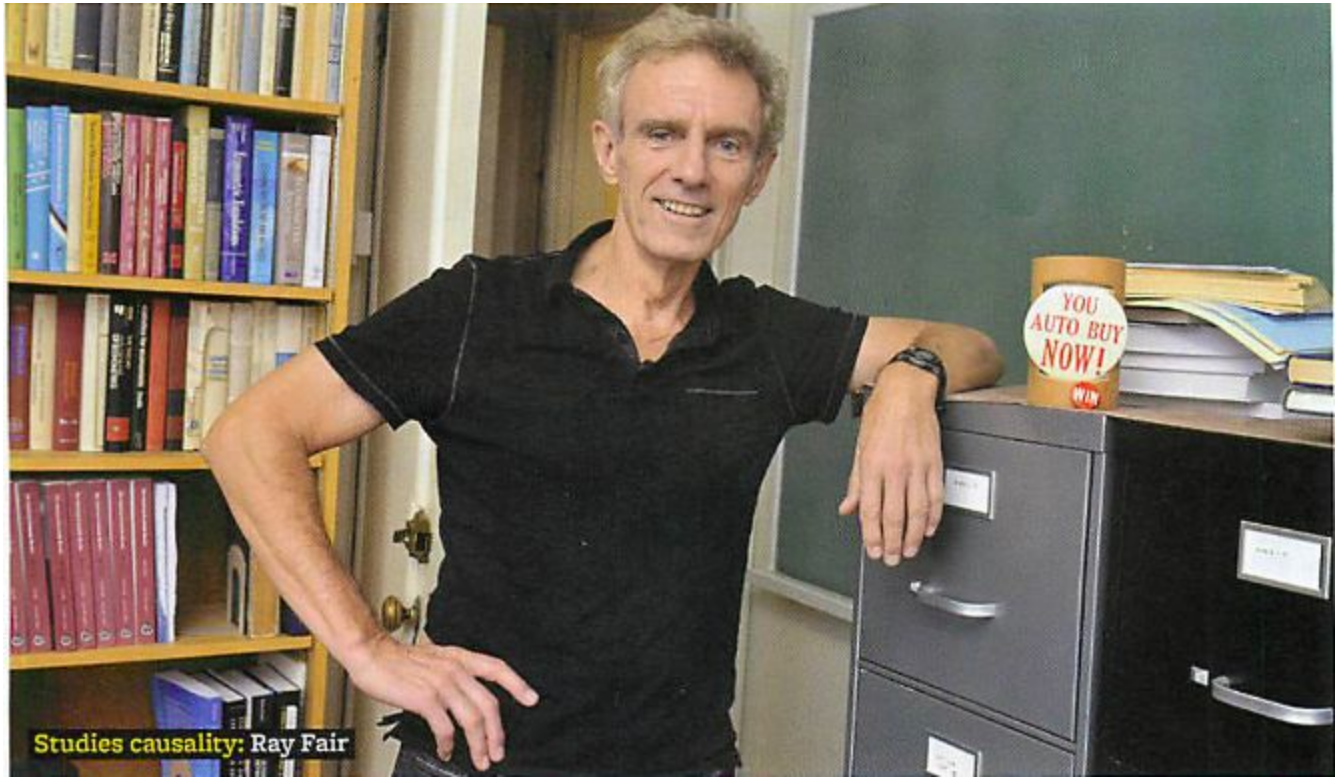
“A Theory of Extramarital Affairs”

Fair, 1978

Why this article?

- Ray C. Fair is a macroeconomist at Yale; this article sticks out on his CV.

- It’s become a standard dataset in economics for illustrating the Tobit estimator.



Studies causality: Ray Fair

Description of the Dataset

- The data comes from a survey conducted by the magazine *Psychology Today* (*PT*) in July, 1969. Responder bias?

- The outcome variable is how often the respondent engaged in extramarital affairs over the previous year:

0 = none, 1 = once, 2 = twice, 3 = 3 times, 7 = 4–10 times, 12 = monthly, 12 = weekly, 12 = daily

  Note that it's a kind of weird count-ordinal hybrid (Fair's analysis is mostly qualitative anyway).

- The explanatory variables are: sex, age, how long the respondent has been married, how religious they are, their level of education, their occupation, and how they would subjectively rate their satisfaction with their marriage. Predictions?

Fair's Estimation Strategy

- "... because many values of y [...] are zero, it would clearly be incorrect to use ordinary least squares to estimate the equations. The obvious technique to use in this case is the Tobit estimator" (pg.57). There is also a little top-censoring: all frequencies equal to or greater than monthly are coded as 12. It seems Fair doesn't acknowledge this (see below).

- Recall Tobin's original problem: large numbers of households spent exactly $0 on luxury goods; he wanted to distinguish the effect of the explanatory variables on people's choosing to buy *any* luxury goods (selection) from the effect of the explanatory variables on *how much* luxury goods someone buys, given that they buy any.

- In this case, 451 of the 601 respondents have a response variable of 0 (no extramarital affairs in the last year). The problem is structurally very similar to Tobin's: how do the explanatory variables affect a person's *selecting* a nonzero response variable, and how do they affect the *level* of the response variable, given that it's nonzero?

