The effect of Gross Domestic Product on food imports

In [49]:
```
#Load packages
import matplotlib.pyplot as plt
import matplotlib
matplotlib.style.use('ggplot')
import patsy
import pandas as pd
#import statsmodels as sm
#import statsmodels.api as sm
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols
import csv
import os
from matplotlib.gridspec import GridSpec
#Set defaults for graphs
pd.set_option('display.mpl_style', 'default') # Make the graphs a bit prettier
figsize(16, 8)
```

In [2]:
```
#Working directory
os.chdir('/Users/duccioa/CLOUD/C07_UCL_SmartCities/QuantitativeMethods/qm_cours
ewrk1')
```

Introduction

In this project, I investigate the determinants of food imports in countries based on the dataset Data_for_Coursework_1_Countries.csv, provided within the context of the course Quantitative Methods at UCL. A first glance at the data reveals an interesting sublinear power law relation between GDP and food imports. Countries with bigger economies import more food then countries with a smaller output but the increase is less important the more the GDP grows. At a second glance, the same relation holds if we take into account population but with an effect of economy of scale due to the population: richer countries import more than poorer ones but spend a smaller fraction of their income in food, bigger countries import less food per capita than smaller ones.

Analysis

In [3]:
```
countries = pd.read_csv('countries.csv', sep = ',', encoding = 'latin1')
countries.columns = ['X', 'Year', 'CountryCode', 'CountryName',
                     'Population', 'GDP', 'FoodImports', 'FuelImports']
```

In this project I aim at investigating the determinant of food import based on the available dataset, which contains data for 190 countries about Gross Domestic Product (GDP), population, food and fuel imports for the year 2005.

In [6]:
```
countries[:3]
```

Out[6]:

	X	Year	CountryCode	CountryName	Population	GDP	FoodImports	FuelImports	Pc
0	1	2005	ABW	Aruba	100031	1.160240e+12	97166150	32335285	11
1	2	2005	AFG	Afghanistan	24860855	6.275076e+09	528341972	461521897	17
2	3	2005	AGO	Angola	16544376	2.823370e+10	1075607744	48218538	16

In [7]:
```
countries['Population_log'] = log(countries['Population'])
countries['FoodImports_log'] = log(countries['FoodImports'])
countries['GDP_log'] = log(countries['GDP'])
countries['GDP_pc'] = countries['GDP']/countries['Population']
countries['FoodImports_pc'] = countries['FoodImports']/countries['Population']
countries['GDP_pc_log'] = log(countries['GDP_pc'])
countries['FoodImports_pc_log'] = log(countries['FoodImports_pc'])
```

Both plots present patterns of correlation with three big outliers: USA, India and China which have respectively very high GDP, population, GDP and population. GDP have a very strong correlation (0.92) with food import and the ln-ln plot strongly suggests a power law relation with values very concentrated on the regression line (the ln-ln plot with population shows higher variation than with GDP).

In [8]:
```
coords = countries[countries['CountryCode'] == 'CHN']
coords = coords.append(countries[countries['CountryCode'] == 'IND'])
coords = coords.append(countries[countries['CountryCode'] == 'USA'])
```

In [61]:

```
#POPULATION AND FOOD IMPORT : analysis and plot
gs = GridSpec(100,100,bottom=0.18,left=0.18,right=0.88)
fig1, ax1 = plt.subplots()
ax1 = fig1.add_subplot(gs[:,0:43])
ax2 = fig1.add_subplot(gs[:,50:99])

ax1.scatter(countries['Population'], countries['FoodImports'])
ax1.annotate(coords['CountryCode'][179],
            xy = (coords['Population'][179], coords['FoodImports'][179]))
ax1.annotate(coords['CountryCode'][34],
            xy = (coords['Population'][34], coords['FoodImports'][34]))
ax1.annotate(coords['CountryCode'][82],
            xy = (coords['Population'][82], coords['FoodImports'][82]))
ax1.annotate('Correlation: ' +
            str(round(countries['Population'].corr(countries['FoodImports'], me
thod='spearman'), 2)),
            xy = (coords['Population'][179]*1.3, coords['FoodImports'][179]*0.6
))

#Log plot
#fig2, ax2 = plt.subplots()
ax2.scatter(countries['Population_log'], countries['FoodImports_log'])
plt.title('Population and Food Imports: Scatter and Ln-ln plot')

#Regression line
mod_pop = ols(formula='FoodImports_log ~ Population_log', data=countries)
res_pop = mod_pop.fit()
par_pop = res_pop.params
ax2.plot(countries['Population_log'], par_pop[0] +
        par_pop[1]*countries['Population_log'], color = 'red')
```

