Tobit model application to extramarital affairs

Nikolaj Nasenko

Introduction

In this paper, we examine extramarital affairs data from the R package AER. Our objective is to understand what creates an incentive for a person to engage in an extramarital sexual intercourse. We estimate the effect of various variables such as gender or length of marriage on how often a person has sexual intercourse outside marriage. As the number of affairs can be thought of as left-censored data, there is plenty of people not having an affair, the Tobit model is used.

Theoretical background

The extramarital affairs data contain information on gender and age. Information on how long the person has been married, whether he or she has children, his or her religion, education, occupation, and self-assessed satisfaction with the marriage. Our expectations are as follows:

H1: Men have an affair more often than women.
H2: Younger people are more frequently having an affair.
H3: The longer the marriage, the more often a person has an affair.
H4: Couples with children are less frequently having an affair.
H5: Religious people are less frequently having an affair.
H6: People who are less satisfied with their marriage are more often having an affair.
We have no clear expectations regarding the effect of education and occupation.

Data and methodology

We use cross-sectional extramarital affairs data from the R package AER from a survey conducted by Psychology Today in 1969. Our dataset includes 601 observations and 9 variables. Each observation represents a person who was married and was married for the first time. Those who were married more than once were excluded because of lack of information on some variables. The survey is also not likely a random sample of the U.S. population (Fair, 1978).

Our dependent variable measures how often the respondent engaged in extramarital sexual intercourse during the past year. There are 451 individuals (75 % of respondents) that had not had an affair. Neither of the variables contain missing values or appear to be highly correlated with each other. The list of all variables considered and their corresponding distributions are presented in Table 1 and Figure 1, respectively.

Figure 1. shows that the majority of respondents are between 20 to 40 years old, have at least some college, with occupation within 1st, 5th or 6th Hollingshead class, believers, have been married for years and are happy with their marriage.

We previously mentioned that the 75 % of respondents have no affairs (our dependent variable), thus we can think of the number of affairs as a left-censored data. Therefore, we can use the Tobit model for estimation.

Table 1: The list of variables

Variable	Description	Mean
affairs	How often engaged in extramarital sexual intercourse during the past year? (discrete) 0 = none, 1 = once, 2 = twice, 3 = 3 times, 7 = 4–10 times, 12 = monthly, weekly, daily	1.46
gender	male or female (dummy)	0.48
age	age in years (discrete) 17.5 = under 20, 22 = 20–24, 27 = 25–29, 32 = 30–34, 37 = 35–39, 42 = 40–44, 47 = 45–49, 52 = 50–54, 57 = 55 or over	32.49
yearsmarried	number of years married (discrete) 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years	8.18
children	Are there children in the marriage? (dummy) yes or no	0.72
religiousness	How religious? (discrete) 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very	3.12
education	level of education (discrete) 9 = grade school, 12 = high school graduate, 14 = some college, 16 = college graduate, 17 = some graduate work, 18 = master's degree, 20 = Ph.D., M.D., or other advanced degree	16.17
occupation	occupation according to Hollingshead classification (reverse numbering) (discrete)	4.19
rating	marriage satisfaction self-assessment (discrete) 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, 5 = very happy	3.93

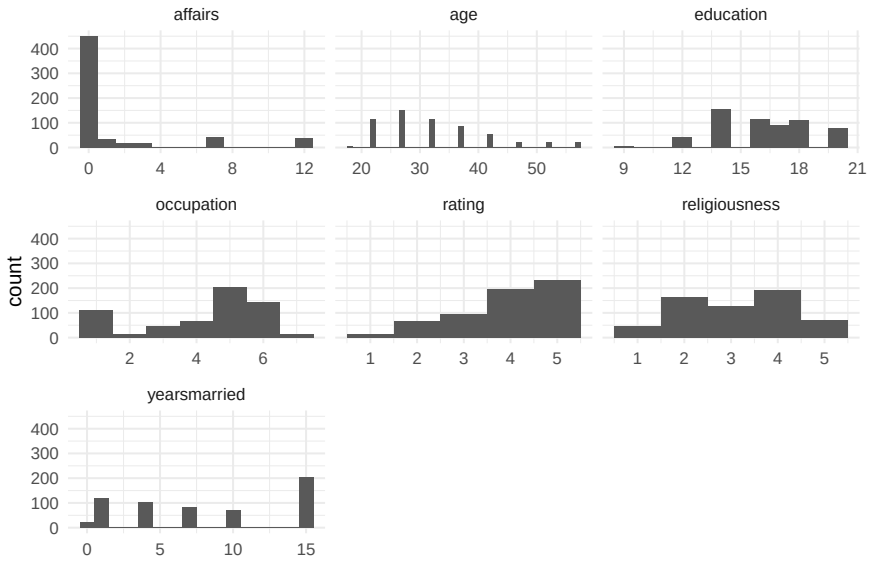

Figure 1: Histograms (binwidth = 1)

