STOCHASTICITY IN STUDENT PROJECT ON LOSS DEVELOPMENT

(The attached PDF file has better formatting.)

{Jacob's question shows the difference between parameters and variables. If a scenario has an inflation rate and stochasticity, one presumes the inflation rate is uncertain. This is not the intention. Distinguish between

- A parameter (β) whose estimator is a random variable.
- The dependent variable, which has a standard error σ.

The dialogue below compare the regression analysis student project template with the time series project templates, which focus on modeling the inflation or interest rate.}

Jacob: The assumptions above use a known inflation rate and geometric decay. In truth, we don't know the inflation rate or the loss payment pattern with certainty. Do we add uncertainty to the geometric decay and the inflation rate?

Rachel: It is true that inflation and loss payment patterns are uncertain.

- The time series course fits ARIMA models to a stochastic time series of inflation rates.
- The observed inflation rates are constrained random draws from a distribution.
- The ARIMA model solves for the distribution and the constraints.

For the regression analysis here, the inflation rate is a parameter, not a variable.

~ We estimate the inflation rate by regression analysis; it is not stochastic. The estimator for the inflation rate is stochastic, but the true inflation rate is not stochastic.
~ The paid losses in each cell are random variables; they are stochastic.

Jacob: Aren't we assuming that the inflation rate or the loss payment pattern changes? Aren't we trying to model the type of change (discrete vs continuous) and its magnitude?

Rachel: An ARIMA process models the time series of inflation rates.

- The inflation rate is the variable.
- The ARIMA process has several parameters that we estimate.

For the project template on loss reserving, inflation is one or more parameters.

- In the project template on loss reserving, the variable is the logarithm of paid losses.
- Based on the residual plots, we form a model with inflation as a parameter.

    - If the residual plot suggests a discrete change, we estimate two inflation rates.

- If the residual plot suggests a continuous change, we estimate an inflation rate and an annual change in the rate.

Jacob: If the inflation rate is a parameter, not a variable, it has a fixed value. That value does not have a variance or a standard deviation, since it is a scalar. But we estimate the value and work out its standard deviation. How can you say it is not a variable?

Rachel: The inflation rate is a parameter, with no variance. The ordinary least squares estimator of the inflation rate is a function of a random variable (the logarithm of paid losses), so it is also a random variable, and it has a variance.

Jacob: Doesn't regression analysis estimate stochastic items?

Rachel: We forecast the future loss payments; these are the stochastic items. From the past realizations of stochastic variables (the past loss payments), we estimate the values of regression coefficients, such as the inflation rate. The ordinary least squares estimator of the inflation rate is a random variable. The estimate itself is a parameter.

- An ARIMA model specifies how the inflation rate changes over time.
- The regression analysis in the project template here estimates the inflation rate. You may deduce a constant inflation rate or a rate that changes by year. But even a rate that changes by year is not stochastic. We estimate a stable annual change, not a change that varies randomly from year to year.

{Focus on the difference between the inflation rate as a parameter of a regression equation and the inflation rate as the observation in an ARIMA process. In both cases, the expected inflation rate may increase from 8% the first year to 12% the last year.

- *Regression parameter:* expected inflation depends on the model, such as 8% the first year and increasing 0.2% a year for 20 years.
- *ARIMA process:* expected inflation in year t depends on the ARIMA process and the observed inflation in year t-1 (and perhaps previous years as well).}

MODELING STOCHASTICITY IN EXCEL

Jacob: How do we add the stochasticity? Do we add a random number to each Y value?

Rachel: We add a random draw from a normal distribution. We have a column of error terms. For each cell in the column, we have several steps:

- The RAND built-in function generates a *random number* from a *uniform distribution* between zero and one. The illustrative spreadsheet puts this number in column H.
- The NORMSINV built-in function generates a *random draw from a standard normal distribution* based on the random number and puts this figure in column I.
  - If the random number is ½, the random draw is 0.
  - If the random number is between ½ and 1, the random draw is positive.
  - If the random number is between 0 and ½, the random draw is negative.
- The random number varies from 0 to 1; the random draw varies from $-\infty$ to $+\infty$.
- This project template assumes the data are homoscedastic. This is not usually true, but it simplifies the math. Multiply the random draw from the standard normal distribution by σ, the standard error of the error term. Put this product in Column J.
- If the regression is heteroscedastic, the variance of the error term varies. To examine the effects of conditional heteroscedasticity, we multiply the value in Column I by some function of the values in Columns B, C, D, or E. For conditional heteroscedasticity, you might multiply Column H by 10% of Column D.

Take heed: The *DATA ANALYSIS* add-in can generate random draws from a normal distribution with any mean or standard deviation. To do this yourself, multiply the random draw from a standard normal distribution by σ (the standard deviation) and add μ (the mean). The *DATA ANALYSIS* add-in can generate random draws from other distributions as well. In many cases, Excel also has built-in functions to do this.

We have 120 observations. On the illustrative worksheet, the observations are in rows 11 through 130. Use the Excel *REGRESSION* add-in to solve for the ordinary least squares estimators of the geometric decay and the inflation rate.

Take heed: The illustrative spread-sheet is an example. Excel has several ways to do regression analysis. If you are proficient at Excel, you may use VBA macros and Excel built-in functions (such as *LINEST, SLOPE, INTERCEPT, GROWTH*).

The *REGRESSION* add-in is simple, but it is static, not dynamic. The Excel built-in functions are dynamic. *You generally want the static result, not a dynamic result.*

Illustration: For a set of 120 data points, you estimate the regression coefficients. After examining the results, you suspect that unusual random fluctuations distort the regression line. You simulate the 120 data points again.

- The results from the *REGRESSION* add-in are values, not functions. These values don't change, even through the input has changed. You must re-run the *REGRESSION* add-in, either over-writing the previous results or placing the output in a different location.
- The Excel built-in functions are dynamic. When you change the simulation and re-calculate the work-sheet, the regression output changes. To save the old output, you must manually convert the formulas to values.

In most cases, you want to keep the old output and compare it to the new output. You can do this most easily with the *REGRESSION* add-in.