

*PROJECT TEMPLATE: DISCRETE CHANGE IN THE INFLATION RATE*

(The attached PDF file has better formatting.)

{This posting explains how to simulate a discrete change in a parameter and how to use dummy variables to test a hypothesis and adjust for this change. Focus on the indexing of the change and the use of cumulative (not incremental) inflationary changes.}

*Jacob:* The illustrative worksheet shows the values of the simulation parameters in the upper left section (Cells D5:F10). Is this for documentation?

*Rachel:* The illustrative worksheet uses a value of 15% for  $\beta_2$ . It places the value in a table of names and uses the name in the cell formulas. Using names keeps your formulas clear.

*Take heed:* Names are not necessary, but they keep your worksheet clearer.

*Jacob:* How do we handle a discrete change in the inflation rate?

Suppose the inflation rate is 20% in 2001 – 2010 and 10% in 2011 – 2015. Change the matrix of names to use BETA2 of 20% and BETA2B of 10%. (See the worksheet *UNSTABLE RATES*). Select the table of names, choose NAMES from the INSERT menu, and chose CREATE from the names menu. The default of LEFT-HAND COLUMN is proper if your table is like the one in the spreadsheet.

*Take heed:* You can make the table other ways, and you can create names in various ways. Many candidates prefer to use Excel's name box in the upper left corner of the worksheet. Use whatever method you are comfortable with.

*Take heed:* By default, Excel creates workbook level names. If you simulate several scenarios, use worksheet level names. Excel 2007 has a name manager that allows use to easily choose the level of the name (workbook or any worksheet). Previous versions of Excel require manual coding of the worksheet name.

For the simulated Y values, change the formulas in the cells. Instead of multiplying the calendar year by beta2, multiply by a combination of beta2 and beta2b. You can use

- *Copy and paste:* low-tech but not efficient.
- *Excel's IF function:* for candidates familiar with Excel but not VBA. This is the simplest method, which most candidates will use.
- A VBA macro with an IF statement: we show illustrative code below. VBA code is useful if you have a large project and you want the macro to perform many simulations. It is not needed for this spreadsheet, but we show the code for candidates who want it.

*Jacob:* The spread-sheet shows the Y value before adding stochasticity; then it adds the stochasticity in a separate step. Can we combine these two steps?

*Rachel:* The illustrative worksheets are written for new Excel users. Computations are done one at a time, so you can follow the sequence. The worksheet shows the Y value before adding stochasticity so you can make sure you properly apply the inflation rates. Verify your equations and formulas before applying them to the data.

Use a stochasticity of  $\sigma = 0$  for the initial simulation. This simulation shows you two things:

- The regression shows the average inflation rate over the 120 observations. If you are unsure of your Excel formula, verify that the *average inflation rate* is correct.
- The residuals are not zero. Some candidates assume the residuals start at zero and decline, or start at zero and increase, or start at zero and first decline then increase or first increase and then decline. Some candidates assume the residuals should start above or below zero, move to zero at the midpoint, and then reverse direction. None of these is correct. *By using a stochasticity of zero, you see the expected values of the residuals.* For the student project, explain why this pattern is expected.

*Jacob:* What do we expect for the residuals? Aren't the expected residuals always zero? Isn't this an assumption of classical regression analysis?

*Rachel:* The average residual over *all* observations is zero. This is always true; it stems from the ordinary least squares method.

- ~ If the regression coefficient is constant, the *expected* residual is zero at each point.
- ~ If the Y values are stochastic, the actual residuals are not zero.

The *slope* of the line segment *connecting the average residuals by calendar year* reveals *if the parameter is not constant*. The example on the illustrative spreadsheet is explained in the project template. For the student project

- Simulate a non-constant regression coefficient.
- Use residual plots to hypothesize and test the proper regression equation.

*Jacob:* Can we observe how stochasticity affects the residual plots?

*Rachel:* Start with a low stochasticity, such as  $\sigma = 0.01$ , and increase  $\sigma$  gradually.

- As  $\sigma$  increases, the residual plots become harder to read.
- For your student project, use moderate stochasticity. Make  $\sigma$  high enough that the residuals vary, but the pattern is still clear.

*To summarize:* If the regression coefficient is constant, the residual plot is a horizontal line. If the regression coefficient has a discrete change, the residual plot is a V or an upside-down V. Use a large discrete change in the regression coefficient and no stochasticity to verify your work. Use more realistic figures to see how stochasticity affects the residual plot. Explain the shape of the residual plot in your write-up.

## DUMMY VARIABLES

*Jacob:* How do we correct for the change in the regression coefficient?

