Daeyoung Lim

Economic Time Series

Fall Semester, 2016

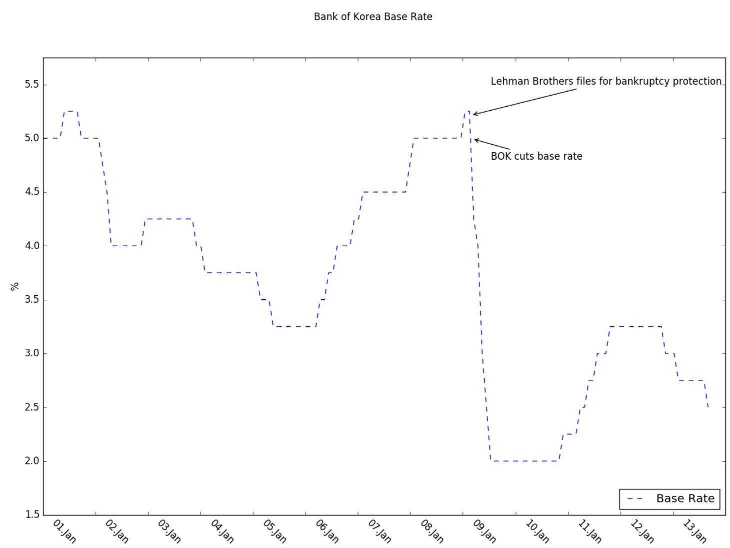
1. Replicate Figures 1,5,6 and 7 in the Matlab exercise. In doing so, make the following modifications to the figures.

   (1) In figure 1, use "plot" instead of "stairs", and use a dotted line instead of a solid line.

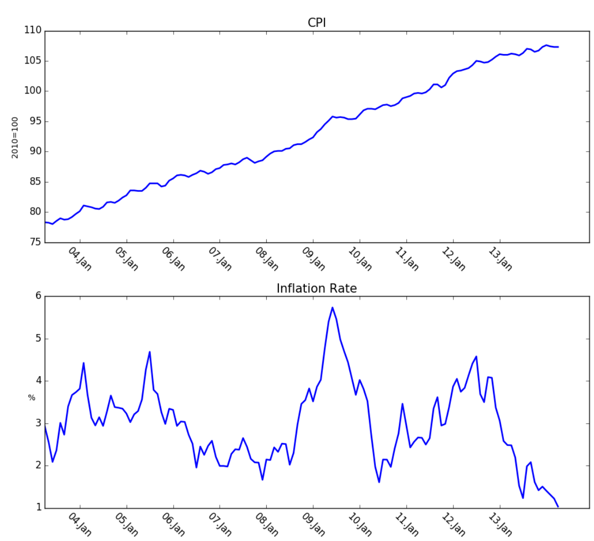
   (2) In figure 5, plot the data from Jan 2004 only and remove all grid lines.

   (3) Put figures 6 and 7 on a common figure frame in 2×2 format.

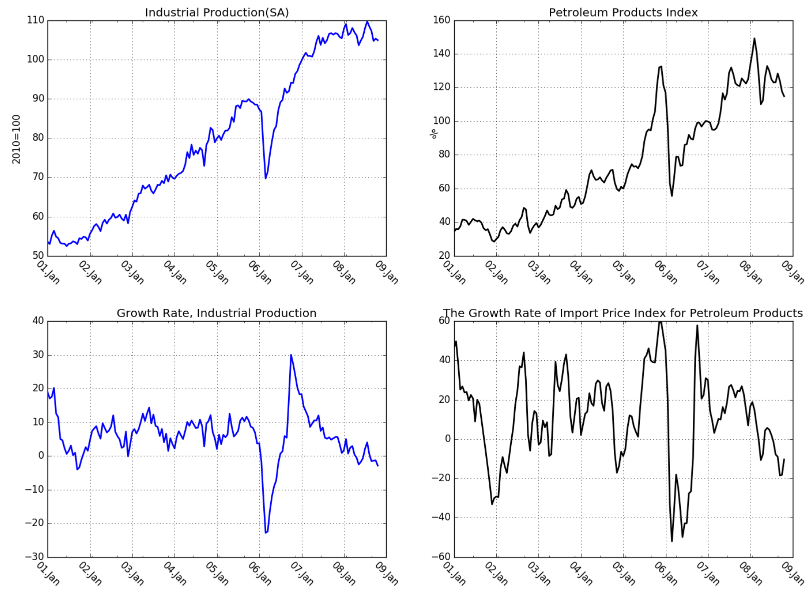
Solution:



