

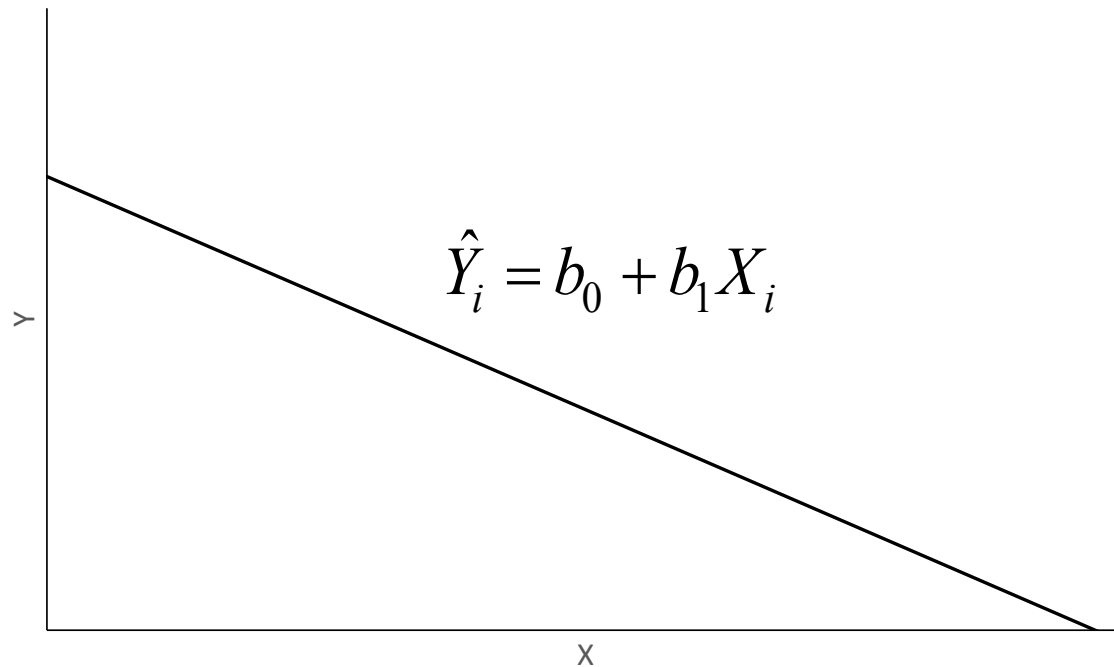
Chapter 7

Predicting in a Nonlinear World

Learning Objectives

- Articulate what it means to “get the model right”
- Learn transformations that can turn a misspecified model into a well- specified model
- Test for nonlinearities in an econometric model

Regression Model Implies **Linear** Relationship Between Y and X



Multiple Regression Model Implies **Linear** Relationship Between Y and X

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_K X_{Ki} + e_i$$

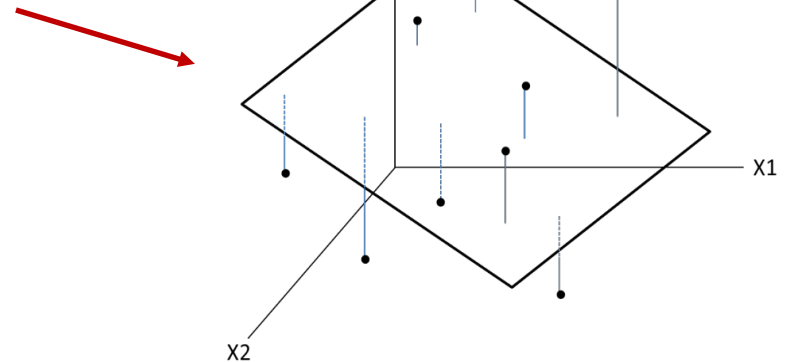


Figure 3.1. Illustration of a regression plane for a model with two explanatory variables, $X1$ and $X2$

