a grant did receive one, while the remaining ones received it in only one of the two years.

After removing missing values for *scrap*, *sales*, and *employ*, there remain 148 observations for 51 firms (for 3 firms, no data is available in year 1987, and for 1 firm, no data is available in years 1987 and 1988). The firms with missing values are smaller, with the mean and median of *sales* and *employ* being roughly 25%-50% lower. However, as long as not reporting values for *scrap*, *sales*, and *employ* is not systematically related to whether a firm received a grant or not, removing these observations does not bias the estimate for the effect of the grant. Given that grants were given on a first-come, first-served basis, this is unlikely to be the case.

Finally, I make the following adjustments to the data. First, because *scrap*, *sales*, and *employ* are very right skewed, I will work with the log of these variables, denoted by *lscrap*, *lsales*, and *lemploy*. And second, I create a lagged *grant* variable to allow for persistent effects of receiving a grant on worker productivity.

Figure 1 below shows how the size of the firm is related to the scrap rate, and how the scrap rate varies over time as well as between firms that did or did not receive a grant. This reveals a positive relation between the number of employees and the scrap rate (in firms with more employees it seems that more items have to be discarded) and the productivity seems to improve over time. However, no immediate effect of the job training grant is visible here, before making use of the panel strucutre that allows us to control for firm specific unobservable characteristics.

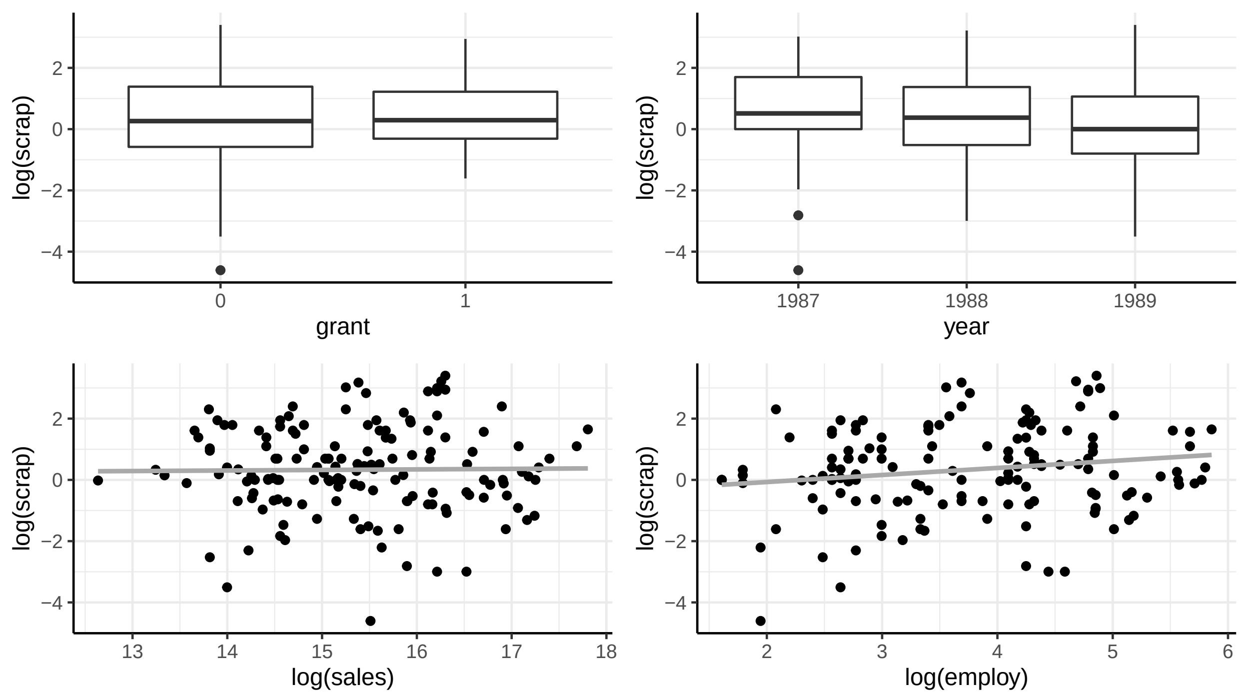

Figure 1: Summary Statistics

3 Model

To analyse the effect of receiving a grant on worker productivity, I estimate the following model:

$$E[lscrap_{i,t}] = fcode_i + year_t + \beta_1 grant_{i,t} + \beta_2 grant_{i,t-1} + \gamma lsales_{i,t-1} + \delta lemploy_{i,t-1} \tag{1}$$

where $i = 1, \ldots, 51$ is the firm index and $t = 1, 2, 3$ the year index. Including $fcode_i$ and $year_t$ allows for different intercepts for each firm and year, that is, for firm-specific effects which are constant over time and for changes over time in the scrap rate independent of the grant. $fcode_i + year_t$ is the intercept for firm i in year t.