(1) Figure 1



(2) Figure 5



(3) Figure 6 and 7

2. Suppose that the data are generated by the following model

$$
\begin{aligned}
y &= X\beta + u && (1)\\
&= X_1\beta_1 + X_2\beta_2 + u && (2)
\end{aligned}
$$

where $X = [X_1, X_2]$. Assume that X is of full column rank, $T^{-1}X'X \xrightarrow{p} Q$, and $\beta_2 \neq 0$. Denote the OLS estimate for (1) by $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Suppose you estimate the regression

$$
y = X_1\beta_1 + e \tag{3}
$$

by OLS and denote the resulting estimate by $\widehat{b}_1$.

a) Show that $\widehat{b}_1$ is inconsistent for β_1, with assuming $\mathrm{E}\left(X_1'X_2\right) \neq 0$.

b) The inconsistency of $\widehat{b}_1$ is an example of the omitted variable bias. A natural estimate would then be based on an instrumental variable procedure. Show that the OLS estimate $\widehat{\beta}_1$ can indeed be given an IV interpretation.

Solution:

a) The omitted variable bias lingers even when the sample size grows large causing the estimators to be inconsistent. We can show this by simply plugging in the true model of y to the estimator $\hat{b}_1$.

$$\hat{b}_1 = \left(X_1'X_1\right)^{-1} X_1'y \tag{4}$$
$$= \left(X_1'X_1\right)^{-1} X_1'\left(X_1\beta_1 + X_2\beta_2 + u\right) \tag{5}$$
$$= \beta_1 + \left(X_1'X_1\right)^{-1} X_1'X_2\beta_2 + \left(X_1'X_1\right)^{-1} X_1'u \tag{6}$$

The problem states that $\mathrm{E}\left(X_1'X_2\right) \neq 0, \beta_2 \neq 0$, preserving the second term. Note that the only *random variable* in the above equation is u, converging in probability to zero as $T \to \infty$ where T is the dimension of y $(T \times 1)$. Thus, there is no reason for the second term $\left(X_1'X_1\right)^{-1} X_1'X_2\beta_2$ to disappear.

$$\hat{b}_1 \xrightarrow{p} \beta_1 + \beta_2\left(X_1'X_1\right)^{-1} X_1'X_2 \tag{7}$$

b) X_2 is by assumption independent of the error term u. Therefore, using the method of moment, if we have $\mathrm{E}\left(x_{i2}u_i\right) = 0$,

$$\frac{1}{T}\sum_{t=1}^{T} x_{i2}u_i = \frac{1}{T}X_2'u = \frac{1}{T}X_2'\left(y - X_1\beta_1\right). \tag{8}$$

Solving with respect to β_1 yields

$$\hat{\beta}_1 = \left(X_2'X_1\right)^{-1} X_2'y \tag{9}$$

which is the IV estimate.

3. Consider the linear regression model:

$$y = X\beta + u \tag{10}$$
$$y, u : T \times 1 \tag{11}$$
$$X : T \times k \tag{12}$$
$$\beta : k \times 1 \tag{13}$$

with the q moment conditions $\mathrm{E}\left(z_t u_t\right) = 0$. Let

$$J_T\left(\beta, W_T\right) = g_T\left(\beta\right)' W_T g_T\left(\beta\right) \tag{14}$$

where $g_T\left(\beta\right) = T^{-1}\sum z_t\left(y_t - x_t'\beta\right)$ and W_T is some weighting matrix. The GMM estimator for β is obtained as the minimizer of $J_T(\beta, W_T)$.

a) How does your choice of W_T affect the GMM estimator? Discuss the implications on the consistency, the asymptotic normality and efficiency.

b) If the model is exactly identified ($k = q$), explain why the choice of W_T becomes irrelevant.

