Purchasing power parity: is it true?

The principle of *purchasing power parity* (PPP) states that over long periods of time exchange rate changes will tend to offset the differences in inflation rate between the two countries whose currencies comprise the exchange rate. It might be expected that in an efficient international economy, exchange rates would give each currency the same purchasing power in its own economy. Even if it does not hold exactly, the PPP model provides a benchmark to suggest the levels that exchange rates should achieve. This can be examined using a simple regression model:

$$\text{Average annual change in the exchange rate} = \beta_0 + \beta_1 \times \text{Difference in average annual inflation rates} + \text{random error.}$$

The appropriateness of PPP can be tested using a cross-sectional sample of countries; PPP is consistent with $\beta_0 = 0$ and $\beta_1 = 1$, since then the regression model becomes

$$\text{Average annual change in the exchange rate} = \text{Difference in average annual inflation rates} + \text{random error.}$$

Data from a sample of 41 countries are the basis of the following analysis, which covers the years 1975 – 2012. The average annual change in exchange rate is the target variable (expressed as U.S. dollar per unit of the country's currency), calculated as the difference of natural logarithms, divided by the number of years and multiplied by 100 to create percentage changes; that is,

$$\frac{\ln(2012 \text{ exchange rate}) - \ln(1975 \text{ exchange rate})}{37} \times 100\%.$$

(Based on the properties of the natural logarithm, this is approximately equal to the proportional change in exchange rate over all 37 years divided by 37, producing an annualized figure.) The predicting variable is the estimated average annual rate of change of the differences in wholesale price index values for the United States versus the country. The data were derived and supplied by Professor Tom Pugel of New York University's Stern School of Business, based on information given in the *International Financial Statistics* database, which is published by the International Monetary Fund. I am also indebted to Professor Pugel for sharing some insights about the economic aspects of the data here, although all errors are my own.

First, let's take a look at the data. Here is a scatter plot of the two variables:

Scatterplot of Exchange rate change vs Inflation difference
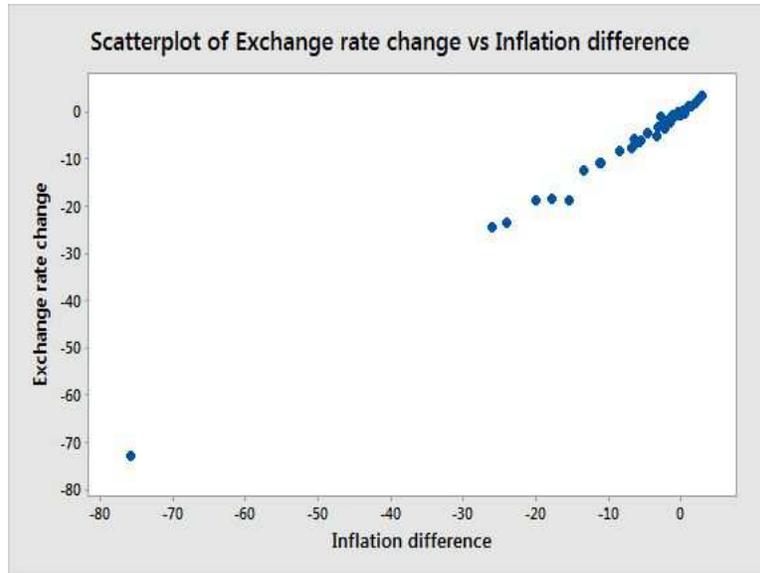
There appears to be a strong linear relationship here, but with a couple of caveats. There is an incredibly obvious point far away from all of the others, which turns out to be Brazil. There is also a point that appears to be a bit below the regression line, which turns out to be Mexico. Let's ignore this for now and do a least squares regression with change in exchange rate as the dependent (target) variable, and difference in inflation rates as the independent (predicting) variable:

```
Analysis of Variance

Source                  DF   Adj SS   Adj MS  F-Value  P-Value
Regression               1  6498.98  6498.98  8035.06    0.000
  Inflation difference   1  6498.98  6498.98  8035.06    0.000
Error                   39    31.54     0.81
Total                   40  6530.52


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.899348  99.52%     99.50%      99.43%


Coefficients

Term                     Coef  SE Coef  T-Value  P-Value   VIF
Constant               -0.306    0.158    -1.94    0.059
Inflation difference   0.9680   0.0108    89.64    0.000  1.00
```

```
Regression Equation
```

```
Exchange rate change = -0.306 + 0.9680 Inflation difference
```