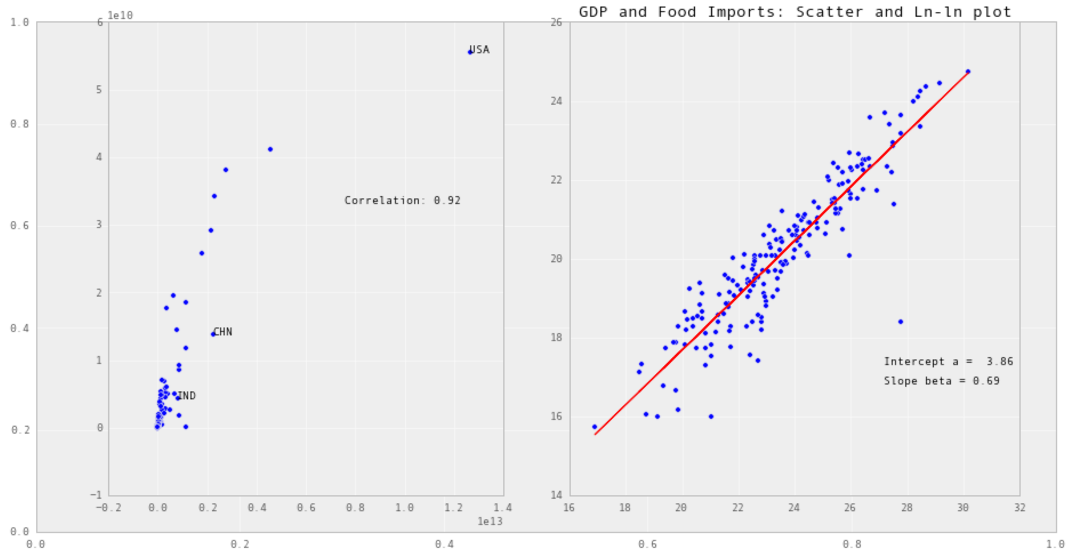
Out[61]:

```
[<matplotlib.lines.Line2D at 0x1146b8fd0>]
```

In [63]:
```
##GDP AND FOOD IMPORT : analysis and plot
#Plot
gs = GridSpec(100,100,bottom=0.18,left=0.18,right=0.88)
fig1, ax1 = plt.subplots()
ax3 = fig1.add_subplot(gs[:,0:43])
ax4 = fig1.add_subplot(gs[:,50:99])


ax3.scatter(countries['GDP'], countries['FoodImports'])
ax3.annotate(coords['CountryCode'][179],
             xy = (coords['GDP'][179], coords['FoodImports'][179]))
ax3.annotate(coords['CountryCode'][34],
             xy = (coords['GDP'][34], coords['FoodImports'][34]))
ax3.annotate(coords['CountryCode'][82],
             xy = (coords['GDP'][82], coords['FoodImports'][82]))
ax3.annotate('Correlation: ' +
             str(round(countries['GDP'].corr(countries['FoodImports'],
                                              method='spearman'), 2)),
             xy = (coords['GDP'][179]*0.6, coords['FoodImports'][179]*0.6))
plt.title('GDP and Food Imports')
#Log plot
ax4.scatter(countries['GDP_log'], countries['FoodImports_log'])
plt.title('GDP and Food Imports: Scatter and Ln-ln plot')
#Regression line
mod_gdp = ols(formula='FoodImports_log ~ GDP_log', data=countries)
res_gdp = mod_gdp.fit()
par_gdp = res_gdp.params
ax4.plot(countries['GDP_log'], par_gdp[0] +
        par_gdp[1]*countries['GDP_log'], color = 'red')
ax4.annotate('Intercept a =   ' + str(round(par_gdp[0], 2)),
             xy = (coords['GDP_log'][179]*0.9, coords['FoodImports_log'][179]*0.
7))
ax4.annotate('Slope beta = '  + str(round(par_gdp[1], 2)),
             xy = (coords['GDP_log'][179]*0.9, coords['FoodImports_log'][179]*0.
68))
```