*Rachel:* We use a dummy variable, breaking at the year in which inflation changes. *The sum of squared residuals decreases if your technique is correct and the residual plot becomes a horizontal line.* The dummy variable is proper for a discrete change in the inflation rate; it does not work well if the change is continuous.

- We show a complete illustration with a dummy variable, with cell formulas, VBA macros, comments, and other documentation.
- The discussion forum provides past student projects using dummy variables. Review these materials as well.

If you do a student project on loss reserving, change the scenario.

- First replicate the illustrative worksheets to be sure you understand the procedures.
- Then change the parameters one by one, to create a different project.

Choose different values for the inflation rates, the development year trend, and  $\alpha$  (alpha). Form the residual plots, explain them in your write-up, and estimate the regression equation with the dummy variable.

### *ILLUSTRATION: PARAMETER INSTABILITY*

*Jacob:* The choice of  $\sigma$  affects the random fluctuation in the simulated data, the variance of the error term, the residual plots, and the significance of the results. Do we use a low  $\sigma$ , such as 0.01, or a realistic  $\sigma$ , such as 0.25?

*Rachel:* The student project uses a moderate stochasticity. The steps using zero or low stochasticity ensure that you do the work correctly.

- Simulate first with a constant inflation rate.
- Use the *REGRESSION* add-in to solve for the ordinary least squares estimators.
- The ordinary least squares estimator for the inflation rate should be close to the simulation parameter.
- The  $s^2$  of the regression should be close to the  $\sigma^2$  used in the simulation.

*Jacob:* If the estimator for the inflation rate does not equal the simulation parameter, how do we know if it is *close enough*?

*Rachel:* The *REGRESSION* add-in gives the standard error of the estimator. The ordinary least squares estimator should be within  $\pm 2$  standard deviations of the true parameter 95% of the time.

- If it is three or four standard deviations away, compare the standard deviation of the  $\beta_2$  parameter with the formula in the textbook. You may have an error in the worksheet.
- If it is two to three standard deviations away, simulate again. If the estimate is now close to the simulation parameter, continue with your project.

*Jacob:* Every time I change a cell in the worksheet, all the figures change. How can we simulate loss triangles if the figures change all the time?

*Rachel:* Your worksheet has calculation set to automatic, which is the default value.

- Click on *TOOLS* → *OPTIONS* → *CALCULATION*.
- Change calculation to *MANUAL*.
- De-select (clear the check box) *RE-CALCULATE WHEN SAVING*.

The worksheet recalculates when you press the F9 key. As you create the worksheet, recalculate whenever you copy one cell formula to another cell. Once you have tested the worksheet, simulate a final time and you no longer need to re-calculate.

*Take heed:* For Excel 2007, click on *DATA* → *OPTIONS* → *CALCULATION*.

*Jacob:* We must form residual plots, which require much calculation. When we calculate the residual matrix and the means or variances of the residuals, the data will change.

*Rachel:* Use separate worksheets to simulate the data and to analyze it.

- Have the *REGRESSION* add-in place the residual output on a *separate sheet*.
- Form the matrix of residuals and residual plots on this separate sheet. You re-calculate several times to form the average residuals and their standard deviations. By using a separate worksheet, this re-calculation does not affect the simulation.

Form the residual plot. The illustrative worksheets have cell formulas and macros that do the number crunching. Focus on the shape of the residual plot.

- The line connecting the average residuals should be horizontal at the X axis.
- If the stochasticity is high, the line may not be horizontal.
- Choose a  $\sigma$  that gives moderate fluctuations but still shows a horizontal residual plot.

*Optional:* The second part of this project template examines the effects of non-constant parameters, using dummy variables and squares of the explanatory variables. Verify that *if the inflation rate is stable*, the ordinary least squares estimators for

- the square of the calendar year
- a dummy variable used with the calendar year

do not differ significantly from zero. Seeing the change in the significance of the squared calendar year variable or the dummy variable verifies your work.

*Take heed:* The project templates frequently refer to earlier or later sections of the work. Ignore these comments on your first reading. Once you have set up the Excel worksheets, these comments help you verify your work and avoid errors.

## CHANGE IN INFLATION RATE

*Jacob:* For the second part of the student project, do we simulate a reasonable change in the inflation rate, such as from 6% to 8%?

*Rachel:* Use a large enough change in the inflation rate, such as from 5% to 35%, that the residual plot becomes V-shaped. Simulate again and create the residual plots. You may reduce the stochasticity to identify the change in the inflation rate.

*Jacob:* If the inflation rate changes, is the ordinary least squares estimator the weighted average of the two rates?