Solution:

a)
- *(Consistency)* Although the problem asked how the choice of W_T affects the *generalized method of moments* estimator, there are a few conditions that should hold in order for the estimator to be consistent.
  - $g_T(\xi)$ should uniformaly converge in probability to the true moment $\mathrm{E}(z_t u_t)$ in L_2 space. That is, if we denote $g_0(\xi) = \mathrm{E}\left(z_t\left(y_t - x_t'\xi\right)\right)$,

$$\sup_{\xi\in\Xi}\|g_T(\xi) - g_0(\xi)\|_2 \xrightarrow{p} 0 \tag{15}$$

    where β is replaced with the notation ξ for the time being and Ξ is the whole parameter space.
  - For all $\xi \in \Xi$ that are $\|\xi - \xi_0\|_2 > \epsilon$, $J_0(\xi, W) - J_0(\xi_0, W) > 0$ where $J_0(\xi, W) = g_0(\xi)' W g_0(\xi)$.
  - $J_T(\xi)$ should be continuous.

  With these assumptions, GMM will be consistent. Note that there is no requirement for W_T except that it should be positive definite. In fact, we could relax this contraint so that W_T is positive semi-definite but for the GMM to be consistent, $W_T \xrightarrow{p} W$ should still hold where W is a positive definite constant matrix.
- *(Asymptotic Normality)* Because we obtain GMM by minimizing $J_T(\xi, W_T)$, we need to specify the conditions that guarantee that there really is a minimum that we can reach. These are the following. Let us first define $\varphi(z_t, \xi) \equiv z_t\left(y_t - x_t'\xi\right)$ where $z_t, \varphi(z_t, \xi) \in \mathbf{R}^q$.
  - The parameter space Ξ is compact.
  - $J_T(\xi, W_T)$ is twice continuously differentiable.
  - If we define the matrix $G = \mathrm{E}\left(\nabla_\xi g_T(\xi)\right)$, then $G'WG$ should be nonsingular, i.e., invertible, since the term is included in the expression of GMM.
  - $\varphi(z_t, \xi) \in L_2$. That is, $\varphi(z_t, \xi)$ should have finite variance.
  - $\mathrm{E}\left(\sup_{\xi\in\Xi}\left\|\nabla_\xi\varphi(z_t, \xi)\right\|\right) < \infty$ so as to apply the uniform law of large numbers.

Again, there is no special requirement for the weight matrix W_T. However, we can talk about the efficiencies of the GMM estimator across different weight matrices. When we compare two matrices, we say matrix A is bigger than B if $A - B$ is positive semi-definite and we denote this relation by $A \succeq B$. If $A - B$ is strictly positive definite, we denote this by $A \succ B$.

With this relation, we can compare the variances of two GMM estimators that only differ in their weight matrices. Let us define $\Omega\left(\xi_0\right)$ to be the covariance matrix $\mathrm{E}\left[\varphi\left(z_t, \xi_0\right)\varphi\left(z_t, \xi_0\right)'\right]$ with regard to the true parameter ξ_0. Then in terms of the asymptotic variance of the GMM estimator, setting $W = \Omega^{-1}$ yields the most efficient result, meaning the variance is the smallest. We can prove this by the following procedure.

$$
\begin{aligned}
&\left(G'WG\right)^{-1} G'W\Omega WG\left(G'WG\right)^{-1} - \left(G'\Omega^{-1}G\right)^{-1} && (16)\\
&= \left(G'WG\right)^{-1}\left(G'W\Omega WG - G'WG\left(G'\Omega^{-1}G\right)^{-1}G'WG\right)\left(G'WG\right)^{-1} && (17)\\
&= \left(G'WG\right)^{-1} G'W\Omega^{1/2}\left(I - \Omega^{-1/2}G\left(G'\Omega^{-1}G\right)^{-1}G'\Omega^{-1/2}\right)\Omega^{1/2}WG\left(G'WG\right)^{-1} \succ 0 && (18)
\end{aligned}
$$

Therefore, Ω^{-1} will always be smaller than any weight matrix W in the asymptotic sense.

b) If the model is exactly identified, i.e., $k = q$, then the first-order condition of the objective function

$$X'ZW_TZ'\left(y - X\xi\right) = 0 \tag{19}$$

whose $X'Z$ term is a square matrix. Thus, if we premultiply both sides by $W_T^{-1}\left(X'Z\right)^{-1}$, the solution gets reduced to the IV estimator. It indicates that the estimator can be chosen independently of the weight matrix W_T if the model is fully identified.