Out[63]: <matplotlib.text.Annotation at 0x1141268d0>



Based on the graphs above the relation between the outcome Food Imports (Y) and the predictor GDP (X) is as following:

$$Y = aX^{\beta}$$

In [66]:
```
df_gdp = pd.DataFrame()
df_gdp['Coef'] = res_gdp.params
df_gdp['P-values'] = res_gdp.pvalues
print 'R-squared: ' + str(round(res_gdp.rsquared, 2))
df_gdp

#print res_gdp.summary()
```

```
R-squared: 0.84
```

Out[66]:

	Coef	P-values
Intercept	3.859762	3.058295e-12
GDP_log	0.691211	3.268875e-77

With a = 3.86 and α = 0.69, the model explains almost 85% of the variation in food import (R-squared = 0.84), which is a good fit. Being the slope less than 1, the relation is sublinear and the impact of GDP on food import declines with the increase of GDP. However, we would expect both GDP and food imports to be strongly increasing in the population. Therefore, it is worth to check that this strong correlation is not driven by their dependence on population. To do so I run the following regression:

$$ln(y) = \alpha_1 + \alpha_2 ln(x) + \alpha_3 ln(L)$$

With $y = \frac{Y}{L}$ (food import per capita) and $x = \frac{X}{L}$ (GDP per capita).
It is worth to note that this formulation encompasses the one above (for $\alpha_2 = 1 - \alpha_3 = \beta$ it is identical, see appendix for details).

In [69]:
```
mod = ols(formula='FoodImports_pc_log ~ GDP_pc_log + Population_log', data=coun
tries)
res = mod.fit()
df = pd.DataFrame()
df['Coef'] = res.params
df['P-values'] = res.pvalues
print 'R-squared: ' + str(round(res.rsquared, 2))
df
```

```
R-squared: 0.81
```

Out[69]:

	Coef	P-values
Intercept	3.860655	3.237017e-12
GDP_pc_log	0.674037	6.845421e-52
Population_log	-0.299762	5.349538e-25

This gives us the following equation:

$$y = ax^{\alpha_2} L^{\alpha_3}$$

With $y = \frac{Y}{L}$ (food import per capita) and $x = \frac{X}{L}$ (GDP per capita), a = 47.49, $\alpha_2 = 0.67$ and $\alpha_3 = 0.3$.
The model has an R-squared = 0.807 and the coefficients are all statistically significant.

The equation above suggests that, holding population constant, the food import per capita increases in a sublinear fashion with the GDP per capita and the increase is less and less pronounced for more productive countries (higher GDP per capita). On the other hand, holding GDP per capita constant, countries with more population import less food per person. The first relation is probably explained by the fact that as people become richer, they consume more food but spend a smaller fraction of their income in buying it. The other relation could be explained as an economy of scale, where increase in population make food allocation more efficient and reduces the relative need of importing food.
The model explains ~80% of the variation and other factors would be worth investigating. To name a few, I would think food trade balance, levels of productivity and employment in agriculture would be a good start to improve the comprehension of the dynamics behind food imports.

Appendix

Identity of the power law

The power law suggested by the graph is

$$\left[Y = aX^{\beta}\right]$$

Dividing by the population L, we can write:

$$\frac{Y}{L} = \frac{aX^{\beta}}{L} = \frac{aX^{\beta}}{L^{\beta}L^{1-\beta}} = a(\frac{X}{L})^{\beta}L^{1-\beta}$$

If I write $y = \frac{Y}{L}$ as the Food Imports per capita and $x = \frac{X}{L}$ as the GDP per capita, I can say that the first equation is equivalent to saying:

$$\left[y = ax^{\beta}L^{1-\beta}\right]$$