Projet Économétrie

Rapport

Sami Bargaoui
Mohammed Benseddik
Fares Zenaidi

May 18, 2017

1 Partie 1 : Régression

1.1 Question 1 : Lire le fichier mroz.txt. Ne sélectionner que les observations pour lesquelles la variable *wage* est strictement positive.

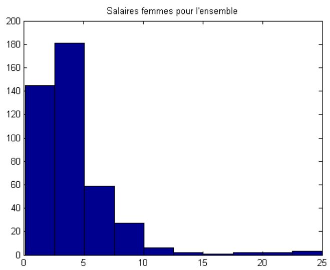
```
% Question 1 :

MROZ = load('MROZ.raw');
positive_wage = MROZ(MROZ(:,7) > 0,:);
```

1.2 Question 2 : Faire les statistiques descriptives du salaire, de l'age et de l'éducation pour l'ensemble des femmes puis, pour les femmes dont le salaire du mari est supérieure à la médiane de l'échantillon, puis pour les femmes dont le salaire du mari est inférieur à la médiane de l'échantillon

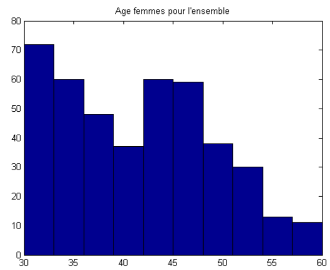
```
% Question 2
function [res] = desc(var,name)
h=figure();
hist(var);
title(name);
hold off
disp('mean and median');
y = [mean(var),median(var)];
disp(y);
disp('min and max');
y = [min(var),max(var)];
disp(y);
end
```

```
female_obs = positive_wage;
```

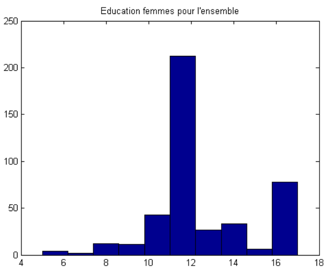
```
female_salaries = female_obs(:,7);
female_age = female_obs(:,5);
female_education = female_obs(:,6);
% On calcule le salaire du mari en multipliant le salaire par heure par le
% nombre d'heures travaillees
positive_wage = [positive_wage,female_obs(:,9).* female_obs(:,12)]
% Ceci rajoute la colonne wage a la 23eme position
female_obs = positive_wage;
female_salaries = female_obs(:,7);
female_age = female_obs(:,5);
female_education = female_obs(:,6);
desc(female_education, 'Education femmes pour l''ensemble')
desc(female_age, 'Age femmes pour l''ensemble')
desc(female_salaries, 'Salaires femmes pour l''ensemble')
female_obs_lower = positive_wage(find(positive_wage(:,23) <= median(positive_wage(:,23))),:);
female_salaries = female_obs_lower(:,7);
female_age = female_obs_lower(:,5);
female_education = female_obs_lower(:,6);
desc(female_education, 'Education femmes dont le salaire est inf a la mediane')
desc(female_age, 'Age femmes dont le salaire est inf a la mediane')
desc(female_salaries, 'Salaires femmes dont le salaire est inf a la mediane')
female_obs_higher = positive_wage(find(positive_wage(:,23) >= median(positive_wage(:,23))),:);
female_salaries = female_obs_higher(:,7);
female_age = female_obs_higher(:,5);
female_education = female_obs_higher(:,6);
desc(female_education, 'Education femmes dont le salaire est superieur a la mediane')
desc(female_age, 'Age femmes dont le salaire est superieur a la mediane')
desc(female_salaries, 'Salaires femmes dont le salaire est superieur a la mediane')
```