The regression is very strong, with an R^2 of 99.5% (that is, roughly 99.5% of the observed variability in exchange rate change is accounted for by inflation difference), and the F–statistic is very significant. The intercept, which is meaningful here, says that given that the annualized difference in inflation between a country and the U.S. is zero (that is, the country has the same inflation experience as does the U.S.), the estimated expected annualized change in exchange rates between the two countries is -0.31%; that is, the currency becomes devalued relative to the U.S. dollar. The slope coefficient says that a one percentage point change in annualized difference in inflation rates is associated with an estimated expected 0.968 percentage point change in annualized change in exchange rates. Although this sort of model would not be used to trade currency (it is based on long-term trends, while currency trading is based primarily on short-term ones), the standard error of the estimate of 0.9 tells us that this model could be used to predict annualized changes in exchange rates to within ±1.8 percentage points, roughly 95% of the time.

Note, by the way, that in all of the discussion above about regression coefficients and the standard error of the estimate the generic term "units" (as in "a one unit increase in annualized difference in inflation rates is associated with an estimated expected .968 unit decrease in annualized change in exchange rates") is **not** used when describing the interpretations of the estimates; rather, the **actual** units (in this case percentage points) are used. You should **always** interpret estimates using the **actual** unit terms of the variables involved, whether those are percentage points, dollars, years, or anything else.

The t–statistic given in the output for `Inflation difference` tests the null hypothesis of $\beta_1 = 0$, which is strongly rejected. There is certainly a relationship between the change in exchange rates and the difference in inflation rates. Of course, it is not the hypothesis of $\beta_1 = 0$ that we're really interested in here. PPP says that $\beta_0 = 0$; the t–statistic for that hypothesis, listed in the line labeled `Constant`, is given as -1.94, which is marginally statistically significant. That is, the foreign currencies appear to have appreciated less than would be predicted by PPP. PPP also says that $\beta_1 = 1$; the t–statistic for this can be calculated manually:

$$t = \frac{0.968 - 1}{.0108} = -2.96.$$

This is statistically significant at the usual Type I error levels, so we reject this hypothesis as well (the tail probability for this test is .0052). Thus, we are seeing violations of PPP with respect to both the intercept and the slope.

The following simple regression scatter plot illustrates the use of **confidence intervals** and **prediction intervals**. Consider again the estimated regression line,

Exchange rate change = -0.306 + 0.9680 Inflation difference.

Two possible uses of this line are: What is our best estimate for the true average exchange rate change for all countries in the population that have inflation difference equal to some value? What is our best estimate for the value of exchange rate change for one particular country in the population that has inflation difference equal to some value? Each of these questions is answered by substituting that value of inflation difference into the regression equation, but the estimates have different levels of variability associated with them. The confidence interval (sometimes called the confidence interval for a fitted value) provides a representation of the accuracy of an estimate of the *average* target value for all members of the population with a given predicting value, while the prediction interval (sometimes called the confidence interval for a predicted value) provides a representation of the accuracy of a prediction of the target value for a *particular* observation with a given predicting value.

Fitted Line Plot
Exchange rate change = - 0.3062 + 0.9680 Inflation difference
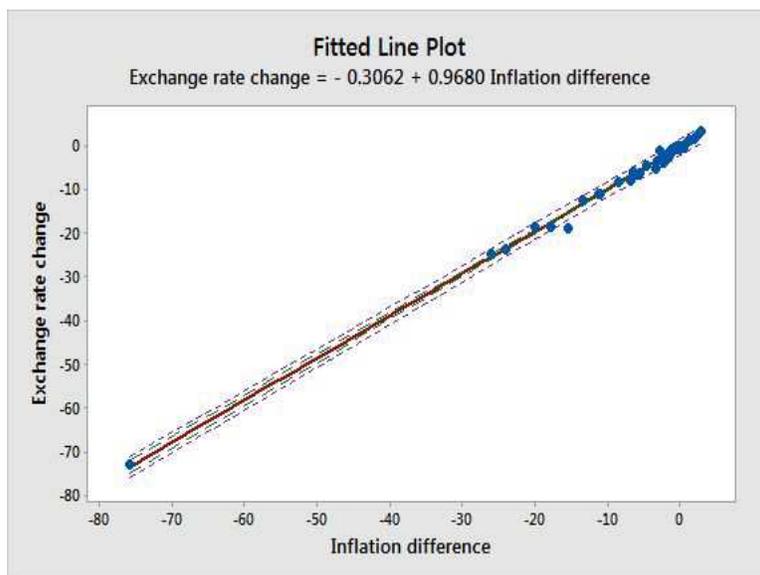
This plot illustrates a few interesting points. First, the pointwise confidence interval (represented by the inner pair of lines) is much narrower than the pointwise prediction interval (represented by the outer pair of lines), since while the former interval is based only on the variability of $\hat{\beta}_0 + \hat{\beta}_1 \times \texttt{Inflation difference}$ as an estimator of $\beta_0 + \beta_1 \times \texttt{Inflation difference}$, the latter interval also reflects the inherent variability of $\texttt{Exchange rate change}$ about the regression line in the population itself. Second, it should be noted that the interval is narrowest in the center of the plot, and gets wider at the extremes; this shows that predictions become progressively less accurate as the predicting value gets more extreme compared with