4. Let $y_t \overset{\text{iid}}{\sim} \mathrm{Exp}\left(\theta\right)$ for $t = 1, \ldots, T$.

   a) Derive the score function, Hessian function and information matrix, using the exponential density.

   b) Derive the MLE for θ. Sketch the proof that the MLE is asymptotically normal. Be specific with the asymptotic variance.

Solution:

a) The likelihood function of $y_1, \ldots, y_T$,

$$L(\theta \mid y_1 \ldots, y_T) = \theta^T \exp\left(-\theta \sum_{t=1}^{T} y_t\right). \tag{20}$$

Taking the logarithm yields

$$\ell(\theta \mid y_1, \ldots, y_T) = T \log\theta - \theta \sum_{t=1}^{T} y_t. \tag{21}$$

By definition of the score function is the first derivative of the log-likelihood.

$$\frac{d\ell}{d\theta} = \frac{T}{\theta} - \sum_{t=1}^{T} y_t. \tag{22}$$

The Hessian gets reduced to the second derivative for a univariate function.

$$\frac{d^2\ell}{d\theta^2} = -\frac{T}{\theta^2} - \sum_{t=1}^{T} y_t. \tag{23}$$

To get the Fisher information,

$$\begin{aligned}
\mathcal{I}_T(\theta) &= \frac{1}{T}\mathrm{E}\left(\frac{T}{\theta^2} + \sum_{t=1}^{T} y_t\right) & (24)\\
&= \frac{1}{\theta^2} + \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}(y_t) & (25)\\
&= \frac{1}{\theta^2} + \frac{1}{\theta} & (26)
\end{aligned}$$

b) We have obtained the first and second derivatives of the log-likelihood already. Recall:

$$\begin{aligned}
\frac{d\ell}{d\theta} &= \frac{T}{\theta} - \sum_{t=1}^{T} y_t & (27)\\
\frac{d^2\ell}{d\theta^2} &= -\frac{T}{\theta^2} - \sum_{t=1}^{T} y_t & (28)
\end{aligned}$$

Using the first-order condition, the MLE comes with a closed-form expression.

$$\widehat{\theta}^{\mathrm{MLE}} = T \bigg/ \left(\sum_{t=1}^{T} y_t\right) \tag{29}$$

The second-order condition validates that the estimator is actually a maximum since

$$\frac{d^2\ell}{d\theta^2} < 0. \tag{30}$$

Now, even as rough a proof as what we will shortly give here takes at least 2 steps: the consistency of MLE and the asymptotic normality of MLE.

- *(Consistency)* Recall that we take the product of every single PDF of y_t through $t = 1, \ldots, T$ to compute the likelihood function, which also means taking the logarithm will convert the product into summation.

$$\ell\left(\theta \mid y_1, \ldots, y_T\right) = \sum_{t=1}^{T} \ell\left(\theta \mid y_t\right) \tag{31}$$

By the strong law of large numbers, we get the following relation:

$$\frac{1}{T}\sum_{t=1}^{T} \ell\left(\theta \mid y_t\right) \xrightarrow{a.s.} \mathrm{E}_{\theta_\circ}\ell\left(\theta \mid y_1\right) \tag{32}$$

for some unknown true parameter value $\theta_\circ$. We can then show that the expected log-likelihood function w.r.t. the true parameter is always greater than that of an arbitrary parameter θ by *Kullback-Leibler divergence*. The KL divergence is defined as follows.

$$\mathrm{KL}\left(f(y_1 \mid \theta_\circ) \,\|\, f(y_1 \mid \theta)\right) = \mathrm{E}_{\theta_\circ}\left[\log \frac{f(y_1 \mid \theta_\circ)}{f(y_1 \mid \theta)}\right] \tag{33}$$

$$= -\int \log \frac{f(y_1 \mid \theta)}{f(y_1 \mid \theta_\circ)}\, f(y_1 \mid \theta_\circ)\, dy_1 \tag{34}$$

By *Jensen's inequality*,

$$\underbrace{-\log \int \frac{f(y_1 \mid \theta)}{\cancel{f(y_1 \mid \theta_\circ)}}\, \cancel{f(y_1 \mid \theta_\circ)}\, dy_1}_{=0} \leq \underbrace{-\int \log \frac{f(y_1 \mid \theta)}{f(y_1 \mid \theta_\circ)}\, f(y_1 \mid \theta_\circ)\, dy_1}_{=\mathrm{KL}(f(y_1 \mid \theta_\circ) \,\|\, f(y_1 \mid \theta))}. \tag{35}$$