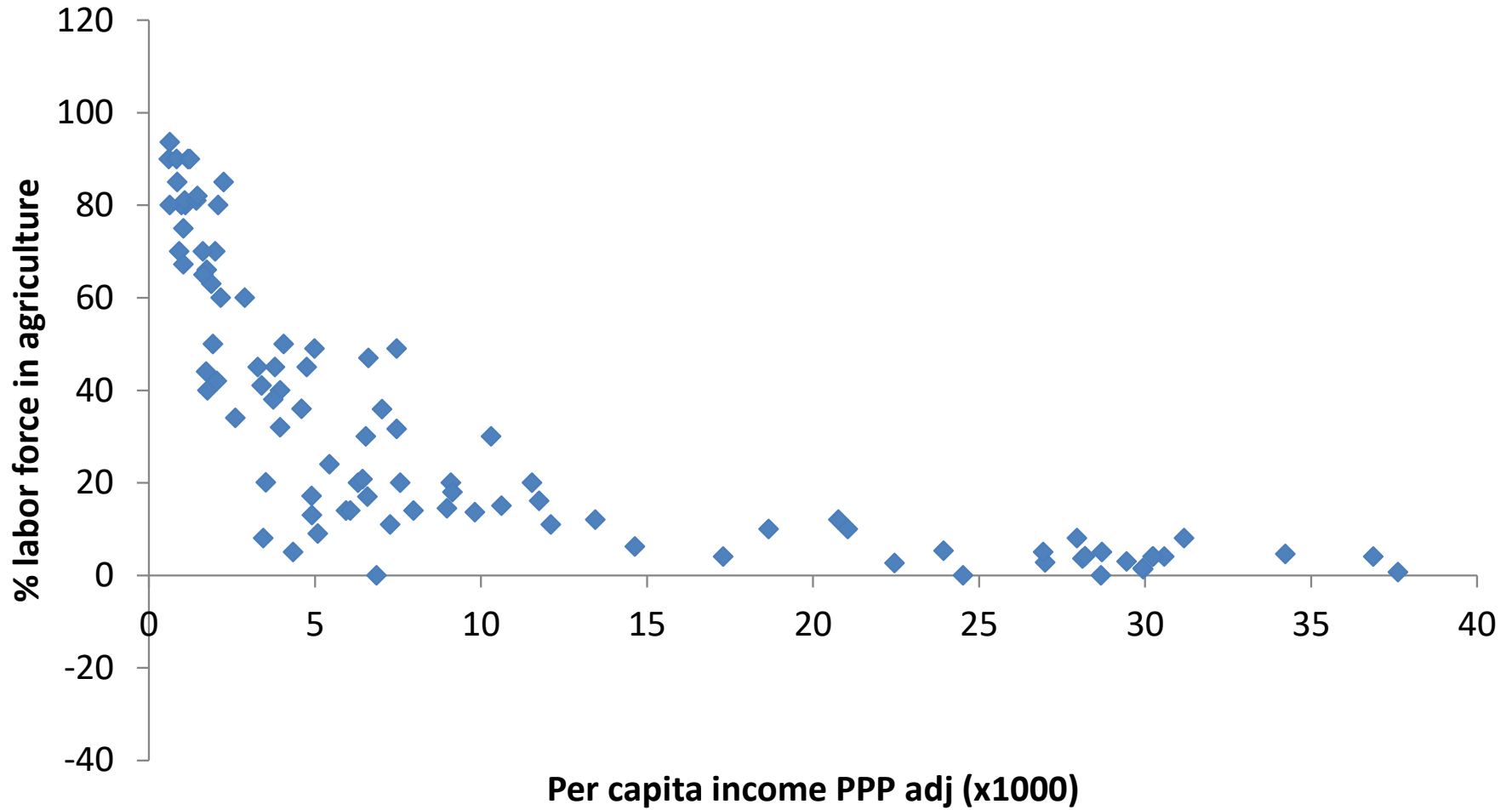
What do we mean by “Linear Relationship”?