the bulk of the points (this is more apparent in the confidence interval than it is in the prediction interval), although we should be clear that what we mean by the "middle" is the middle of the points (more specifically, the intervals are narrowest at the mean value of the predictor). Finally, it is clear that the points noted earlier are quite unusual, with Brazil very isolated to the left and Mexico outside the 95% prediction interval.

The graph obviously cannot be used to get the confidence and prediction interval limits for a specific value of inflation difference, but `Minitab` will give those values when a specific value is given. As an example, here are the values when the inflation difference is -1.50, which is roughly the value for Norway:

```
Prediction for Exchange rate change

Regression Equation

Exchange rate change = -0.306 +  0.9680  Inflation difference



Variable              Setting
Inflation difference    -1.5



    Fit    SE Fit         95% CI                  95% PI
-1.75814  0.150881  (-2.06333, -1.45296)  (-3.60267, 0.0863819)
```

As we know must be true, the prediction interval is a good deal wider than the confidence interval; while our guess for what the *average* exchange rate change would be for *all* countries with inflation difference equal to -1.5 is $(-2.063, -1.453)$, our guess for what the exchange rate change would be for *one* country with inflation difference equal to -1.5 is $(-3.603, 0.086)$.

Diagnostic plots also can help to determine whether the points we noticed earlier are truly unusual, and to check whether the assumptions of least squares regression hold. Here is a "four in one plot" that gives plots of residuals versus fitted values, residuals in observation order, a normal plot of the residuals, and a histogram of the residuals:

5

Residual Plots for Exchange rate change
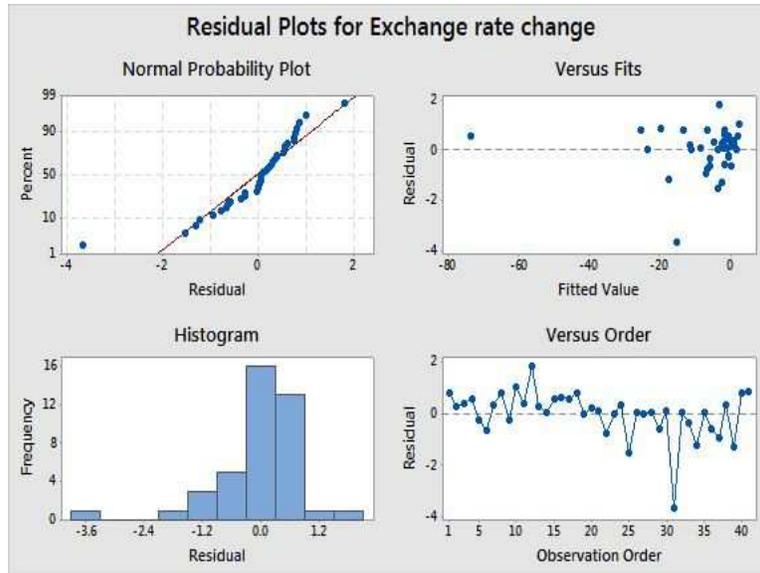
There are three unusual observations in these plots. In the plot of residuals versus fitted values one point is isolated all the way to the left of the plot, corresponding to a very unusually low fitted value (estimated exchange rate change). This is Brazil, which has an annualized inflation rate more than 75 percentage points higher than that of the United States. This is a *leverage point*, and it is a problem. A leverage point can have a strong effect on the fitted regression, drawing the line away from the bulk of the points. It also can affect measures of fit like R^2, t–statistics, and the F–statistic. What is going on here? Brazil had a period of hyperinflation from 1980 to 1994, a time period during which prices went up by a factor of roughly **1 trillion** (that is, something that cost 1 Real at the beginning of the 1980s cost roughly 1 trillion Reals in the mid-1990s). This hyperinflation was caused by an expansion of the money supply; the government financed projects not through taxes or borrowing but simply by printing more money, a crisis triggered by the worldwide energy crisis of the 1970s and political instabilities of the Brazilian military dictatorship. It is worth noting that this shows how changing the definition of "long term" can change the numbers for a country dramatically (the numbers for Brazil would be noticeably different if the starting time period was 1985 rather than 1975, and would be dramatically different if it was 1995 instead of 1975), because of the dependence of the annualized rates on the first and last values.