*Rachel:* It is a weighted average, but the weights are not simple. If the inflation rate is 10% the first ten years and 20% the last five years, the average inflation rate is **not**  $(10 \times 10\% + 5 \times 20\%) / 15 = 13.33\%$ .

- The 13.33% is used in pricing analyses to estimate the loss cost trend if each year has equal weight.
- For the reserve analysis here, the years do not have equal weight.

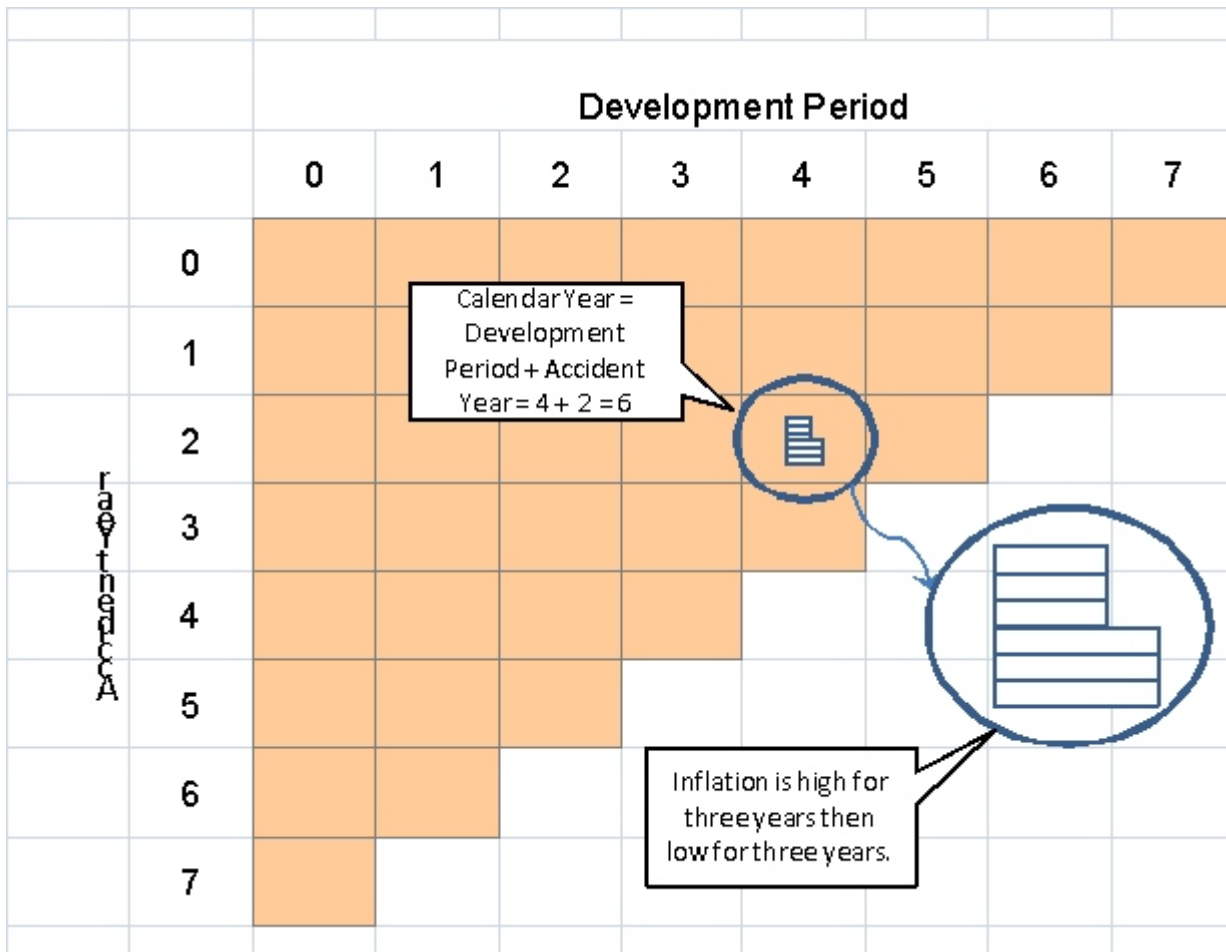
We examine how the inflation rates affect the cells of the loss triangle. The loss triangle has 120 cells with observed values, not 15 cells for calendar years.

- ~ Of the 120 cells, 1 cell is calendar year 0, 2 cells are calendar year 1, 3 cells are calendar year 2, ..., and 15 cells are calendar year 14.
- ~ The calendar year 1 cells have one year of 10% inflation. The calendar year 14 cells have 9 years of 10% inflation and 5 years of 20% inflation.

To see the average inflation rate, use a regression with zero stochasticity ( $\sigma = 0$ ).