- A change in X is associated with the same change in Y no matter what the initial value of X
- **Potential point of confusion**
 - In Ch 5, we discuss the linearity of the OLS estimator – **this is different**

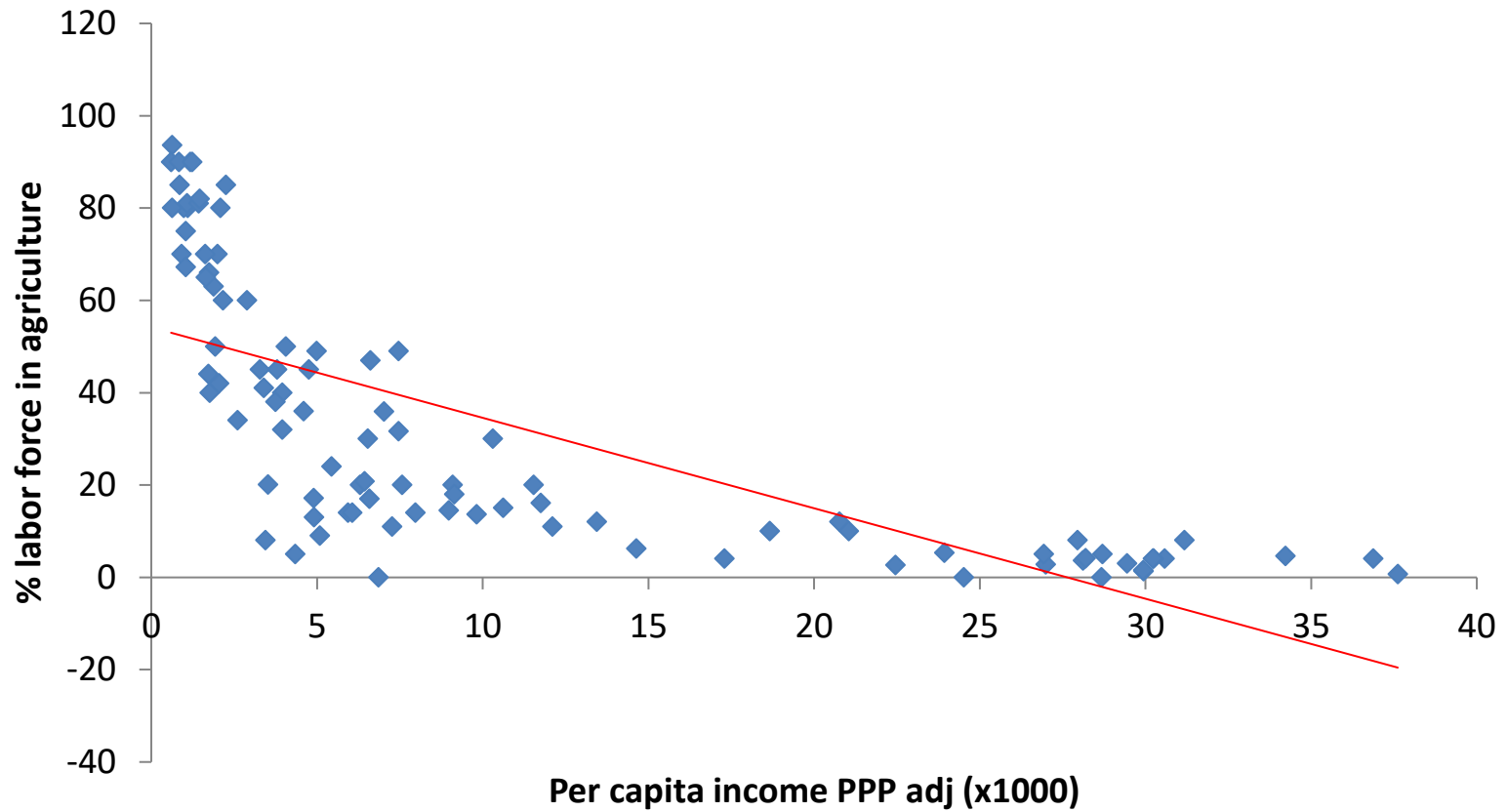
- **Linear model:**
$$Y_i = b_0 + b_1 X_i + e_i$$

- **Linear estimator:**
$$b_1 = \sum_{i=1}^N w_i Y_i$$

Moving off the Farm



Moving off the Farm



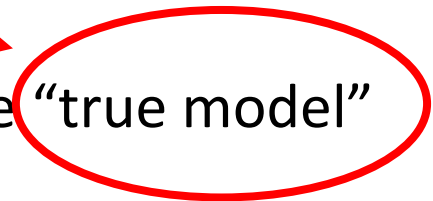
Getting the Model Right

- We aim to represent the relationship between Y and the X 's *reasonably well*
- In old econometrics books, the goal is to find the “true model”
- Remember Step 1 in Chapter 1
 - “What do you want to do?”
 - If you start with Step 1, you will have a good idea of what *“reasonably well”* means

