Economic Models and Empirical Methodology

This handout applies theoretical economic models to empirical tests using regression. This will give you a preview of how your data should be organized and how your economic model relates to the data.

Sacramento housing prices

Cross section study: To what extent do housing prices in Sacramento reflect home size, number of bedrooms, lot size, location, and other variables?

Time series study: What are the primary sources of the dramatic increases in Sacramento's housing prices over the last 15 years? How do general economic conditions and housing prices in surrounding areas affect Sacramento housing?

This is an example of a research topic that is driven by data, rather than a fundamental economic model. However, there is a basic model of demand and supply in the background. The cross section study relies on a demand-supply approach to determine the value of a house, but it is difficult to separate the demand from supply factors. In this case, the researcher will spend a good deal of time discussing why individual explanatory variables are the most important from an intuitive point of view. For instance, in a time series study, how would you expect Sacramento's median income level to affect house prices? There should be a positive effect because this would lead to an increase in the demand for houses.

Example of regressions for this research question are listed below

<u>Cross section</u>

Dependent variable:	house price (P)
Explanatory variables (cross section):	lot size measure in tenths of an acre (LOT), house square footage (SQF), number of bedrooms (BED), number of bathrooms (BATH)

Cross section regression: $P_i = \beta_0 + \beta_1 * LOT_i + \beta_2 * SQF_i + \beta_3 * BED_i + \beta_4 * BATH_i$

where P_i is the house price for house *i*, and BED_i is the number of bedrooms in this house (house *i*). The subscript *i* is commonly used to indicate "individual" in cross section regressions. So, β_1 measures how a one-tenth acre increase in lot size affects the house price. One would expect that β_1 would be a positive number because a bigger lot usually increases property value.

<u>Time series</u>

Dependent variable:	area (median or average) house price (P)
Explanatory variables (time series):	interest rate (R), area income (INC), population (POP), crime rates (CRIME), unemployment (UNEMP)

Time series regression: $P_t = \beta_0 + \beta_1 * R_t + \beta_2 * INC_t + \beta_3 * POP_t + \beta_4 * CRIME_t$

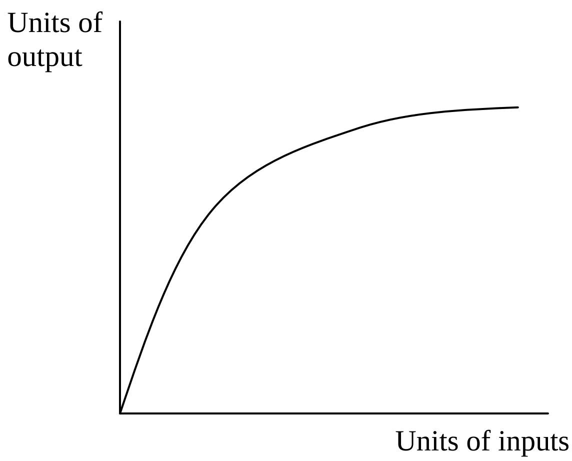
where P_t is the median or average house price at time *t*, and INC_t is the area income level at time *t*. The subscript *t* is commonly used to indicate "time" in time series regressions. So, β_1 measures how a one-percentage point increase in the interest rate affects the area house price. One would expect that β_1 would be a negative number because a higher interest rate means a higher cost of borrowing – reducing demand for homes (for purchase).

Education production functions and student performance

Do changes in school expenditures cause changes in student performance? For instance, one may believe that increased school spending reduces student-teacher ratios, allows schools to attract higher quality teachers, and/or improves upon school administration capabilities.

This researcher topic uses a production function approach as its theoretical backbone. The inputs are the researcher's explanatory variables (expenditures, class size, etc.) with the output being student performance (the dependent variable measured by test scores or some other quality measure). Chapter 7 of *the Economics of Education* by Cohn and Geske is a good resource for how to link the theory to the empirical test.

If the researcher assumes the production function has nice properties, such as diminishing marginal returns to inputs (curvature in the production function) and that an increase in inputs leads to higher output (a positive slope in the production function).



Example of a regression for this research question:

Dependent variable:	Student performance at an individual school (measured by a test score, TEST)
Explanatory variables (cross section):	Expenditures in dollars (EXP), class size (SIZE), computers per student (COMP), teacher salary (SAL),

Cross section regression: $TEST_i = \beta_0 + \beta_1 * EXP_i + \beta_2 * SIZE_i + \beta_3 * COMP_i + \beta_4 * SAL_i$

where $TEST_i$ is the average test score at school *i* (this could be generalized to individual districts rather than schools), and $SIZE_i$ is the number of students at this school (school *i*). Again, the subscript *i* is commonly used to indicate "individual" in cross section regressions. So, β_1 measures how a one-dollar increase in expenditures affects average test score. One would expect that β_1 would be a positive number because we would like to think that more expenditures translates into improved student performance.

Predicting stock prices

Mutual fund managers and investors generally look at data on price-dividend ratios, or price-earnings ratios to determine whether or not a stock will gain value. Does monetary policy matter? Has the relationship between these variables changed over time?

An important model in the determination of stock prices is the efficient markets model. This expresses the current stock price as being equal to its fundamental value, usually measured as the present value of expected future dividends on the stock.

$$P_t = \frac{D^e_{t+1}}{(1+i)} + \frac{D^e_{t+2}}{(1+i)^2} + \frac{D^e_{t+3}}{(1+i)^3} + \cdots$$

$$P_t = \sum_{j=1}^{\infty} \frac{D^e_{t+j}}{(1+i)^j}$$

An important implication from the efficient markets model is that investors cannot "beat the market" because all available information has already been incorporated into the current stock price. As an investor, if you see a riskless profit opportunity, so will someone else, causing an increase in the stock price and the elimination of potential capital gains. Another implication is that stock prices follow a random walk – meaning that future changes in stock prices cannot be predicted in advance (the stock price immediately reacts to the arrival of new information).

How to test the validity of this model? There are a few approaches, one is to look at the relationship between dividends and stock prices. Since the current stock price is simply an average of expected future dividends, it should be less variable than the dividends themselves. In the data, this does not hold true (a problem known as excess volatility), posing a challenge to the validity of this model.

The other approach is to see if there are in fact investment strategies that will beat the market, such as using technical analysis or other active stock trading strategies used by mutual fund managers. Many of these investment strategies involve regression to predict future stock prices.

Finally, one could test the random walk hypothesis of stock prices. Is it possible to predict future changes in stock prices? One could use regression to test this.

Example of a regression to test the random walk:

Dependent variable: Current stock price (P_t)
Explanatory variable: Stock price last period (P_{t-1}), time trend (this is called, a "drift" to account for growth in stock prices, *t*)

$$P_t = \beta_0 + \beta_1 * P_{t-1} + \beta_2 * t$$

The time trend takes the value of one in the first period, and two in the second, and so on. According to the random walk hypothesis, changes in the stock price should be random, so (ignoring the time trend) if $\beta_1 = 1$:

$$P_t = \beta_0 + 1 * P_{t-1}$$
$$(P_t - P_{t-1}) = \beta_0 + \beta_2 * t$$

So, if β_1 is equal to one, the change in the stock price predicted by the model is a constant (plus some growth trend inherent in stock prices). Therefore, any movements in the data would be attributed to random error (the "residual" from the regression).