Results

Table. 2 presents the estimated marginal effects at the mean, where the reference person is a childless woman. The results suggest that the frequency at which a person has an affair is linked to age, length of marriage, religiousness and self-assessed marital satisfaction. As per our expectations, younger people engage in extramarital sexual intercourse more frequently. The same holds for more religious people and people satisfied with their marriage. Furthermore, as we assumed, the results show that the longer a person is married, the more frequently the person has an affair.

We found no statistically significant relationship for gender, presence of children, education or occupation (assuming the 5 % significance level). Although gender and presence of children turned out to be insignificant,

the 95 % confidence intervals indicate that their influence may be relatively large and positive from time to time (Figure 2.).

Table 2: Marginal effects at the mean (MEM)

	MEM
gendermale	0.220 (0.246)
age	−0.045 (0.019)*
yearsmarried	0.124 (0.034)***
childrenyes	0.237 (0.296)
religiousness	−0.395 (0.093)***
education	0.006 (0.053)
occupation	0.049 (0.075)
rating	−0.528 (0.096)***
Num. obs.	601

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

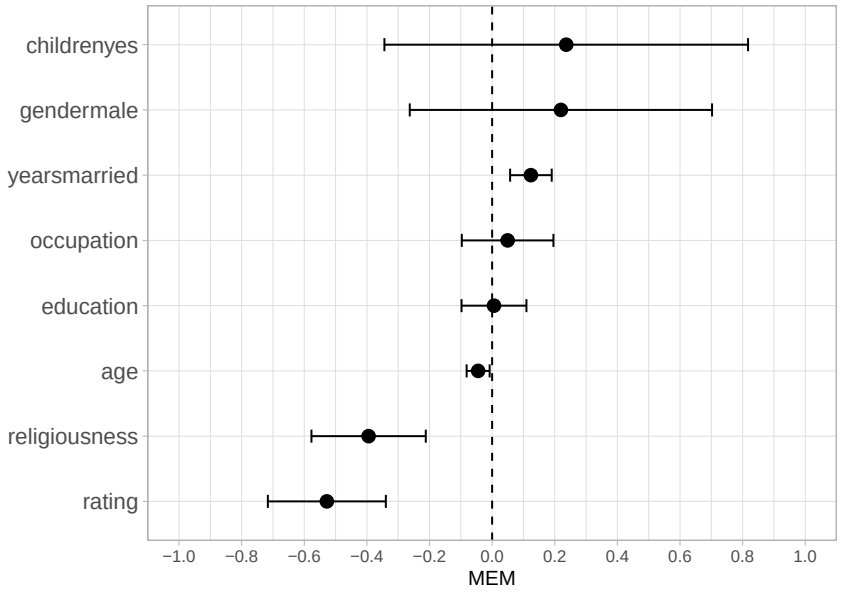

Figure 2: MEM 95 % confidence intervals

Conclusion

In this paper, we investigated what personal characteristics might be related to a person having an affair. We concluded that the frequency with which a person has an affair is linked to age, length of marriage, religiousness, self-assessed marital satisfaction, and occasionally gender and presence of children. Younger people engage more often in affairs. The more religious a person is and the more satisfied he or she is with the marriage, the less frequently he or she has an affair. The effect of gender and the effect of having children was estimated positive, but with high uncertainty regarding the magnitude and direction of the relationship.