Getting the Model Right

- We aim to represent the relationship between Y and the X's *reasonably well*
- In old econometrics, the goal was to find the "true model"
- Remember Step 1 . Chapter 1
 - "What do you want to do?"
 - If you start with Step 1, you will have a good idea of what *"reasonably well"* means

**THERE IS NO
SUCH THING!**



Getting the Model Right

- We aim to represent the relationship between Y and the X 's *reasonably well*
- In old econometrics books, the goal is to find the “true model”
- Remember Step 1 in Chapter 1
 - “What do you want to do?”
 - If you start with Step 1, you will have a good idea of what *“reasonably well”* means

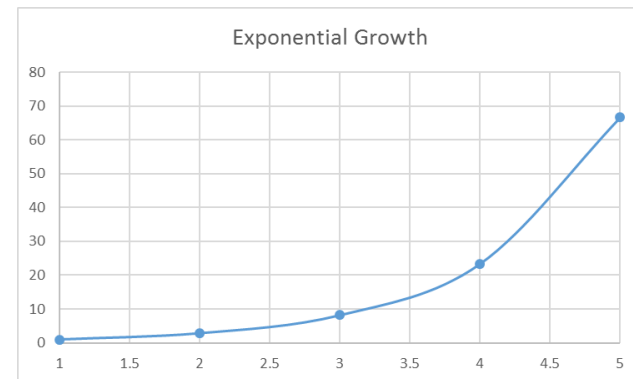
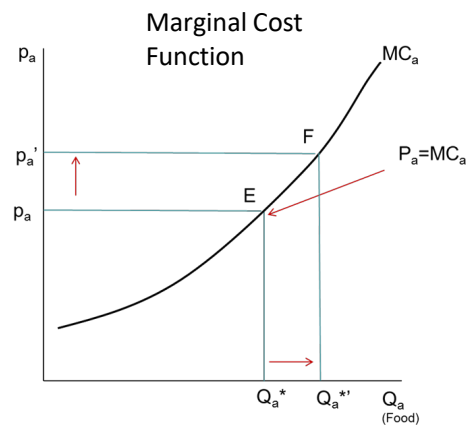
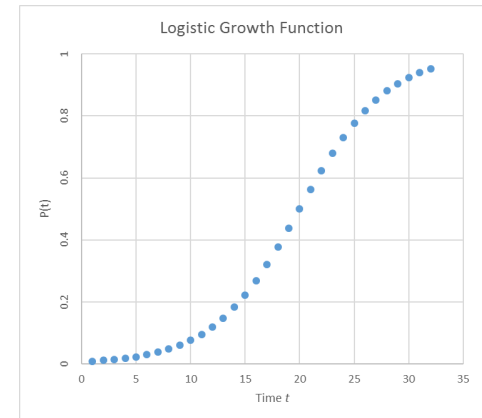
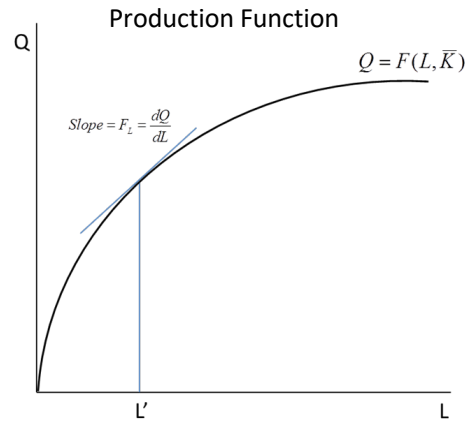
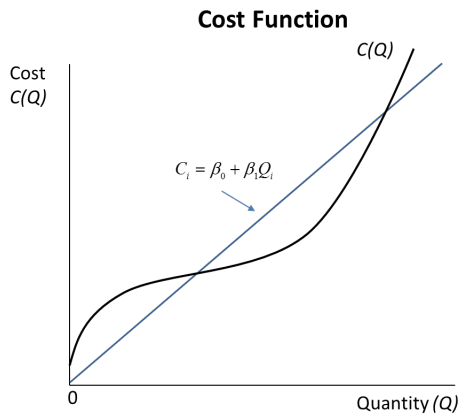
How Can the Model be Wrong?