What should we do about this case? The unusual point has the potential to change the results of the regression, so we can't simply ignore it. We can remove it from the data set and analyze without it, *being sure to inform the reader about what we are doing.* That is, we can present results without Brazil, while making clear that the implications of the model don't apply to Brazil, or presumably to other countries with a similar unsettled economic situation.

What about the other two unusual values? We could examine them and see their effects on the regression as well, but because Brazil is so extreme, it is possible that once it is removed those points will not look as outlying, so we'll start by omitting Brazil only. Here is the scatter plot now:
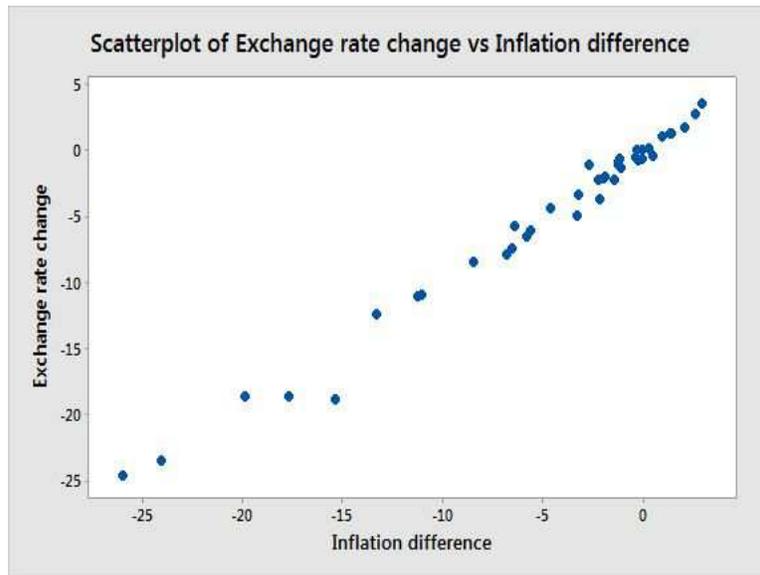
Mexico shows up as unusual, and Poland is marginally unusual as well. Here is regression output:

```
Regression Analysis: Exchange rate change versus Inflation difference

Analysis of Variance

Source                 DF   Adj SS   Adj MS  F-Value  P-Value
Regression              1  1982.27  1982.27  2478.84    0.000
  Inflation difference  1  1982.27  1982.27  2478.84    0.000
Error                  38    30.39     0.80
Total                  39  2012.66


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
0.894247  98.49%     98.45%      98.24%


Coefficients

Term                  Coef  SE Coef  T-Value  P-Value   VIF
Constant            -0.223    0.171    -1.30    0.201
```

```
Inflation difference  0.9880    0.0198    49.79    0.000  1.00
```

Regression Equation

```
Exchange rate change = -0.223 +  0.9880  Inflation  difference
```