References

Fair, R.C. (1978). *A Theory of Extramarital Affairs. Journal of Political Economy*, 86, 45–61. [cited 2022-05-07]. https://fairmodel.econ.yale.edu/rayfair/pdf/1978a200.pdf

Henningsen, A. (2017). *Estimating Censored Regression Models in R using the censReg Package.* [cited 2022-07-05]. https://mran.revolutionanalytics.com/snapshot/2016-01-01/web/packages/censReg/vignettes/censReg.pdf

R

```
library(AER)
library(skimr)
library(tidyverse)
library(knitr)
library(kableExtra)
library(censReg)
library(texreg)

# load data
data("Affairs")
str(Affairs)

### Variables table
------------------------------------------------------------

# recode factors and calculate means
means <- Affairs %>%
  mutate(gender = ifelse(gender=="male",1,0), children = ifelse(children=="yes",1,0)) %>%
  summarise_all(mean)

# list of variables table
table1 <- tibble(
  Variable = names(Affairs),
  Description = c(
    "How often engaged in extramarital sexual intercourse during the past year? (discrete)\n0 = none, 1
    "male or female (dummy)",
    "age in years (discrete)\n17.5 = under 20, 22 = 20-24, 27 = 25-29, 32 = 30-34, 37 = 35-39, 42 = 40-4
    "number of years married (discrete)\n0.125 = 3 months or less, 0.417 = 4-6 months, 0.75 = 6 months-1
    "Are there children in the marriage? (dummy)\nyes or no",
    "How religious? (discrete)\n1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very",
    "level of education (discrete)\n9 = grade school, 12 = high school graduate, 14 = some college, 16 =
    "occupation according to Hollingshead classification (reverse numbering) (discrete)",
    "marriage satisfaction self-assessment (discrete)\n1 = very unhappy, 2 = somewhat unhappy, 3 = avera
  Mean = means %>% round(2) %>% as.numeric())

table1$Description <- linebreak(table1$Description, align = "l")

kable(table1, escape = FALSE, booktabs = TRUE, caption = "The list of variables") %>%
  column_spec(2,"13cm") %>%
  kable_styling(font_size = 7)


### Histograms
------------------------------------------------------------

# Histograms
Affairs %>%
  select_if(is.numeric) %>%
  pivot_longer(everything(), names_to = "variable", values_to = "value") %>%
  ggplot(aes(value)) +
  geom_histogram(binwidth = 1) +
```

```
  facet_wrap(~variable, scales = "free_x") +
  xlab("") +
  theme_minimal()


### Descriptive statistics (dependent variable):
------------------------------------------------------------

# affairs no. missing values
skim(Affairs) # no missing values

# affairs histogram
ggplot(Affairs, aes(affairs)) + geom_histogram(binwidth = 1) + ggtitle("Histogram (binwidth=1)")

# number of zeroes
table(Affairs$affairs, useNA = "always") # 451 zeroes
sum(Affairs$affairs == 0)/length(Affairs$affairs) # 75 % zeroes


### Descriptive statistics (independent variables):
------------------------------------------------------------

skim(Affairs) # no missing values, mostly non-normally distributed variables
Affairs %>% select(-affairs, -gender, -children) %>% pairs() # no obvious correlation


### Tobit model
------------------------------------------------------------

# tobit
mod <- censReg(affairs ~ ., data = Affairs)

# marginal effects at the mean
mem <- margEff(mod) %>% summary()


### Regression output table
------------------------------------------------------------

# censReg package is unsupported by texreg
tr <- createTexreg(coef.names = rownames(mem),
                   coef = mem[,1],
                   se = mem[,2],
                   pvalues = mem[,4],
                   gof.names = "Num. obs.",
                   gof = nobs(mod),
                   gof.decimal = FALSE)

# table
texreg(tr, custom.model.names = "MEM",
       digits = 3, single.row = TRUE, center = TRUE,
       caption = "Marginal effects at mean (MEM)",
       caption.above = TRUE, float.pos = "h")
```

```r
# !!! r chunk options must include results="asis" !!!
# Otherwise there will be just a raw output in PDF


### Coefficients plot
-----------------------------------------------------------------

# confidence intervals
beta.ci <- data.frame(MEM = mem[,1], SE = mem[,2])
beta.ci <- beta.ci %>% mutate(lower = MEM - 1.96*SE, upper = MEM + 1.96*SE)

# coefficients plot
ggplot(beta.ci) +
  geom_vline(xintercept = 0, linetype = 2) +
  geom_point(aes(MEM, reorder(rownames(mem), MEM)), size = 3) +
  geom_errorbar(aes(x = MEM, y = rownames(mem), xmin = lower, xmax = upper), width = 0.2) +
  labs(y="") +
  scale_x_continuous(breaks = seq(-10, 10, 0.2), limits = c(-1,1)) +
  theme_light() +
  theme(axis.text.y  = element_text(size = 12))

# !!! out.width="70%", fig.cap="MEM 95 \\% confidence intervals" !!!
# Percentage must be written as \\%
```