- Omitted variables
- Non-linearity
 - ... in the variables
 - ... in the parameters
- Think of this as an opportunity – including previously omitted variables or incorporating nonlinearity can improve our analysis

Omitted Variables

- You specify $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$
- When you could have specified $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
- When will this matter?

Nonlinear Models Are All Over Economics!



Nonlinearity in the Variables

- Examples

$$Y_i = \beta_0 + \beta_1 \frac{1}{X_{1i}} + \varepsilon_i$$

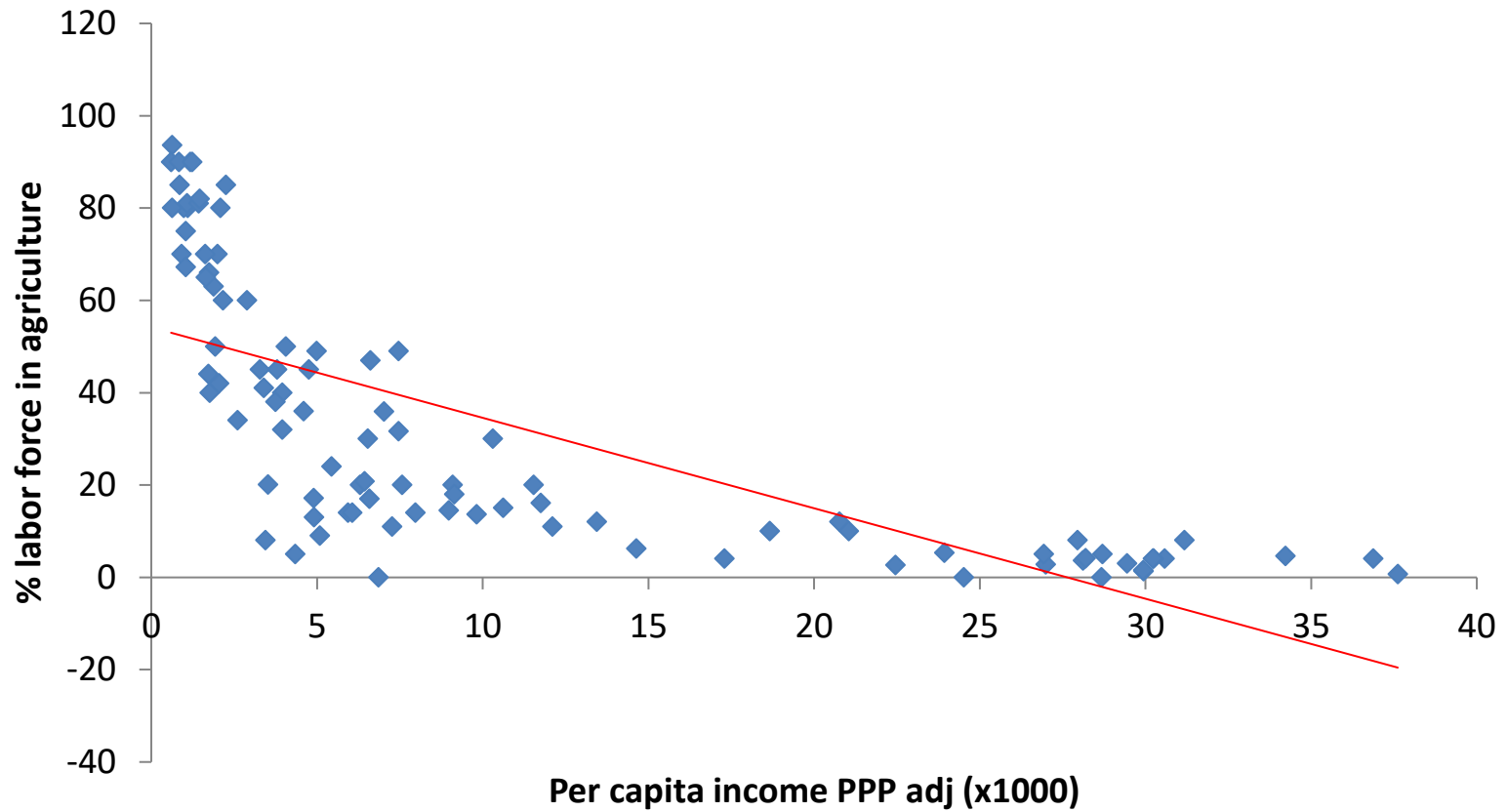
$$Y_i = \beta_0 + \beta_1 \ln(X_{1i}) + \varepsilon_i$$