- Inflation in calendar year 14 affects 15 cells: the lower right diagonal.
- Inflation in calendar year 13 affects  $14 + 15 = 29$  cells: the two lowest diagonals.
- Inflation in calendar year 12 affects  $13 + 14 + 15 = 42$  cells: the three lowest diagonals.

The figure below shows an 8 × 8 loss triangle, with inflation of 10% for the first 3 years and 6% for the next 5 years. The inflation affecting each cell is depicted as a stack of bars.



- The calendar years run from 0 to 7.
- Accident year 2 (third row) and development period 4 (fifth column) is circled. It is calendar year 6, so it is affected by six annual inflation rates. The first 3 are 10% (lower layers); the next three are 6% (upper layers).

*Jacob:* The number of layers in each cell depends on the chosen base year.

- If the base is the value of the dollar at the *beginning* of the 15 calendar years, calendar year 0 has 0 layers of inflation and calendar year 14 has 14 layers of inflation.
- If the base is the value of the dollar at the *end* of the 15 calendar years, calendar year 0 has 14 layers of deflation and calendar year 14 has 0 layers of deflation.

The base year is arbitrary. But different bases give different inflation in each cell. Do we get different results depending on the base year?

*Rachel:* The results are the same. Suppose calendar year 0 is the base year,  $\alpha = 12$ ,  $\beta_1 = \text{zero}$  (for simplicity), and  $\beta_2$  (the inflation rate) = 10%. If calendar year 14 is the base year, then  $\alpha = 12 + 15 \times 10\% = 13.5$  and  $\beta_2$  (the deflation rate) =  $-10\%$ . The regression results are the same.

*Jacob:* As we increase  $\sigma$ , how do the regression results change?

*Rachel:* Repeat the regression with moderate stochasticity, such as  $\sigma = 0.10$ . You find:

- The estimate of  $\sigma^2$  increases.
- The *expected* value of  $\beta_2$  does not change.
- The variance of  $\beta_2$  increases, so the estimated  $\beta_2$  may change.
- Some forecasts are too high and some are too low.
- The residual plot looks like a V or an upside-down V, but the lines are slightly jagged.
- The sum of squared residuals increases.
  - TSS and ESS increase by about the same amount.
  - RSS does not change materially.
  - Since  $TSS > ESS$ , the percentage change in  $ESS >$  the percentage change in TSS.
  - $R^2 = 1 - ESS/TSS$  decreases.

*Take heed:* Your final student project uses figures that are realistic but still give a residual plot that is easy to interpret.

*Jacob:* How do we correct for the change in the inflation rate? *Rachel:* Redo the regression with a dummy variable  $D$ . If inflation rate in 2010, the dummy variable is 0 for calendar years 2001 – 2010 and 1 for calendar years 2011 – 2015.

Re-read the chapter in the text on dummy variables if you need to. Much actuarial work with regression analysis and generalized linear models uses dummy variables for class dimensions like male/female, urban/rural, smoker/non-smoker, and married/unmarried.

*Jacob:* If the inflation rate changes in 2010, does that year get a 0 or a 1?

*Rachel:* The answer depends on the definitions you use.

- If the inflation rate in Year  $T$  means from  $T$  to  $T+1$ , the loss payments in Year  $T$  are before the change in the inflation rate.
- If the inflation rate in Year  $T$  means from  $T-1$  to  $T$ , the loss payments in Year  $T$  are after the change in the inflation rate.
- If the inflation rate in Year  $T$  means from January 1 to December 31 of Year  $T$ , and the loss payments occur on December 31, the loss payments are after the change in the inflation rate.

Any definition above is fine. Specify in your write-up what each item means.



*Take heed:* The illustrative worksheets use the first definition above. Inflation in Year T means from T to T+1, so loss payments in Year T are before the change in inflation.

*Jacob:* How can we check if we use the proper dummy variables for each year?

*Rachel:* With the dummy variable, the residual plot should be a horizontal line. If it is not, you may be applying the dummy variable incorrectly. Redo the regression with a  $\sigma$  of zero.

- If the dummy variable is applied correctly, the residuals should be zero.
- If the dummy variable is not applied correctly, the residuals differ from zero.

*Jacob:* Do you describe how to use dummy variables in the project template?

*Rachel:* The statistical techniques are described in the textbook and in the course modules. This project template assumes you know how to use dummy variables.

*Take heed:* The textbook has several variations of dummy variables. The dummy variable may affect the intercept ( $\alpha$ ), the slope ( $\beta$ ), or both.

- Use a dummy variable for the slope coefficient  $\beta_2$  *after* a given year.
- You may write this as a dummy variable for  $\beta_2$  for all years offset by a change in  $\alpha$  when the dummy variable = 1.