Therefore, it always follows that the KL divergence is nonnegative. In fact, it is strictly positive if $f(y_1 \mid \theta) \neq f(y_1 \mid \theta_\circ)$. This indicates that

$$\theta_\circ = \sup_{\theta \in \Omega} \mathrm{E}_{\theta_\circ}\ell\left(\theta \mid y_1\right). \tag{36}$$

Recall the following:

$$\widehat{\theta}^{\mathrm{MLE}} = \sup_{\theta \in \Omega} \frac{1}{T}\sum_{t=1}^{T} \ell\left(\theta \mid y_i\right). \tag{37}$$

Therefore by 32, $\widehat{\theta}^{\text{MLE}} \xrightarrow{p} \theta_\circ$ for a finite parameter space Ω. We can also prove this for a compact parameter space by starting from the lemma that

$$\frac{1}{T}\sum_{t=1}^{T}\ell\left(\theta \mid y_t\right) \xrightarrow{\text{uniformly convergent}} \int \ell\left(\theta \mid y_1\right) f\left(y_1 \mid \theta_0\right) dy_1 \quad \left(= \mathrm{E}_{\theta_0}\ell\left(\theta \mid y_1\right)\right) \tag{38}$$

which is equivalent to

$$\Pr\left(\sup_{\theta\in\Omega}\left|\frac{1}{T}\sum_{t=1}^{T}\ell\left(\theta \mid y_t\right) - \mathrm{E}_{\theta_0}\ell\left(\theta \mid y_1\right)\right| > \epsilon\right) \xrightarrow{a.s.} 0, \quad \forall \epsilon > 0. \tag{39}$$

Pointwise convergence is not enough with infinite parameter spaces because the convergence at one parameter of the log-likelihood function as a function of $y_{1:T}$ does not guarantee that the log-likelihood function with another set of $y_{1:T}$ generated with a different parameter value is also close to the true expected log-likelihood. Thus, the convergence at one parameter value does not generalize and we end up having to prove for infinitely many parameters which is not possible. Uniform convergence addresses such a problem all at once. Anyway, the proof for the consistency of MLE ends here.

- *(Asymptotic Normality)* We approximate the score function with its first-order Taylor expansion around the true parameter $\theta_\circ$ and apply the mean value theorem.

$$\frac{d\ell}{d\theta} \approx \left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ} + \left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\left(\theta - \theta_\circ\right) \tag{40}$$

where $\overline{\theta}$ lies somewhere between θ and $\theta_\circ$. Since MLE is the value which sets the first derivative to zero, we can think of the following identity.

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ} + \left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\left(\widehat{\theta}^{\text{MLE}} - \theta_\circ\right) = 0 \tag{41}$$

This, in turn, translates to the following relationship.

$$\widehat{\theta}^{\text{MLE}} - \theta_\circ = -\left(\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ}\right)\Bigg/\left(\left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) \tag{42}$$

with $\overline{\theta} = s\widehat{\theta}^{\text{MLE}} + (1-s)\,\theta_\circ,\ s \in [0,1]$. Let's slowly examine the RHS of 42. First, the numerator can be expressed as a summation of the log-likelihood of a single observation.

$$\left.\frac{d\ell\left(\theta \mid y_{1:T}\right)}{d\theta}\right|_{\theta=\theta_\circ} = \sum_{t=1}^{T}\left.\frac{d\ell\left(\theta \mid y_t\right)}{d\theta}\right|_{\theta=\theta_\circ} \tag{43}$$

By the *Central Limit Theorem*,

$$\mathrm{E}\left(\left.\frac{d\ell\left(\theta \mid y_1\right)}{d\theta}\right|_{\theta=\theta_\circ}\right) = 0 \tag{44}$$

$$\mathrm{Var}\left(\left.\frac{d\ell\left(\theta \mid y_1\right)}{d\theta}\right|_{\theta=\theta_\circ}\right) = T\mathcal{I}_1\left(\theta\right) \tag{45}$$

$$\left.\frac{d\ell\left(\theta \mid y_{1:T}\right)}{d\theta}\right|_{\theta=\theta_\circ} \xrightarrow{d} \mathcal{N}\left(0, T\mathcal{I}_1\left(\theta_\circ\right)\right) \tag{46}$$

where $\mathcal{I}_1$ is the Fisher information for a single observation y_1. The calculations for the expectation and the variance are given in the end. Now the denominator behaves like the following which makes use of the weak law of large numbers. (As is the case with consistency, if the parameter space is infinite, we need the uniform law of large numbers since we cannot say that the convergence around one parameter value does not guarantee the convergence around another. Anyway, we skip this part.)