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_{1i}) + \varepsilon_i$$

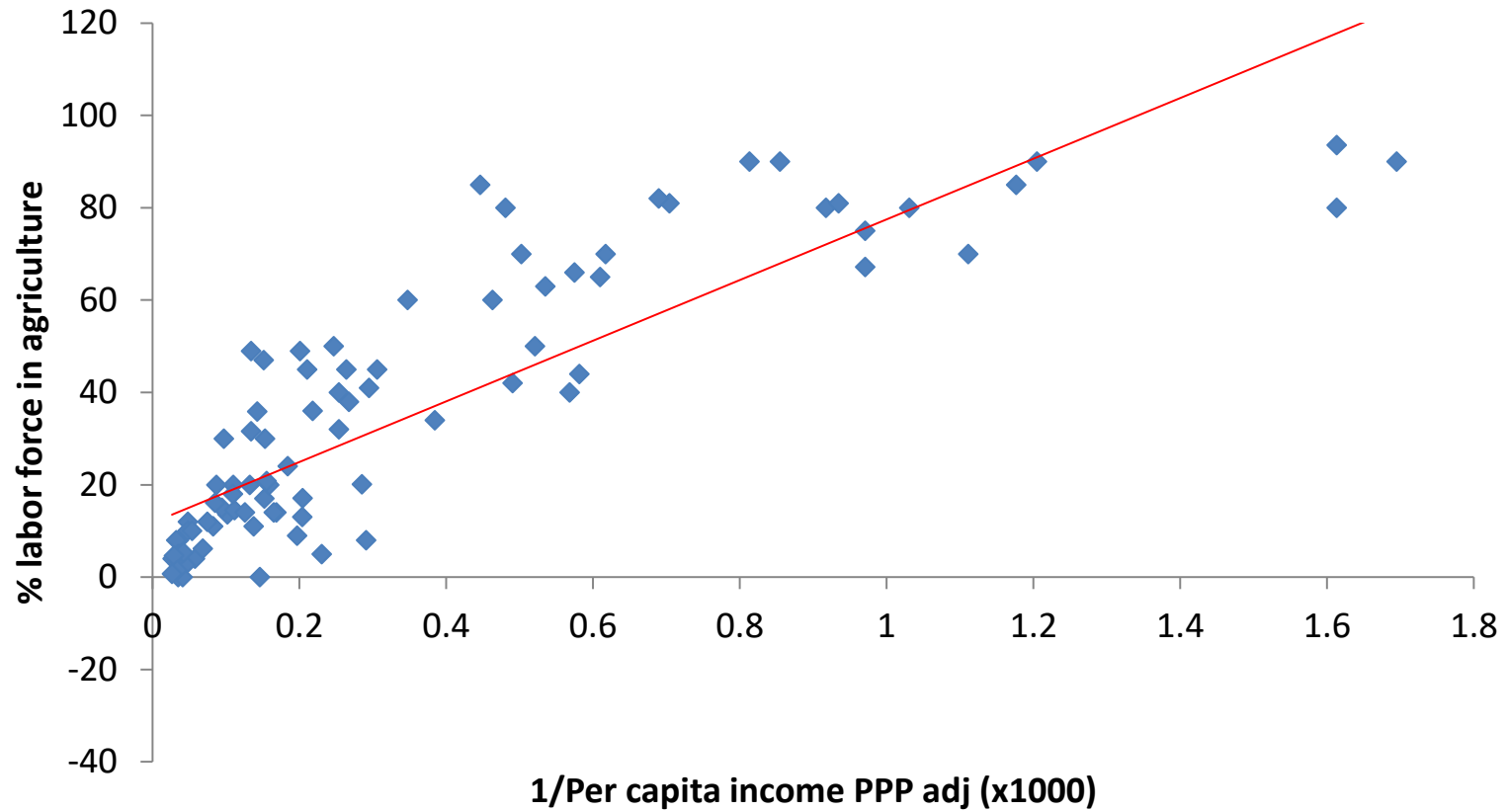
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \varepsilon_i$$

- Can we still estimate with OLS?