*Jacob:* How might we do each step?

*Rachel:* One illustrative spreadsheet shows residual plots for a discrete change in  $\beta_2$ .

- Use  $\sigma = 0.01$  to see the effects and verify that you use the dummy variable correctly.
- Use a moderate or high  $\sigma$  to show how stochasticity makes the analysis realistic.

The illustrative worksheet uses

- ~ A geometric decay of  $-0.25$  for all development years.
- ~ An inflation rate of 35% for the first 9 years and 5% for the remaining 5 years.

*Take heed:* The illustrative worksheet uses an enormous change in inflation with a low  $\sigma$  to make the procedures clear. Once you understand the concepts, use realistic figures.

*Jacob:* The simulation has 15 years. Do you mean the first 9 years or the first 10 years?

*Rachel:* The first year (calendar year = 0) has no inflation. The first inflation rate is from the first calendar year to the second calendar year.

- It is easy to make errors when indexing years and trends.
- Indexing errors make your regression seem incorrect, but they are easy to fix.

- To spot errors, run a regression using a dummy variable with  $\sigma = 0$ . If the residuals are not all zero, you have an indexing error.

*Jacob:* Are these suitable parameters for the student project?

*Rachel:* With these parameters, the change in the inflation rate is clear in the residual plot.

The illustrative worksheet has  $\beta_2 = 35\%$  and  $\beta_{2B} = 5\%$ .

- First compute the Y values (logarithms of paid losses) in each cell *with no stochasticity*.
- The starting value for Y is 10 (development period = 0 and calendar year = 0) in the illustrative workbook; you may choose a more realistic base.

We can compute the cell values several ways: manual, IF statement, and VBA macro.

- Manual: write two cell formulas: one for calendar years 0 - 9 and one for years 10 - 14.
- IF statement: =IF(year < 10, statement 1, statement 2).
- VBA macro: Let the macro write the cell formula

The VBA macro for this illustrative worksheet reads:

```
Range(ActiveCell, ActiveCell.End(xlDown)).Select
For Each Ocell In Selection
  If Ocell.Offset(rowoffset:=0, columnoffset:=-1) < 10 Then
    Ocell.FormulaR1C1 = "=alpha+beta1*RC[-2]+beta2*RC[-1]"
  Else: Ocell.FormulaR1C1 = "=alpha+beta1*RC[-2]+beta2*9+beta2b*(RC[-1]-9)"
  End If
Next Ocell
```

The macro computes the expected Y values. Select the range of Y values several ways:

1. Select the range of expected Y values. You may name the column of expected Y values as *Expected* and click on Expected in the Name box.
2. Place the cursor on the first cell in the column of expected Y values and supply the number of data points, such as 120 for a 15 × 15 loss triangle.
3. If the column of expected Y values already has values, place the cursor in the top cell. The macro takes that cell and the other cells with values below it.

With the first two methods, the column of expected Y values may be blank or have values. For the third method, the column can not be blank.

## EXCEL IF STATEMENTS

{The project templates use several built-in functions and add-ins: if, offset, index, match, average, stdev, solver, regression. We explain these built-in functions so that you do not waste time with coding errors. If you are familiar with Excel, skip these explanations.}

The Excel IF statement gives one value if the condition is true and another if it is false.

*Illustration:* On the illustrative worksheet,

- The development period is in Column C.
- The calendar year is in Column D.
- The square of the calendar year is in Column E (not used for a discrete change).
- The expected Y values are in Column F.

Data values are on Rows 22-141 for a 15×15 loss triangle. The worksheet uses names.

- Alpha is the intercept.
- Beta1 is the development period trend.
- Beta2 is the calendar year trend for Year 0 – 9.
- Beta2b is the calendar year trend for Year 10 – 14.

If the inflation rate is constant (beta2b = beta2), the formula in Cell F22 is:

$$= \text{alpha} + \text{beta1} * \text{B22} + \text{beta2} * \text{C22}$$

If the inflation rate is **not** constant (beta2b ≠ beta2), we use this formula if the value in Column C (the calendar year) is less than 10. Otherwise we use

$$= \text{alpha} + \text{beta1} * \text{B22} + \text{beta2} * 9 + \text{beta2b} * (\text{C22} - 9)$$

We write the cell formula as

=IF(C22<10, alpha + beta1 \* B22 + beta2 \* C22, alpha + beta1 \* B22 + beta2 \* 9 + beta2b \* (C22 - 9) )

*Take heed:* Use whatever method you are comfortable with. The illustrative worksheets provide cell formulas and macros so you can focus on the statistical analysis.

