spm9155 15 November 2012
due date 30 November 2012

Assignment 2a: Use of data

---

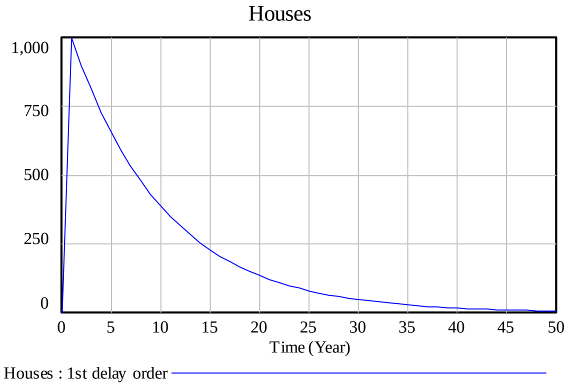
3. Estimating parameters using a model equation: estimating the order of a delay and the delay time

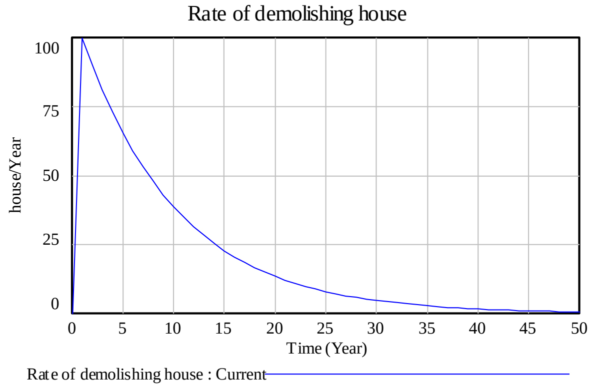
The case of house demolishing could be taken as an example of estimating delay time and delay order.

According to Sterman (2000), there are two approaches to determine the delay order and delay time: (1) If numerical data is available, it is possible to use econometric or statistical tools (2) If numerical data is not available, the only way is to estimate the current case based on analogue cases.

Let's assume we have somehow housing database, so the 1st method is chosen. We need to treat the building house and the demolishing house as a single and aggregated process, with one variable stock and 2 flows, and with a material delay. This means every intermediate delay (administrative process, communication process, construction process, etc…) is combined into one delay. The material delay could be pipeline delay, first or higher order delay. Moreover, we need to assume that the delay time is constant.

To make it easy, we only consider the houses which finished building in the same certain year, and record the data of year they are demolished, or how many units among those one are demolished every year. This mean we consider the input flow as a pulse. Then by looking at the graph of behavior of stock and outflow, we could deduce the delay order and delay time. And obviously, in this case, the delay could not be pipeline delay because in reality, houses have different existing duration. The demolishing house is a mixing and variable individual process.

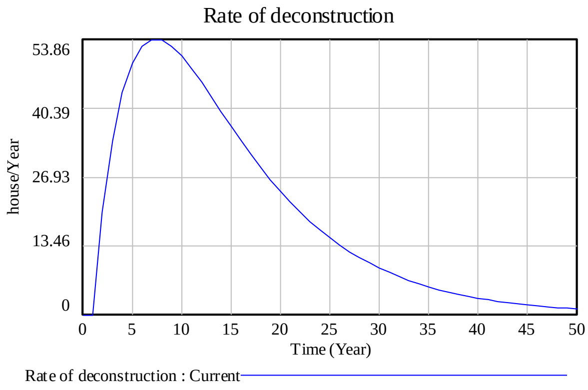




If the distributions of house and demolished house every year are "more or less" similar to above figures, the model must have the 1st order of delay. We could derive the delay time through the plot of number of house every year. As the exponential decay is given by $S = S_0 exp(-t/D)$ so when $t = D$, the stock S has fallen to $exp(-1) = 0.37$ of its initial level (Sterman, 2000). In this case, the S_0 is 1000 houses. At time ~ 10, S is around 370 houses. Therefore, the delay time is 10 years.

If the distribution of demolished house every year is "more or less" similar to figure 2, the model must have the 2nd or higher order of delay.

From here, by using Erlang distribution, we could calculate the mean delay time and delay order of the case.