Fair's Results

* α = 0.1
** α = 0.05
*** α = 0.01

Variable		Coeff. Est.	*t*-Stat.	Coeff. Est.	*t*-Stat.
Constant	...	7.60	1.92 **	8.17	2.96 **
Occupation	z_7	.213	.67	.326	1.29 *
Education	z_6	.0252	.11	...	...
Husband's occupation	...	...	...	...	...
Marital happiness	z_8	−2.27	−5.48 ***	−2.28	−5.61 ***
Age	z_2	−.193	−2.37 ***	−.179	−2.26 **
No. years married	z_3	.533	3.63 ***	.554	4.13 ***
Children	z_4	1.02	.79	...	...
Degree of religiosity	z_5	−1.70	−4.15 ***	−1.69	−4.14 ***
Sex	z_1	.945	.88	...	...
	$\hat{\sigma}^2_{PT}$	8.26	...	8.25	...
No. observations		601‡		601‡	

There are two sets of coefficient estimates: if you look at the second column, you see that it drops some of the variables from the first model.

Replication in R

```
library(AER)
model <- tobit(y ~ sex + age + years_married + children + how_religious +
                 education + occupation + rate_marriage, data = data,
               left = 0)
summary(model)
```

Replication in R

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	7.60849	3.90599	1.948	0.051426	.
sex	0.94579	1.06287	0.890	0.373548	
age	-0.19270	0.08097	-2.380	0.017316	*
years_married	0.53319	0.14661	3.637	0.000276	***
children	1.01918	1.27957	0.797	0.425741	
how_religious	-1.69900	0.40548	-4.190	2.79e-05	***
education	0.02536	0.22767	0.111	0.911304	
occupation	0.21298	0.32116	0.663	0.507220	
rate_marriage	-2.27328	0.41541	-5.472	4.44e-08	***

Other than rounding, the point estimates are identical to Fair's (thankfully). Because the degrees of freedom is so high, R uses a normal approximation rather than a t-distribution: this has no effect on the statistical significance (R's star notation is a bit different from mine).

New Contribution: Heckman Model

Fair seems not to have been aware of Heckman selection models (Heckman's first paper on the subject was about two years earlier). In fact, Heckman models are perhaps more useful for the purpose discussed earlier, since they have a distinct *selection equation* and *outcome equation*.

This allows us to move towards a *theory* of censoring / selection.

New Contribution: Heckman Model

```
### Estimate a Heckman model.
library(sampleSelection)

# We have to define a binary outcome variable for the selection equation.

thing <- numeric(601)
for (i in 1:length(data$y)) {
  if (data$y[i] > 0) {
    thing[i] <- 1
  }
  if (data$y[i] == 0) {
    thing[i] <- 0
  }
}

# Then estimate the model.

model2 <- heckit(selection = thing ~ sex + age + years_married + children + how_religious +
                   education + occupation + rate_marriage,
                 outcome = y ~ sex + age + years_married + children + how_religious +
                   education + occupation + rate_marriage,
                 data = data)
summary(model2)
```

New Contribution: Heckman Model

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.77940    0.51255   1.521 0.128895
sex            0.17346    0.13799   1.257 0.209253
age           -0.02458    0.01042  -2.360 0.018612 *
years_married  0.05434    0.01881   2.889 0.004005 **
children       0.21664    0.16517   1.312 0.190154
how_religious -0.18547    0.05163  -3.593 0.000355 ***
education      0.01126    0.02952   0.382 0.702930
occupation     0.01367    0.04140   0.330 0.741421
rate_marriage -0.27179    0.05347  -5.083 5.03e-07 ***
Outcome equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.57504   11.15810   0.410    0.682
sex            2.65669    4.55357   0.583    0.560
age           -0.51448    0.57499  -0.895    0.371
years_married  1.35905    1.21799   1.116    0.265
children       3.49256    6.06633   0.576    0.565
how_religious -4.42642    4.13525  -1.070    0.285
education      0.04795    0.63916   0.075    0.940
occupation     0.56871    0.87044   0.653    0.514
rate_marriage -5.90339    5.63577  -1.047    0.295
Multiple R-Squared:0.1876,      Adjusted R-Squared:0.1353
   Error terms:
              Estimate Std. Error t value Pr(>|t|)
invMillsRatio   27.089     27.318   0.992    0.322
sigma           23.261         NA      NA       NA
rho              1.165         NA      NA       NA
```

New Contribution: Heckman Model

Takeaways:

- The same variables are estimated to have a significant effect on *selection* as in Fair's original analysis: age, years married, religiosity, and self-described marital satisfaction.
- As with the analysis of the Mroz data in the lecture: *none* of these variables is significant in the outcome equation. This suggests that these features influence *whether* to engage in an extramarital affair: not the *extent* of such behaviour conditional upon it being nonzero.
- The inverse Mills ratio (lamda) is not statistically significantly different from zero: we can't reject the null hypothesis that there is no selection bias effect.

The End.