I assume a normal likelihood for *lscrap_{i,t}*, where each firm i has it's own mean and a common standard deviation, and for the firm and year indicators.

$$lscrap_{i,t}|\boldsymbol{x_{i,t}}, \beta_1, \beta_2, \gamma, \delta \overset{ind}{\sim} N(\mu_{i,t}, \sigma^2) \tag{2}$$

$$fcode_i \overset{iid}{\sim} N(a, b^2) \tag{3}$$

$$year_t \overset{iid}{\sim} N(c, d^2) \tag{4}$$

where $\boldsymbol{x_{i,t}} = fcode_i, year_t, grant_{i,t}, grant_{i,t-1}, lsales_{i,t-1}, lemploy_{i,t-1}$.

To complete the hierarchical model, I assume the following priors for $fcode_i, year_t, a, b, c, d, \beta_1, \beta_2, \gamma, \delta$ and σ:

$$a \sim N(0, 100),\ b \sim IG(0.5, 0.5),\ c \sim N(0, 100),\ d \sim IG(0.5, 0.5), \tag{5}$$

$$\beta_1 \sim N(0, 100),\ \beta_2 \sim N(0, 100),\ \gamma \sim N(0{,}100),\ \delta \sim N(0, 100),\ \sigma \sim IG(0.5, 0.5) \tag{6}$$

Here, I assume a prior sample size of 1 and a prior guess for the variance of 1 when parametrising the inverse gamma priors for the variance of *lscrap*, *fcode*, and *year*. This will let the data dominate the posterior. The priors for a and c imply the assumption that on average the *year* and *fcode* have no effect on the scrap rate. Finally, I also center the remaining priors around 0, i.e. being "skeptical" about the effect of receiving a grant and assuming that firm size should not have a particular effect on the scrap rate.

4 Results

The frequentist estimation of (1) by ordinary least squares produces the following results, where the standard errors of the coefficients are reported in parentheses:

Table 2: Frequentist Analysis

	Dependent variable:
	lscrap
grant	-0.297^{*}
	(0.157)
grant_1	-0.536^{**}
	(0.224)
lsales	-0.087
	(0.260)
lemploy	-0.076
	(0.350)
firm fixed effects	Yes
year fixed effects	Yes
Observations	148
R^2	0.067
Adjusted R^2	-0.508
Note:	*p<0.1; **p<0.05; ***p<0.01

This suggests that obtaining a grant in year 1988 lowers the scrap rate by about $1 - e^{-0.297} = 25.7\%$ and, interestingly, the lagged effect of receiving a grant in year $t-1$ on the scrap rate in year t is even larger ($1 - e^{-0.536} = 41.5\%$). Not shown here is that the year fixed effects both are negative. So, the scrap rate on average declined for all firms, including the ones which did not receive a grant, highlighting the importance to include the year fixed effects.

Next, I fit the Bayesian hierarchical model, using 1,000 burn-in iterations followed by 20,000 iterations that I monitor. While convergence looks okay for the two main parameters of interest, β_1 and β_2, in particular a and c (the means for the firm and year intercepts) show strong autocorrelation. Nonetheless, inference for the effect of receiving a grant will be accurate.

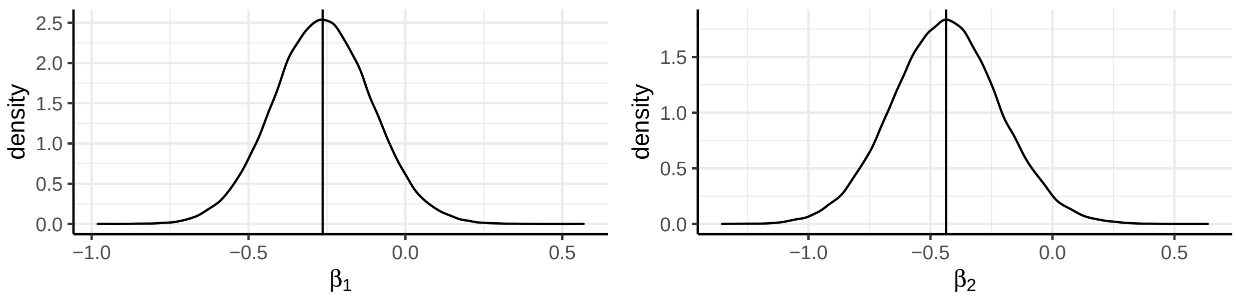

Figure 2: Posterior Densities

Hence, with the chosen priors, the results are similar to the frequentist analysis. Receiving a grant seems to strongly reduce the scrap rate, with the effect in the year after being even stronger. Both parameters are slightly smaller, with posterior means of -0.26 for β_1 and -0.44 for β_2, most likely due to the prior which was centered at zero.

Although the posterior means are similar to what we've seen in Table 2, it's now further possible to say that the probability that receiving a grant reduces the scrap rate $P(\beta_1 < 0) = 0.95\%$.

5 Conclusions

To conclude, by choosing relatively noninformative priors, I arrived at similar results for the effect of receiving a job training grant on the scarp rate of a manufacturing firm. However, the Bayesian analysis bring the advantage of taking better into account the uncertainty related to the model and parameter values.

References

Holzer, Harry J., Richard N. Block, Marcus Cheatham, and Jack H. Knott. 1993. "Are Training Subsidies for Firms Effective? The Michigan Experience." *Industrial and Labor Relations Review* 46 (4): 625–36.

Wooldridge, Jeffrey M. 2013. *Introductory Econometrics: A Modern Approach, 5th Edition.* South-Western: Cengage Learning.