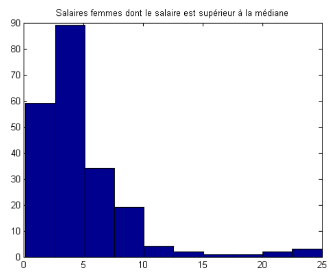
(a) fig 1



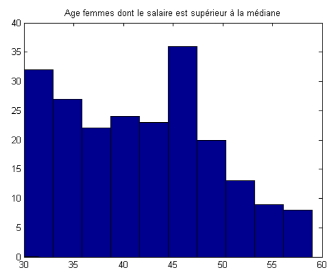
(b) fig 2



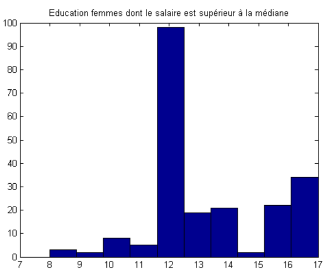
(c) fig 3



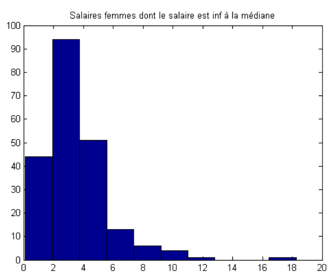
(d) fig 4



(e) fig 5

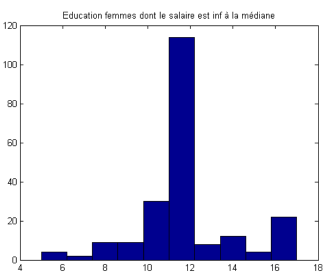


(f) fig 6



(g) fig 7

(h) fig 8



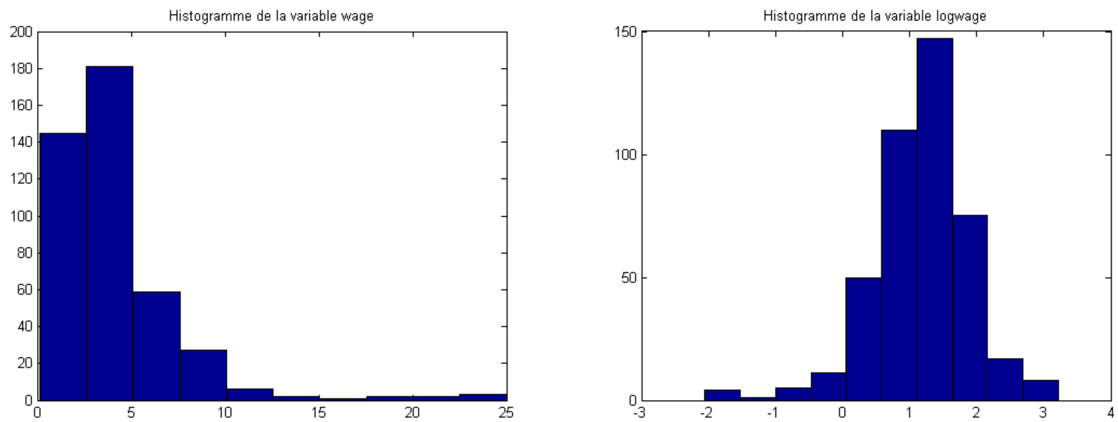
(i) fig 9

1.3 Question 3 : Faire l'histogramme de la variable *wage*. Calculer le log de *wage* et faire l'histogramme. Comparez les deux histogrammes et commentez.

La variable logwage présente le profil d'une gaussienne. Ceci sera utile après pour réaliser les différents tests d'hypothèses (Voir graphes a et b ci dessous 1)

Table 1 – Statistiques descriptives

	Mean	*Median*	*Min*	*Max*
Education femmes	12.6589	12	5	17
Age femmes	41.9720	42	30	60
Salaires femmes	4.177	3.4819	0.1282	25
Education f/salaire est sup à la méd	4.8522	3.9098	0.1616	25
Age f/salaire est sup à la méd	41.7383	12	30	59
Salaires f/salaire est sup à la méd	13.3364	12	8	17
Education f/salaire est inf à la méd	11.9813	12	5	17
Age f/salaire est inf à la méd	42.2056	43	30	60
Salaires f/salaire est inf à la méd	3.5031	2.9827	0.1282	19.2670



(a) Histogramme de wage **(b)** Histogramme de logwage

Figure 1 – Histogrammes

1.4 Question 4 : Calculer les corrélations motheduc et fatheduc. Commentez. Il y a-t-il un problème de multicollinéarité si l'on utilise ces variables comme variables explicatives ?

```
% Question 4 :
correlation = corr(female_obs(:,15), female_obs(:,16))
```

Corrélation = 0.5541. Les deux variables ne sont pas très corrélées.

On peut a priori utiliser ces deux variables car ça ne va pas réduire le rang de la matrice des variables explicatives. Cependant, il faut vérifier si il y a d'autres variables qui peuvent introduire une colinéarité par combinaison linéaire.

1.5 Question 5 : Faites un graphique en nuage de point entre wage et educ, wage et exper, wage et fatheduc. Commentez. S'agit-il d'un effet "toute chose étant égale par ailleurs ?"