The loop in this macro calculates the Y value before adding stochasticity. For each cell in Column E, we examine the calendar year in Column D.

- If the calendar year is < 10, the inflation rate for each year is  $\beta_2$ .
- If the calendar year is > 9, the inflation rate is  $\beta_2$  for the first 9 years and  $\beta_{2B}$  afterwards.

*Jacob:* Is the macro off by one year? For calendar year 10, don't we have 8 years of  $\beta_2$  inflation (years 1 through 9) and then one year of  $\beta_2$  inflation?

*Rachel:* We use an origin of zero, not of one. When you calculate the inflation and geometric decay, check to make sure your origin is consistent with your calculation.

Columns F, G, and H add stochasticity to the simulation. This is the same for a stable as for a changing inflation rate. We use the Excel *REGRESSION* add-in to compute the ordinary least squares estimators and residual plot.

The ordinary least squares estimator for the geometric decay is close to  $-25\%$ . With a  $\sigma$  of 0.01, the estimator should be between  $-24.8\%$  and  $-25.2\%$ . The estimator has an expected value of  $-25\%$ , so a value of  $-25\%$  doesn't necessarily mean everything is correct, but a value of  $-22\%$  or  $-28\%$  means something is wrong.

The ordinary least squares estimator for the inflation rate is close to the average figure. We have  $\frac{1}{2} \times 9 \times 10 - 1 = 54$  observations with 35% and  $120 - 55 = 65$  observations with a mix of 35% and 5%. We get an estimator between 21% and 22%.

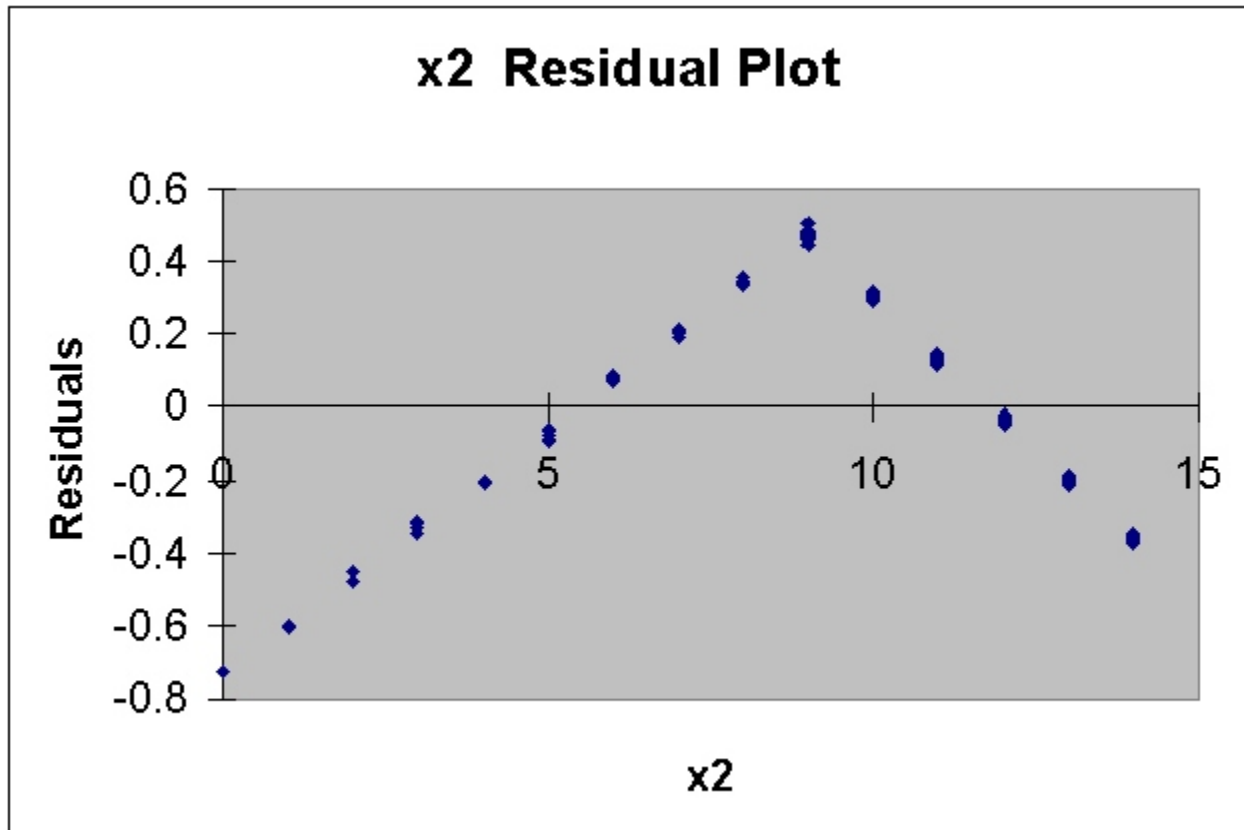
*Jacob:* How do we verify this figure?

