Lecture 3

Types of Data that Economists Analyze

Let y_{it} denote an observation on a dependent variable y for an i-th "individual" (person, state, country, firm, etc.) obsered at time t (year, month, quarter, etc.). Likewise define $X_{it1}, X_{it2}, \cdots, X_{itK}$ as the i-th individual's observation on the K explanatory variables $X_1, X_2, \cdots, X_K$ taken at time t.

Now consider the following data types:

Cross-section Data

Cross-section data consists of observations on a dependent variable, y, and explanatory variables $X_1, X_2, \cdots, X_K$ for N "individuals," $i = 1, 2, \cdots, N$ taken at one point in time, $t = t_0$. Therefore, the matrix of observations on such a data set would look like

$$
\begin{array}{ccccc}
y_{1t_0} & x_{1t_01} & x_{1t_02} & \cdots & x_{1t_0K} \\
y_{2t_0} & x_{2t_01} & x_{2t_02} & \cdots & x_{2t_0K} \\
\vdots & \vdots & \vdots & & \vdots \\
y_{Nt_0} & x_{Nt_01} & x_{Nt_02} & \cdots & x_{Nt_0K}
\end{array}
$$

when there is no chance of misunderstanding, we can drop the t_0 index and represent the cross-section data set by

$$
\begin{array}{ccccc}
y_1 & x_{11} & x_{12} & \cdots & x_{1K} \\
y_2 & x_{21} & x_{22} & \cdots & x_{2K} \\
\vdots & \vdots & \vdots & & \vdots \\
y_N & x_{N1} & x_{N2} & \cdots & x_{NK} .
\end{array}
$$

Example:

y_i = per-capita real GDP of developing countries observed in 1998

x_{i1} = literacy rate of population in i-th country

x_{i2} = degree of development of property rights in i-th country

$\vdots$

x_{iK} = mortality rate of children in i-th country

Time Series Data

In contrast a time series data set is a set of observations on one individual, $i = i_0$, observed over time, $t = 1, 2, \dots, T$. The data array for such data is

$$
\begin{array}{lllll}
y_{i_0 1} & X_{i_0 11} & X_{i_0 12} & \cdots & X_{i_0 1K} \\
y_{i_0 2} & X_{i_0 21} & X_{i_0 22} & \cdots & X_{i_0 2K} \\
\vdots & \vdots & \vdots & & \vdots \\
y_{i_0 t} & X_{i_0 t1} & X_{i_0 t2} & \cdots & X_{i_0 tK} \\
\vdots & & & & \\
y_{i_0 T} & X_{i_0 T1} & X_{i_0 T2} & \cdots & X_{i_0 TK}
\end{array}
$$

This data array can be simplified as long as it is understood that we are talking about a particular individual, $i = i_0$ and we can drop the i_0 subscript resulting in

$$
\begin{array}{lllll}
y_1 & X_{11} & X_{12} & \cdots & X_{1K} \\
y_2 & X_{21} & X_{22} & \cdots & X_{2K} \\
\vdots & \vdots & & & \\
y_t & X_{t1} & X_{t2} & \cdots & X_{tK} \\
\vdots & \vdots & & & \\
y_T & X_{T1} & X_{T2} & \cdots & X_{TK}
\end{array}
$$

Example:

y_t = monthly rate of return on IBM stock

X_{t1} = monthly rate of return on S&P 500 stock index.

(Market Model used in the field of Finance)

Panel Data

In panel data we have observations on N individuals, $i=1,2,\cdots,N$ over T time periods, $t=1,2,\cdots,T$. The data for such data is

$$\left.\begin{array}{ccccc} y_{11} & X_{111} & X_{112} & \cdots & X_{11K} \\ y_{21} & X_{211} & X_{212} & \cdots & X_{21K} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{N1} & X_{N11} & X_{N12} & \cdots & X_{N1K} \end{array}\right\} \begin{array}{l}\text{cross-section} \\ \text{of } \underline{\text{same}} \ N \text{ individuals} \\ \text{at time } t=1\end{array}$$

$$\vdots$$

$$\left.\begin{array}{ccccc} y_{1t_0} & X_{1t_01} & X_{1t_02} & \cdots & X_{1t_0K} \\ y_{2t_0} & X_{2t_01} & X_{2t_02} & \cdots & X_{2t_0K} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{Nt_0} & X_{Nt_01} & X_{Nt_02} & \cdots & X_{Nt_0K} \end{array}\right\} \begin{array}{l}\text{cross-section} \\ \text{of } \underline{\text{same}} \ N \text{ individuals} \\ \text{at time } t=2\end{array}$$

$$\vdots$$

$$\left.\begin{array}{ccccc} y_{1T} & X_{1T1} & X_{1T2} & \cdots & X_{1TK} \\ y_{2T} & X_{2T1} & X_{2T2} & \cdots & X_{2TK} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{NT} & X_{NT1} & X_{NT2} & \cdots & X_{NTK} \end{array}\right\} \begin{array}{l}\text{cross-section} \\ \text{of } \underline{\text{same}} \ N \text{ individuals} \\ \text{at time } t=T.\end{array}$$

An equivalent way of organizing the above panel data would be by time-series observation by each individual.

$$\left.\begin{array}{ccccc} y_{11} & X_{111} & X_{112} & \cdots & X_{11K} \\ y_{12} & X_{121} & X_{122} & \cdots & X_{12K} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{1T} & X_{1T1} & X_{1T2} & \cdots & X_{1TK} \end{array}\right\} \begin{array}{l} T \text{ time series observations} \\ \text{on the first individual} \\ i=1. \end{array}$$

$$\vdots$$

$$\left.\begin{array}{ccccc} y_{i_0 1} & X_{i_0 11} & X_{i_0 12} & \cdots & X_{i_0 1K} \\ y_{i_0 2} & X_{i_0 21} & X_{i_0 22} & \cdots & X_{i_0 2K} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{i_0 T} & X_{i_0 T1} & X_{i_0 T2} & \cdots & X_{i_0 TK} \end{array}\right\} \begin{array}{l} T \text{ time series observations} \\ \text{on the } i=i_0 \text{ individual} \end{array}$$

$$\vdots$$

$$\left.\begin{array}{ccccc} y_{N1} & X_{N11} & X_{N12} & \cdots & X_{N1K} \\ y_{N2} & X_{N21} & X_{N22} & \cdots & X_{N2K} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{NT} & X_{NT1} & X_{NT2} & \cdots & X_{NTK} \end{array}\right\} \begin{array}{l} T \text{ time series observations} \\ \text{on the } i=N \text{ individual} \end{array}$$

Example: Real GDP growth rates observed across $i=1,2,\cdots,N$ countries (the same countries) over time $t=1,2,\cdots,T$.

Pooled Cross-Section and Time Series Data

The only difference between PCSTS data and panel data is that the individuals in each cross-section can change over time.