4. Estimating parameters using a statistical relationship between two (or more) variables

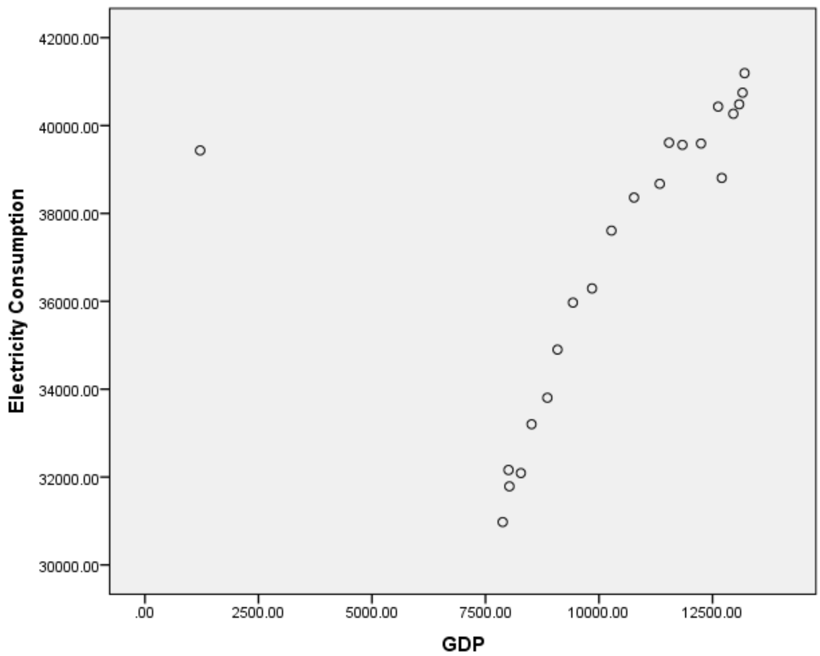
The statistical relation between "Real Gross Domestic Product" and "Consumption for Electricity Generation by Energy Source" of the United Nations of American from 1989 – 2010.

The descriptive statistics of those 2 variables:

Descriptive Statistics

	Mean	Std. Deviation	N
GDP	10222.6364	2778.19003	22
Electricity Consumption	37089.0455	3394.02080	22

The scatter plot between those 2 variables:



The Bivariate Correlation test:

Correlations

	GDP	Electricity Consumption

GDP	Pearson Correlation	1	.550**
	Sig. (1-tailed)		.004
	N	22	22
Electricity Consumption	Pearson Correlation	.550**	1
	Sig. (1-tailed)	.004	
	N	22	22

**. Correlation is significant at the 0.01 level (1-tailed).

- The r-value = 0,550 which means there is "more or less" a correlation between those 2 variables, though not too strong. The value of r-value is positive, which means those 2 variables have proportional relation.
- The "**" means we could reject the Null Hypotheses (reject hypotheses of "no actual correlation)
- The p-value = 0,04 < 0,5 which means we could reject the Null Hypotheses. So We have evidence that we found a true relationship.

The analysis of linear regression:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Electricity Consumption	.	Enter

a. All requested variables entered.

b. Dependent Variable: GDP

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.550[a]	.302	.267	2378.27964	.302	8.656	1	20	.008

a. Predictors: (Constant), Electricity Consumption

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	30225.186	2413.795		12.522	.000			
	GDP	.671	.228	.550	2.942	.008	.550	.550	.550

a. Dependent Variable: Electricity Consumption

Because the p-value = 0,008 < 0,05 so we could reject H0.

<u>Our regression equation becomes</u>:

$$\boxed{\textit{Electricity Consumption} = 0.671 \times \textit{GDP} + 30225{,}186}$$

Data

Source: The U.S. Energy Information Administration (EIA)

Year	Real Gross Domestic Product	Consumption for Electricity Generation by Energy Source: Total (All Sectors)	
	Billion Chained (2005) Dollars (Rounded up)	Billion Btu	Trillion Btu (Rounded Up)
1989	7879	30.975.709	30976
1990	8027	31.787.701	31788
1991	8008	32.160.248	32161
1992	8280	32.090.536	32091
1993	8516	33.202.865	33203
1994	8863	33.803.057	33804
1995	9086	34.900.988	34901
1996	9426	35.970.632	35971
1997	9846	36.292.777	36293
1998	10275	37.607.340	37608
1999	10771	38.361.742	38362
2000	1216	39.433.430	39434
2001	11338	38.672.033	38673
2002	11543	39.609.541	39610
2003	11836	39.559.105	39560
2004	12247	39.590.547	39591
2005	12623	40.430.104	40431
2006	12959	40.267.618	40268
2007	13206	41.192.733	41193
2008	13162	40.746.536	40747
2009	12703	38.808.093	38809
2010	13088	40.484.954	40485

STERMAN, J. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Irwin/McGraw-Hill c2000. .