*Rachel:* We have two observations with one calendar year of 35% inflation; three observations with two calendar years of 35% inflation; ...; and ten observations with nine calendar years of 35% inflation. We then have 65 observations with nine years of 35% inflation and 1, 2, 3, 4, or 5 years of 5% inflation. Over the full 15 calendar years, we have 9 years of 35% inflation and 5 years of 5% inflation.

*Jacob:* The  $t$  statistic is very high and the  $p$ -value is almost zero. This normally means that the estimator is correct. But we know that the estimator is not correct here; what happened?

*Rachel:* The  $t$  statistic says that *if the inflation rate is constant, the ordinary least squares estimator is not zero*. The estimated value is very close to its true value. Since the inflation rate is not constant, the  $t$  statistic is no help. The  $p$ -value is also no help, since the inflation rate is not constant.

We examine the residual plot, which appears below:



With low stochasticity, the residual plot appears as two line segments. For the first nine years, the actual inflation rate of 35% is greater than the fitted inflation rate of 21.6%. The residual is the actual Y value minus the fitted Y value. The actual Y values increase more rapidly than the fitted Y values, so the residuals *increase* each year. The stochasticity is so low that all the residuals by calendar year have about the same value, so we can read the average residual directly from the plot.

*Jacob:* Can we verify the figures in the residual plot?

*Rachel:* The residual is  $-0.75$  for calendar year = 0. For calendar year = 1, the actual inflation rate is 35% and the fitted inflation rate is about 21.6%, so the average residual is  $-0.75 + 0.35 - 0.216 = -0.616$ . The same 0.14 increase in the average residual occurs each calendar year.

*Jacob:* What happens after the tenth calendar year?

*Rachel:* For the last five years, the actual inflation rate of 5% is less than the fitted inflation rate of 21.6%. The actual Y values increase less rapidly than the fitted Y values, so the residuals *decrease* each year.

*Jacob:* Is the rate of decrease the same as the rate of increase?

*Rachel:*  $35\% - 21.6\% = 13.40\%$ ;  $21.6\% - 5\% = 16.60\%$ . The rate of decrease is slightly more rapid than the rate of increase. As you work through a statistical project, pause at each step to verify that your results are expected.

*Jacob:* Why do the average residuals start below zero, at  $-0.75$  for the first calendar year? At the first calendar year, we have not experienced any inflation yet, so there should be no difference between the fitted and actual values? Shouldn't the residuals start at zero, increase for 9 years, then decrease for 5 years back to zero?

*Rachel:* The average of all 120 residuals must be zero. This is a constraint of linear regression. You have described the pattern of the residuals. That pattern gives a positive overall average. All the residuals are moved down so that the overall average is zero. We examine the slope of the line connecting the average residuals by calendar year, not their absolute value.

*Jacob:* Is the average of the 15 average residuals exactly zero?

*Rachel:* The average of the 15 averages is not the average of the 120 observations, since the 15 averages do not have the same number of observations. Verify with the average of the 120 observations, not the average of the 15 averages. If the average of the 120 residuals is not zero, you made an error. This project template says how you can verify your work in each step. You don't have to write up each verification for the write-up. As you work through the project, jot down what you are doing. When you are done, this log (these jottings) are the write-up. You don't need to submit the entire log; just include the relevant points.

The average of the 120 residuals is a weighted average of the 15 averages, where the weights are the number of data points in each average.

*Jacob:* What happens when we make the simulation more realistic?

*Rachel:* The large difference from 35% to 5% causes a large change in the slope of the two line segments. With a discrete change in the inflation rate, both parts of the residual plot are straight lines. When we make the simulation more realistic, the following happens:

- As the change in the inflation rate becomes smaller, such as 25% to 15%, both line segments become flatter (more horizontal). If the change in the inflation rate is small, such as 22% to 18%, it is hard to distinguish the two line segments.
- If the change in the inflation rate is continuous, such as one percentage point each year, the residual plot becomes a smooth curve.
- As the stochasticity increases (as  $\sigma$  becomes larger), the average residuals by calendar year differ from their expected values.
  - The two parts of the residual plot no longer have constant slopes. With a change from 25% to 15% and a high stochasticity, the upward sloping line segment may have a change of  $+0.09$  one year and  $-0.02$  the next year. This is especially true for the early calendar years, which have few observations.