$$\left.\frac{d^2\ell\left(\theta \mid y_{1:T}\right)}{d\theta^2}\right|_{\theta=\overline{\theta}} = \sum_{t=1}^{T}\left.\frac{d^2\ell\left(\theta \mid y_t\right)}{d\theta^2}\right|_{\theta=\overline{\theta}} \xrightarrow{p} T\mathrm{E}\left(\left.\frac{d^2\ell\left(\theta \mid y_1\right)}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) = -T\mathcal{I}_1\left(\overline{\theta}\right) \tag{47}$$

With 44, 47, and the *Slutsky's theorem*, we can conclude that 42 converges in distribution to a normal distribution.

$$-\left(\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ}\right)\Big/\left(\left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{T\mathcal{I}_1\left(\theta_\circ\right)}{\left(T\mathcal{I}_1\left(\overline{\theta}\right)\right)^2} = \frac{\mathcal{I}_T\left(\theta_\circ\right)}{\left(\mathcal{I}_T\left(\overline{\theta}\right)\right)^2}\right) \tag{48}$$

Finally, we know that $\overline{\theta} \in \left[\widehat{\theta}^{\mathrm{MLE}}, \theta_\circ\right]$ if $\widehat{\theta}^{\mathrm{MLE}} < \theta_\circ$ or $\overline{\theta} \in \left[\theta_\circ, \widehat{\theta}^{\mathrm{MLE}}\right]$ if $\widehat{\theta}^{\mathrm{MLE}} \geq \theta_\circ$. As $T \to \infty$, $\widehat{\theta}^{\mathrm{MLE}} \xrightarrow{p} \theta_\circ$ which also means $\overline{\theta} \xrightarrow{p} \theta_\circ$. Thus,

$$\widehat{\theta}^{\mathrm{MLE}} \xrightarrow{d} \mathcal{N}\left(\theta_\circ, \left(\mathcal{I}_T\left(\theta_\circ\right)\right)^{-1}\right) \tag{49}$$

There are 3 different ways to compute the Fisher information and all three are equivalent under regularity conditions. The three are

$$\mathcal{I}(\theta) = \mathrm{E}\left(\left(\frac{d\ell}{d\theta}\right)^2\right) \tag{50}$$

$$= -\mathrm{E}\left(\frac{d^2\ell}{d\theta^2}\right) \tag{51}$$

$$= \mathrm{Var}\left(\frac{d\ell}{d\theta}\right) \tag{52}$$

A Code for Figure 1

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.text as txt
import types

def rate(y,lag):
  T = len(y)
  g = (np.log(y[lag:T])-np.log(y[0:(T-lag)]))*100.
  return g

data = np.loadtxt(                                        ,skiprows=1)
date = data[:,0]
base = data[:,1]
cpi  = data[:,2]
inf  = rate(cpi,12)
imp_price = data[:,3]
dimp_price = rate(imp_price,12)
ip = data[:,5]
dip = rate(ip,12)

labels = [        ,        ,        ,        ,        ,        ,        ,        ,
           ,        ,        ,        ,        ]
fig,ax = plt.subplots()
fig.suptitle(                        )
line, = ax.plot(base[12:],label=           ,linestyle=    )
plt.ylim([1.5,np.max(base[12:])+.5])
#ax = plt.axes()
plt.xticks(list(range(9,165,12)),labels)
# start,end = ax.get_xlim()
# ax.xaxis.set_ticks(np.arange(start,end,(end-start)/13.))
```