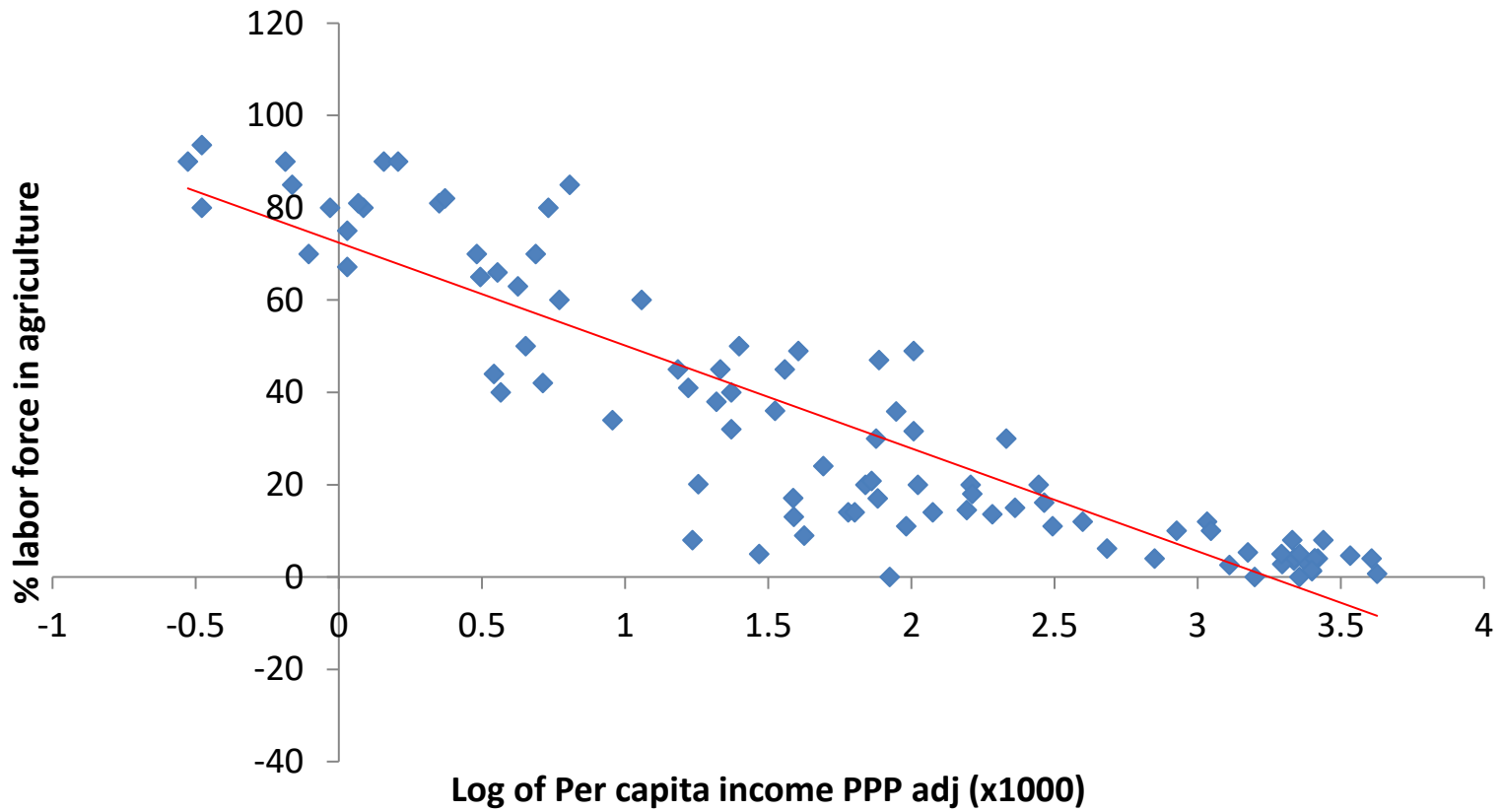
Moving off the Farm



Moving off the Farm



Moving off the Farm



Nonlinearity in the Parameters

- Example: Cobb-Douglas Production Function

$$Q_i = F(L_i, K_i) = \beta_0 L_i^{\beta_1} K_i^{\beta_2}$$

- or, with constant returns to scale

$$Q_i = F(L_i, K_i) = \beta_0 L_i^{\beta_1} K_i^{1-\beta_1}$$

- The β coefficients are elasticities

$$\ln(Q_i) = \ln(\beta_0) + \beta_1 \ln(L_i) + \beta_2 \ln(K_i) + \varepsilon_i$$

Dummy Variables

- A dummy variable takes the value zero or one.
- Do men earn more than women?

$$\ln(E_i) = \beta_0 + \beta_1 FE_i + \varepsilon_i$$

- The Mincer Earnings Model:

$$\ln(E_i) = \beta_0 + \beta_1 ED_i + \beta_2 EX_i + \beta_3 EX_i^2 + \varepsilon_i$$