- With a continuously changing inflation rate and a high stochasticity, the residual plot is a jagged line that is hard to interpret.

*Jacob:* What values do we use for the student project?

*Rachel:* For the student project, we begin with a large, discrete change and low stochasticity. Once you understand what we examine, reduce the change in the inflation rate and see how the slopes change. Increase the stochasticity and see how the random fluctuations obscure the residual plot. For the final version, choose values with moderate stochasticity but still a clear pattern in the residual plot.

*Jacob:* What else affects the residual plot?

*Rachel:* If you comfortable with Excel, you can use 20 years with quarterly observations. This gives an 80 by 80 square, for a loss triangle of  $\frac{1}{2} \times 80 \times 81 = 3,240$  observations. A large number of observations lets you see the residual plot even with higher stochasticity.

*Take heed:* If you do calculations in VBA, the number of observations doesn't affect the work. If you do the work by cut and paste, you can't use large arrays.

## DUMMY VARIABLES

*Jacob:* Do we examine both discrete changes and continuous changes?

*Rachel:* Examine either one. If you are good with Excel, examine both. If you examine a discrete change, use a dummy variable to correct for the change, as follows.

Suppose the inflation rate changes at the end of the tenth year.

- If the calendar year is less than 10, the dummy variable (D) is zero.
- If the calendar year is 10 or more, the dummy variable is one.

The calendar year index starts at 0, so year 10 is the *eleventh* year.

- The regression equation is  $Y = \alpha_1 + \beta_1 \times X_1 + \beta_2 \times X_2 + D \times \alpha_2 + D \times \beta_3 \times X_2$ .
- The independent variables are  $X_1$ ,  $X_2$ , D, and  $D \times X_2$ .

*Jacob:* The starting value doesn't depend on the inflation rate. Why isn't  $\alpha_2$  equal to zero?

*Rachel:*  $\alpha_2$  is 9 years of the difference in the inflation rates:  $9 \times (\beta_2 - \beta_3)$ . Rewrite the regression equation to get rid of the variable  $\alpha_2$ . The three independent variables are combinations of  $X_2$  and D (along with  $X_1$ ). The student project requires you to work out the proper combinations, based on the simulation you use.

*Illustration:* Suppose inflation is 25% in the first ten years and 5% in the last five years. For clarity, assume  $\beta_1 = 0$ , so we can ignore  $X_1$ .

- For  $X_2 = 0$  through 9, accumulated inflation is  $X_2 \times 25\% = X_2 \times \beta_2 \Rightarrow \beta_2 = 25\%$
- For  $X_2 = 10$  through 14, accumulated inflation is  $9 \times 25\% + (X_2 - 9) \times 5\%$ .

Since  $D = 1$ , this equals  $\beta_2 \times X_2 + \alpha_2 + \beta_3 \times X_2$ . We solve

- $\beta_2 \times X_2 + \beta_3 \times X_2 = 5\% \times X_2 \Rightarrow \beta_2 + \beta_3 = 5\% \Rightarrow \beta_3 = -20\%$ .
- $\alpha_2 = 9 \times 25\% - 9 \times 5\% = 9 \times 20\%$ .

*Take heed:* It is easy to err and use an incorrect parameter or subscript. To verify your work, set  $\sigma = 0$  so the residuals are all zero. Check that the simulated Y values are the expected values. Once you understand the method, use realistic figures.

*Jacob:* How does the spreadsheet change for three independent variables?

*Rachel:* Add a column for the third independent variable,  $(X_2)^2$ .

- On the illustrative worksheet,  $X_1$  is in Column C and  $X_2$  is in Column D.
- Use Column E for  $(X_2)^2$ .



- If the first observation is on Row 13, type " $=D13^2$ " in Cell E13.
- Copy Cell E13 to the rest of this column.

For the regression analysis, treat Column E as a separate explanatory variable.

*Jacob:* One candidate used 20% inflation for the first 10 years and 10% inflation for the last 5 years. She used Excel's *IF* built-in function:

- ~ if calendar year  $< 10$ ,  $Y = \alpha - \beta_1 \times \text{development year} + 20\% \times \text{calendar year}$
- ~ if calendar year  $> 9$ ,  $Y = \alpha - \beta_1 \times \text{development year} + 10\% \times \text{calendar year}$

The residual plot had two upward sloping line segments, not a V shape. What is wrong?

*Rachel:* Look at her cumulative inflation by calendar year. The first row is the calendar year index and the second row is the cumulative inflation.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	1.0	1.1	1.2	1.3	1.4

What she wanted is

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	1.9	2.0	2.1	2.2	2.3

*Take heed:* This is a common error for this project template. Verify your work with  $\sigma = 0$ .