```
ax.set_xticklabels(labels,rotation=-45)
ax.annotate(                                                ,xy=(100.441,5.21404),
    xytext=(105,5.5),
            arrowprops=dict(facecolor=       ,arrowstyle=    ))
ax.annotate(                    ,xy=(100.688,4.99432),xytext=(105,4.8),
            arrowprops=dict(facecolor=       ,arrowstyle=    ))
legend = plt.legend(handles=[line],loc=4)

plt.ylabel(   )
for lab in ax.xaxis.get_majorticklabels():
  lab.customShiftValue = 5.
  lab.set_x = types.MethodType(lambda self,x:  txt.Text.set_x(self, x+self.
      customShiftValue),
                                    lab)
plt.show()
```

B Code for Figure 5

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.text as txt
import types

def rate(y,lag):
  T = len(y)
  g = (np.log(y[lag:T])-np.log(y[0:(T-lag)]))*100.
  return g

data = np.loadtxt('/Users/daeyounglim/Downloads/data.txt',skiprows=1)
date = data[:,0]
base = data[:,1]
cpi  = data[:,2]
inf  = rate(cpi,12)
imp_price = data[:,3]
dimp_price = rate(imp_price,12)
ip = data[:,5]
dip = rate(ip,12)


# labels = ['01.Jan','02.Jan','03.Jan','04.Jan','05.Jan','06.Jan','07.Jan','08.Jan
    ','09.Jan','10.Jan','11.Jan','12.Jan','13.Jan']
labels = ['04.Jan','05.Jan','06.Jan','07.Jan','08.Jan','09.Jan','10.Jan','11.Jan
    ','12.Jan','13.Jan']
fig,(ax1,ax2) = plt.subplots(2,1,sharex=False)
```

```
# line, = ax1.plot(cpi[12:],linewidth=2.)
line, = ax1.plot(cpi[36:],linewidth=2.)
line2, = ax2.plot(inf[24:],linewidth=2.)
ax1.xaxis.set_ticks(list(range(9,129,12)))
ax1.set_xticklabels(labels,rotation=-45)
ax2.xaxis.set_ticks(list(range(9,129,12)))
ax2.set_xticklabels(labels,rotation=-45)
ax1.set_ylabel('2010=100',fontsize=10)
ax2.set_ylabel('%',fontsize=10,rotation=0)
ax1.set_title('CPI',fontsize=15)
ax2.set_title('Inflation Rate',fontsize=15)
plt.show()
```

C Code for Figure 6 and 7

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.text as txt
import types

def rate(y,lag):
  T = len(y)
  g = (np.log(y[lag:T])-np.log(y[0:(T-lag)]))*100.
  return g

data = np.loadtxt('/Users/daeyounglim/Downloads/data.txt',skiprows=1)
date = data[:,0]
base = data[:,1]
cpi  = data[:,2]
inf  = rate(cpi,12)
imp_price = data[:,3]
dimp_price = rate(imp_price,12)
ip = data[:,5]
dip = rate(ip,12)

labels = ['01.Jan','02.Jan','03.Jan','04.Jan','05.Jan','06.Jan','07.Jan','08.Jan
    ','09.Jan','10.Jan','11.Jan','12.Jan','13.Jan']
plt.figure(figsize=(2,2))
ax1 = plt.subplot(221)
ax2 = plt.subplot(222)
ax3 = plt.subplot(223)
ax4 = plt.subplot(224)
# fig,(ax1,ax2,ax3,ax4) = plt.subplots(2,2,sharex=False)
line1, = ax1.plot(ip[12:],'b',linewidth=2.)
```

```
line2, = ax3.plot(dip,'b',linewidth=2.)
line3, = ax2.plot(imp_price[12:],'k',linewidth=2.)
line4, = ax4.plot(dimp_price,'k',linewidth=2.)
ax1.set_title('Industrial Production(SA)')
ax2.set_title('Petroleum Products Index')
ax3.set_title('Growth Rate, Industrial Production')
ax4.set_title('The Growth Rate of Import Price Index for Petroleum Products')
ax1.set_ylabel('2010=100')
ax2.set_ylabel('%',rotation=-45)
ax1.set_xticks(list(range(9,165,12)),labels)
ax1.set_xticklabels(labels,rotation=-45)
ax1.grid(True)
ax2.set_xticks(list(range(9,165,12)),labels)
ax2.set_xticklabels(labels,rotation=-45)
ax2.grid(True)
ax3.set_xticks(list(range(9,165,12)),labels)
ax3.set_xticklabels(labels,rotation=-45)
ax3.grid(True)
ax4.set_xticks(list(range(9,165,12)),labels)
ax4.set_xticklabels(labels,rotation=-45)
ax4.grid(True)

plt.show()
```