The regression relationship is still very strong, and now neither the slope nor the intercept are statistically significantly different from the PPP-hypothesized values. Unfortunately, Mexico shows up as quite unusual in the residual plots:
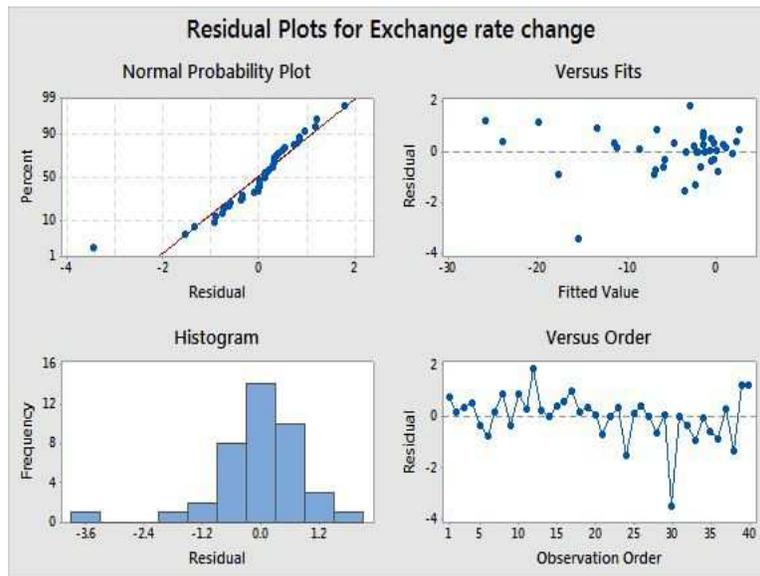

Mexico's story is similar to that of Brazil, but (in a sense) flipped around. At the beginning of the time period Mexico had an exchange rate that was arguably overvalued, because of an oil bubble related to the energy crisis (Mexico at that time was heavily dependent on oil production for national income and exports). Oil prices collapsed in the early 1980s, and Mexico had a severe international debt crisis. Thus, the Mexican Peso is weaker than expected relative to the U.S. dollar over the entire time period because it appeared to be stronger than it really was in 1975.

With an unusual response value given its predictor value, Mexico is an *outlier*, which is again a problem. In addition the points made earlier about the Brazil leverage point (that the outlier can have a strong effect on the fitted regression, measures of fit like R^2, t–statistics, and the F–statistic), we also must recognize that we can't very well claim to have a model that fits all of the data when this point isn't fit correctly!

We could now omit Mexico to see what effect it has had on the regression. Before we do that, however, we need to recognize a different problem with this analysis. The sample of 41 countries used here actually consists of two samples: a set of 16 industrialized (developed)

countries, and a set of 25 developing countries. Might the PPP relationship be different for these two groups? This is not an unreasonable supposition, particularly since PPP is based on the principle of easy movement of goods, services, and money across international boundaries, which we could imagine is more likely in developed countries than in developing ones. First, here are results for developed countries:
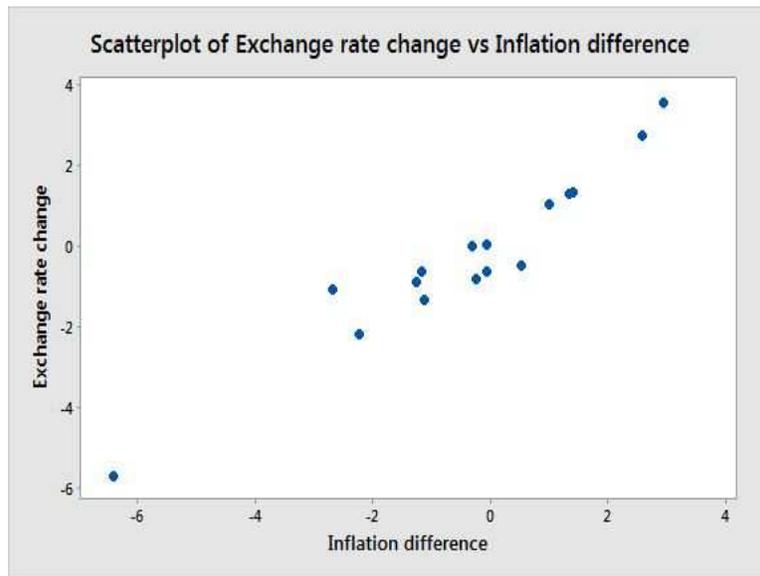
Greece is clearly an unusual observation (higher inflation and a weaker currency than those of the other developed countries), and the trials and tribulations of the Greek economy since 2007 make this unsurprising. We could pursue this further, but the sake of expediency I'll ignore it here. Here are the results of the regression for the developed countries:

Regression Analysis: Exchange rate change versus Inflation difference

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	61.850	61.8496	183.49	0.000
Inflation difference	1	61.850	61.8496	183.49	0.000
Error	14	4.719	0.3371		
Total	15	66.569			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.580583	92.91%	92.40%	90.91%

```
Coefficients
```

```
Term                    Coef  SE Coef  T-Value  P-Value   VIF
Constant               0.086    0.147     0.59    0.567
Inflation difference  0.9024   0.0666    13.55    0.000  1.00
```

```
Regression Equation
```

```
Exchange rate change = 0.086 + 0.9024 Inflation difference
```

Neither the slope nor the intercept are significantly different from the PPP-assumed values, but we should recognize that part of that comes from the relatively small sample size. Residual plots are similarly difficult to parse with a small sample, and Greece certainly does show up as unusual:
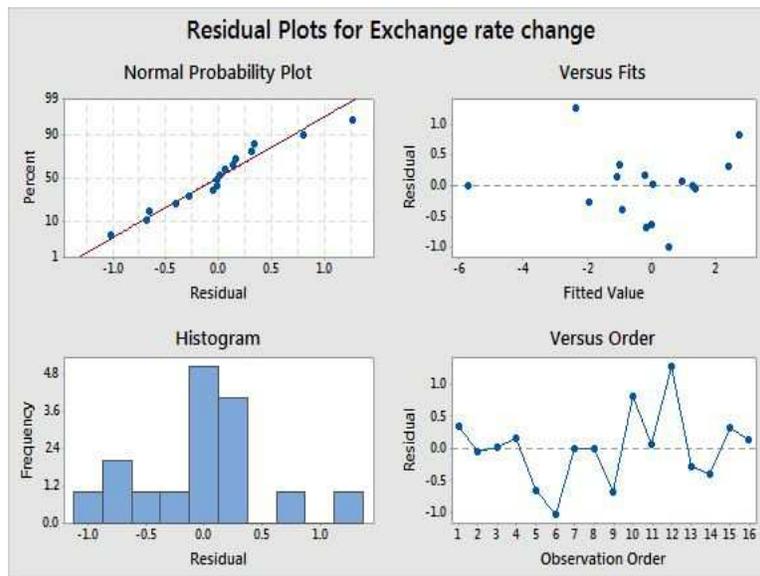
Here is a scatter plot for developing countries (I've already removed the mega-leverage point Brazil). Poland and especially Mexico show up as unusual, with currencies relative to the dollar weaker than inflation differences would imply.

