

ECN 3972

Lab 7

Download “TRACTOR.xls” and load it into R. Also load mosaic.

In R:

```
library(mosaic)
```

Let’s look at the variables in our data:

$price_i$ = price of tractor i (in dollars)

HP_i = horsepower of tractor i

age_i = age of tractor i (in years)

$hours_i$ = engine hours of use for tractor i

$diesel_i$ = 1 if tractor i uses diesel fuel; = 0 otherwise

fwd_i = 1 if tractor i has four-wheel drive; = 0 otherwise

$manual_i$ = 1 if tractor i has a manual transmission; = 0 otherwise

$deere_i$ = 1 if tractor i is manufactured by John Deere; = 0 otherwise

$enccab_i$ = 1 if tractor i has an enclosed cab; = 0 otherwise

To estimate different functional forms, we need to create natural log versions of two of the continuous variables:

In R:

```
TRACTOR$lnprice = log(TRACTOR$price)
```

```
TRACTOR$lnage = log(TRACTOR$age)
```

Note that, to add a variable to the TRACTOR data set, we have to put “TRACTOR\$” (which is “dataname\$”) before the new variable name. Note also that we have to specify which variable to take logs of by also using the “dataname\$” form.

Now we can try a few functional forms. Let’s start with a double-log model:

$$\ln price_i = \beta_0 + \beta_1 \ln age_i + \varepsilon_i$$

In R:

```
model1 = lm(lnprice~lnage, data = TRACTOR)
```

```
summary(model1)
```

How do we interpret β_1 ?

Let's try two different semi-log models:

$$\ln price_i = \beta_0 + \beta_1 hours_i + \varepsilon_i$$

In R:

```
model2 = lm(lnprice~hours, data=TRACTOR)
summary(model2)
```

How do we interpret β_1 ?

Note: we have estimated these individually, but they can be combined. You can have double-log and semi-log together in the same equation if you have more than one independent variable.

Or the other semi-log:

$$price_i = \beta_0 + \beta_1 \ln age_i + \varepsilon_i$$

In R:

```
model3 = lm(price~lnage, data=TRACTOR)
summary(model3)
```

How do we interpret β_1 ?

Let's think about a simple model...a polynomial:

$$price_i = \beta_0 + \beta_1 Age_i + \beta_2 HP_i + \beta_3 HP_i^2 + \varepsilon_i$$

First, we need to create a variable for HP_i^2 :

In R:

```
TRACTOR$HPsq = (TRACTOR$HP)^2
```

Then, the model:

In R:

```
model4 = lm(price~age + HP + HPsq, data=TRACTOR)
summary(model4)
```

Why would we add HP_i^2 to the model?

How can we determine the true effect of HP_i given the polynomial?

Okay, last model for now.

What if we believe that the effects of Age_i on $Price_i$ may be different for John Deere tractors? We would create an interaction between the variables Age_i and $Deere_i$ as follows:

In R:

```
TRACTOR$deereage = TRACTOR$age * TRACTOR$deere
```

Now, we want to include this in our equation. We will do a model with it and without it.

$$price_i = \beta_0 + \beta_1 Age_i + \beta_2 deere_i + \varepsilon_i$$

$$price_i = \beta_0 + \beta_1 Age_i + \beta_2 deere_i + \beta_3 deereage_i + \varepsilon_i$$

In R:

```
model5 = lm(price~age+deere, data = TRACTOR)
```

```
summary(model5)
```

```
model6 = lm(price~age+deere+deereage, data = TRACTOR)
```

```
summary(model6)
```

How do we interpret β_3 ?