# Exercise 1: Geometric Distributed Lag (GDL) Parameters

In this exercise, you will investigate the performance of an econometric model of residential energy demand as a function of fuel price and gross domestic product (GDP). You will experiment with different model specifications (functional forms) and techniques for estimating parameters. The goal is to develop a basic understanding of the influence of various parameters and the considerations involved in using such a model with national data.

1. Set-up and Orientation

The first step in preparing for the exercise is to open Microsoft Excel and ensure that the Analysis ToolPak add-in is activated. The Analysis ToolPak is a library of Excel extras that support quantitative analysis (including ordinary least squares [OLS] regression, which you will use in this exercise). The procedure for activating the ToolPak depends on the version of Excel you have installed. In Excel 2010 and Excel 2013, click on the File tab, then select Options. A window for Excel Options should pop up. Select Add-Ins, then choose “Excel Add-ins” in the Manage field and press the Go… button.



In the Add-Ins window, check the box for Analysis ToolPak and press OK. Back in the main Excel window, click on the Data tab. On the far right you should now see a button for Data Analysis.  If these steps do not work for you, please check with the instructors.

Now open the workbook for Exercise 1 (“exercise\_01\_kz.xlsx”). You should see two tabs. The Demand Model tab has several parts:

* Historical energy demand, GDP, and average fuel price in the **Education services** sector in Kazakhstan are shaded in gray.
* The cells below the gray area show projected demand, GDP, and average fuel price. Initially, there is no projected demand (the value is set to 1) because you have not yet specified parameters for the econometric model.
* Parameters are entered in the Model Coefficients box (header shaded in green).
* A graph shows the energy demand output by the model.

On the Parameter Estimation tab, historical demand, GDP, and fuel price are reproduced, and the natural logarithms of these variables are given. The Year column appears twice to facilitate using it with the Analysis ToolPak regression function (this will be demonstrated in Section 3 below). The tab also contains two scatterplots of ln(Energy Demandt) versus ln(GDPt) and ln(Average Fuel Pricet), respectively. Both show values from the historical period.

1. Graphical Parameter Estimation

As a first attempt to define parameters for the demand model, you will estimate them graphically using the scatterplots on the Parameter Estimation tab. Because the plots show the natural logarithms of demand and the two main demand drivers, GDP and price, the slope of the best-fit line on each is an estimate of the elasticity of demand with respect to the driver. These elasticities can be directly entered as parameters for the econometric model. Note that in this case, each elasticity will be determined *without* controlling for the other driver: for example, the GDP elasticity of demand will be estimated without controlling for fuel price. This approach mimics the results you might get if you used elasticities from the literature in the model. Often such elasticities are based on the simple correlation between two variables and do not control for other factors.

Find the best-fit line in each scatterplot and estimate its slope.

***Hint:*** *You can use Excel’s Chart Tools to add a linear trendline to each graph and display its slope.*

Then go back to the Demand Model tab, and enter the two slopes as the coefficients for GDPt and Average Fuel Pricet. Observe that because you are leaving the coefficients for Energy Demandt-1 and Year as 0, the demand model does not include these terms. Thus, you are projecting demand using a simple model based on contemporaneous GDP and price (without any lags or yearly trend).

You should obtain a demand graph like the following.



The projections are still unrealistically high because you have not specified a constant term for the model. Experiment with different values for the constant term (i.e., enter different values in cell L5) until you arrive at a projection whose 2013 value is relatively consistent with the historical value from 2012. Now you should see an upward-sloping projection like this.



Notice the sensitivity of the model to small changes in the constant term. Try varying the GDP and price elasticities as well, and assess the model’s response.

1. Parameter Estimation Using Regressions

Now you will estimate parameters for a few variants of the econometric model using OLS regressions. Note that you will be conducting a basic regression analysis only—using OLS to obtain point estimates of model coefficients. A more thorough analysis including regression diagnostics is beyond the scope of this exercise.

Return to the Parameter Estimation tab. To begin, you will perform a regression for the simple, no-lags model you worked with in Section 2. Click on Excel’s Data tab and press the Data Analysis button. In the Data Analysis pop-up window, choose “Regression” and press OK. A Regression window should appear. Select the data in the ln(Energy Demandt) column as the Input Y Range and the data in the ln(GDPt) and ln(Average Fuel Pricet) columns as the Input X Range. Include the column headers in the selection, and check the Labels box. Specify that you want the regression output in a New Worksheet Ply.



When you press the OK button, Excel should carry out the regression and print the results in a new workbook tab. In this tab, find the estimated coefficients for ln(GDPt) and ln(Average Fuel Pricet) as well as the estimated intercept (i.e., the constant term). Notice how much each coefficient estimate has changed by controlling for the other driver.

Transfer the estimated coefficients to the Demand Model tab and review the results. You should obtain a projection that resembles the projection in Section 2 even though the elasticity parameters are somewhat different.



Now re-run the regression, but this time include the lagged demand term [ln(Energy Demandt-1)] in addition to GDP and price. In this way you will be estimating a GDL model for GDP and price (without an exogenous time trend).

***Hint:*** *Because there is no lagged demand value for 2000, you must estimate the model using data from 2001-2012 only. Adjust your selections for Input Y Range and Input X Range accordingly, and clear the Labels check box since the input ranges for the regression function must each be contiguous.*



Switch to the new tab containing the regression output, and copy the estimated coefficients into the Demand Model tab. Note that the x variables shown in the regression output are listed in the same order as on the Parameter Estimation tab (GDP first, fuel price second, lagged demand third). You should find that this variant of the model projects lower demand growth than the previous one, as illustrated below.



Next conduct one final regression for the GDL model plus a yearly trend term. Include ln(GDPt), ln(Average Fuel Pricet), ln(Energy Demandt-1), and Year as x variables, and base the regression on data from 2001-2012. After entering the estimated coefficients in the Demand Model tab, you should see a projection like the following.

This projection is clearly too low, illustrating the danger of assuming that a trend based on recent history will continue indefinitely! It also raises a corollary question: if this projection is unreasonable, which other projection should be used?

The answer to this question depends on statistical testing and modeling judgment. In terms of statistics, if you wanted to use this demand model for policy analysis, you would certainly assess the goodness-of-fit of each model variant and the statistical significance of its parameter estimates. You should also check for violations of the OLS assumptions using standard regression diagnostics. These steps might not yield a conclusive result, though, in which case the choice could be based on input from the modeling team or other stakeholders. Which projection is the most plausible over the long run? If two or more seemed possible, each could be evaluated as a separate scenario. Such an approach acknowledges that no single model variant or set of parameters can predict the future—each is merely an extrapolation of what could happen in the future given certain assumptions.

For the purposes of this workshop, we will assume that the GDL model without a trend term is considered the most realistic. In Exercise 2, you will implement this model in LEAP.