Scatterplot of Exchange rate change vs Inflation difference
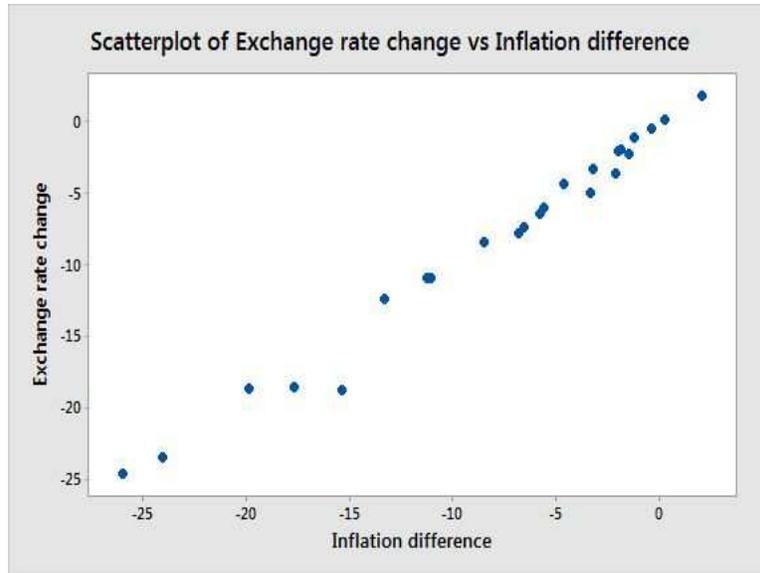
Here are the regression results. The intercept is significantly different from the PPP-hypothesized value of 0, but Mexico shows up as a clear outlier:

```
Regression Analysis: Exchange rate change versus Inflation difference

Analysis of Variance

Source                  DF   Adj SS   Adj MS  F-Value  P-Value
Regression               1  1309.89  1309.89  1359.33    0.000
  Inflation difference   1  1309.89  1309.89  1359.33    0.000
Error                   22    21.20     0.96
Total                   23  1331.09


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.981645  98.41%     98.33%      98.06%


Coefficients

Term                     Coef  SE Coef  T-Value  P-Value   VIF
Constant               -0.641    0.287    -2.23    0.036
Inflation difference   0.9637   0.0261    36.87    0.000  1.00


Regression Equation
```
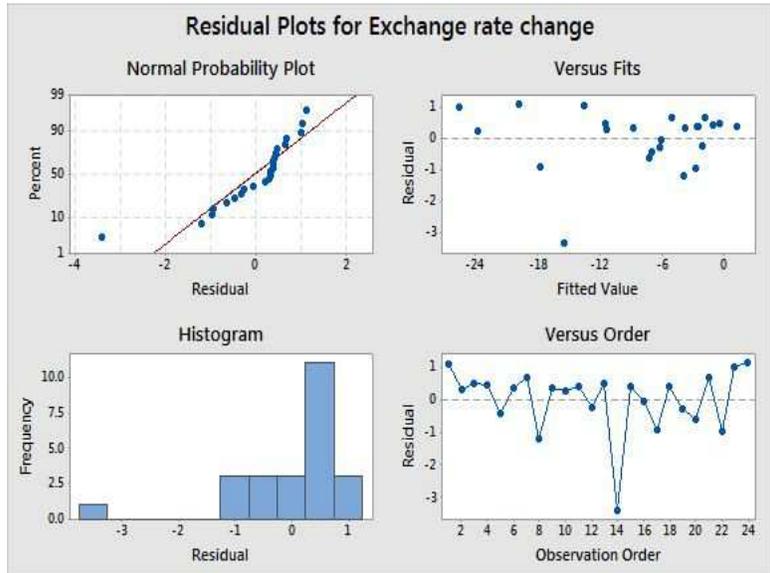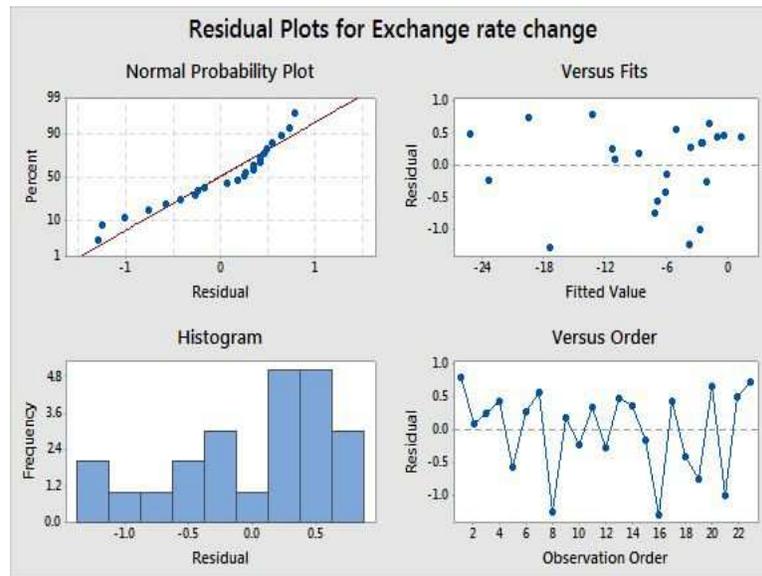
Exchange rate change = -0.641 + 0.9637 Inflation difference

Residual Plots for Exchange rate change

Would omitting Mexico change anything?

Regression Analysis: Exchange rate change versus Inflation difference

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1205.86	1205.86	2932.51	0.000
Inflation difference	1	1205.86	1205.86	2932.51	0.000
Error	21	8.64	0.41		
Total	22	1214.50			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.641253	99.29%	99.26%	99.14%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.641	0.188	-3.41	0.003	
Inflation difference	0.9442	0.0174	54.15	0.000	1.00

Regression Equation

```
Exchange rate change = -0.641 + 0.9442 Inflation difference
```

The strength of the relationship has increased, but so has the rejection of PPP (slope $t = -3.21$, intercept $t = -3.41$). The residual plots are only okay, but with such a small sample, there's not really much that would be done at this point.