- How would you test whether the return to education is the same for women as men?

More Challenging Nonlinearity in the Parameters

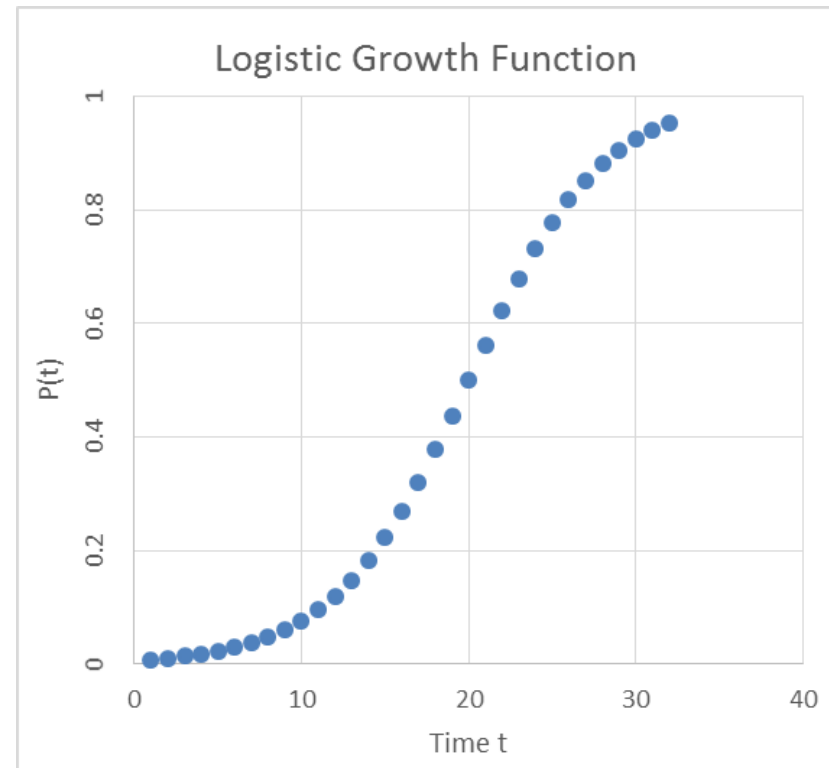
Logistic Growth Function

$$Y_t = 1 / (1 + e^{-(\beta_0 + \beta_1 t)})$$

Transform for OLS:

$$1 / Y_t = (1 + e^{-(\beta_0 + \beta_1 t)})$$

$$\ln(1 / Y_t - 1) = \beta_0^* + \beta_1^* t$$



What We Learned

- An econometric model will almost never be literally right in the sense that it describes precisely how one variable depends on a set of other variables.
 - A good econometric model is as simple as possible and fits the data well.
- Taking logs of the variables, fitting a polynomial, adding interaction terms, and adding right-hand-side variables are easy ways to improve model specification.