Residual Plots for Exchange rate change

Thus, support for the principle of purchasing power parity is decidedly mixed here. For developed countries, changes in inflation difference do seem to be balanced by exchange rate changes, but it's difficult to be sure with such a small sample, and Greece is potentially problematic. For developing countries, the case for PPP is considerably weaker; depending on how we choose to handle Brazil and Mexico, there is evidence that both the intercept and slope are not consistent with PPP. We also see that PPP is not necessarily a robust principle in any event, in the sense that unusual economic or political conditions, particularly at the beginning or end of the time period in question, can have a strong effect on inflation and exchange rates.

Minitab commands

To construct a histogram, click on `Graph → Histogram`. Choose the type of histogram you want (Simple being the default). Enter the variable name under `Graph variables`. A technically more correct version of the plot can be obtained by first clicking on `Tools → Options → Individual graphs → Histograms`, and setting `Type of intervals` to `CutPoint`. Note that variable names can be picked up and placed in the dialog boxes by double clicking on them.

To construct a boxplot, click on `Graph → Boxplot` (Simple boxplot). Enter the variable name under `Graph variables`. To obtain descriptive statistics, click on `Stat → Basic Statistics → Display Descriptive Statistics`. Enter the variable name under `Variables`.

To construct a scatter plot, click on `Graph → Scatterplot`, and choose the type of plot wanted (Simple is the default). Enter the target variable for the regression under `Y variables` and the predicting variable under `X variables`.

To fit a linear least squares regression line, click on `Stat → Regression → Regression → Fit Regression Model`. Enter the target variable under `Responses:` and the predicting variable under `Continuous predictors:`. Click on `Graphs`. If you want standardized residuals, click `Standardized` in the drop-down menu under `Residuals for Plots:`. Under `Residual Plots` click `Normal plot of residuals`, `Residuals versus fits`, and `Residuals versus order` to create the three residual plots mentioned earlier. Clicking the radio button next to `Four in one` will give all three plots, along with a histogram of the residuals.

To calculate the two–tailed tail probability for an observed t–statistic, click on `Calc → Probability Distributions → t`. Click on the radio button next to `Cumulative probability`, and enter the appropriate value for the degrees of freedom next to `Degrees of freedom:`. Click on the radio button next to `Input constant:`, and input the absolute value of the observed t–statistic. The tail probability is $2 \times (1 - p)$, where p is the value given under `P(X <= x)`.

To calculate the critical value for a $100 \times (1 - \alpha)\%$ t–based confidence interval, click on `Calc → Probability Distributions → t`. Click on the radio button next to `Inverse cumulative probability`, and enter the appropriate value for the degrees of freedom next to `Degrees of freedom:`. Click on the radio button next to `Input constant:`, and input the number corresponding to $1 - \alpha/2$.

To create a regression plot with pointwise confidence and prediction intervals superimposed, click on `Stat → Regression → Fitted Line Plot`. Enter the target variable under `Response (Y):` and the predictor under `Predictor (X)`. Click on `Options`, and click on

`Display confidence interval` and `Display prediction interval` under `Display Options`.

You can identify the unusual point in the scatter plot using *brushing*. While the scatter plot is displayed, right click on it and click on `Brush`. A little hand will appear on the graph; when the hand points to the point you're interested in, click, and the case number will appear in a box that you can then read off. To omit the observation, the variables in your data set must be copied over to new columns, with the observation(s) omitted. Click on `Data → Copy → Columns to Columns`. Enter the variables being used under `Copy from columns:` and choose where you wish the new columns to go; enter new variable names (or columns) by unclicking the box next to `Name the columns containing the copied data` and entering variable names under the location you are copying to. Click on `Subset the data`; enter the case numbers or conditions that define the points to be omitted. You can omit the observation from all variables (and then continue to use the same variable names) by clicking on the case number(s) in the `Data` window and then pressing the *Del* button, but this omits the case permanently (you can get it back, of course, by reading in the data file again). You can create a new worksheet within your project that omits observations by clicking on `Data → Subset Worksheet`. Click the radio button next to `Specify which rows to exclude`, and define the condition or row numbers that identifies the rows you wish to omit in the panel below.

A prediction interval and a confidence interval for average y for a given predictor value are obtained as a followup to the regression fit. Click on `Stat → Regression → Regression → Predict` after you have fit your model. Enter the appropriate value(s) under `Enter individual values` to get the intervals. If you want to get intervals for more than one observation, you can put all of the variable values needed in columns, and put in the column names rather than individual value(s) after changing the drop-down menu to `Enter columns of values`.