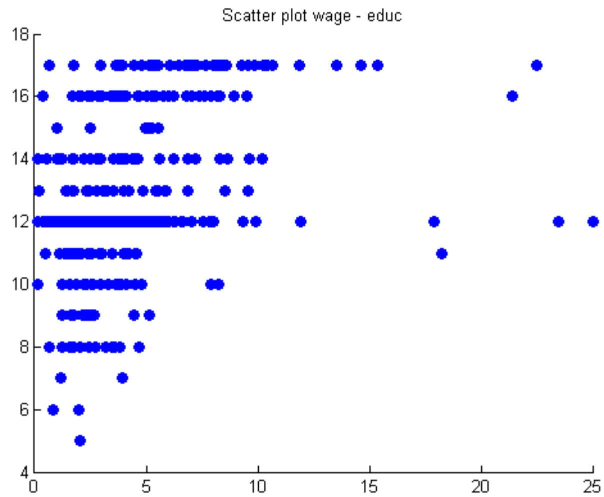


Figure 2 – Scatter wage - educ

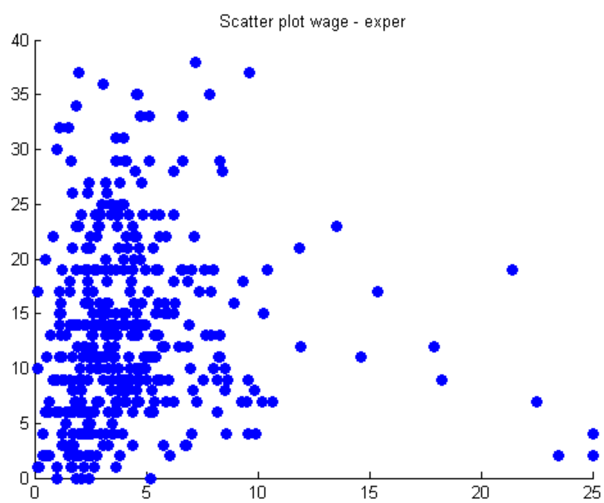


Figure 3 – Scatter wage - exper

```
h=figure();
scatter(positive_wage(:,7), positive_wage(:,6), 'filled');
title('Scatter plot wage - educ');
```

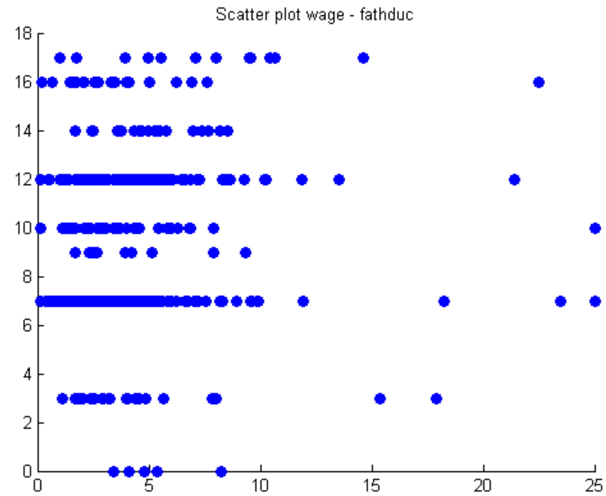


Figure 4 – Scatter wage - fatheduc

```
h=figure();
scatter(positive_wage(:,7), positive_wage(:,19), 'filled');
title('Scatter plot wage - exper');
h=figure();
scatter(positive_wage(:,7), positive_wage(:,16), 'filled');
title('Scatter plot wage - fathduc');
```

Il ne s'agit pas d'un effet ' toute chose étant égale par ailleurs, car on a plusieurs valeurs de y qui varient pour le même x.

1.6 Question 6 : Quelle est l'hypothèse fondamentale qui garanti des estimateurs non biaisés ? Expliquer le biais de variable omise.

L'hypothèse fondamentale qui garantit des estimateurs non biaisés est le fait de ne pas avoir une corrélation entre les variables observées et les variables non-observées.

Si une variable est omise elle introduira un biais dans l'estimateur sauf si cette dernière est orthogonale avec le reste des variables explicatives qu'on a considérées.

1.7 Question 7 : Faire la régression de wage en utilisant les variables explicatives un constante, city, educ, exper, nwifeinc, kidslt6, kidsgt6. Commentez l'histogramme des résidus.

```
y = positive_wage(:,7);
[n,k] = size(y);
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
[n,k] = size(X);
beta = inv(X'*X)*X'*y;
```

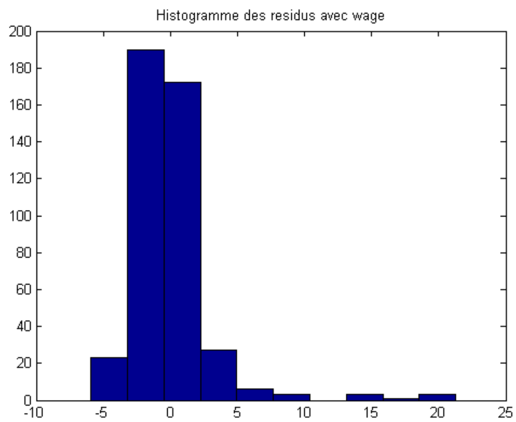


Figure 5 – Histogramme des résidus de la regression (wage)

```
u=y-X*beta;
figure();
hist(u);
title('Histogramme des residus avec wage');
```

L'espérance du bruit n'est pas nulle, on observe une certaine déviation.

1.8 Question 8 : Faire la régrssion de lwage sur une constante, city, educ, exper, nwifeinc, kidslt6, kidsgt6. Comparer l'histogramme obtenu à celui de la question 7.

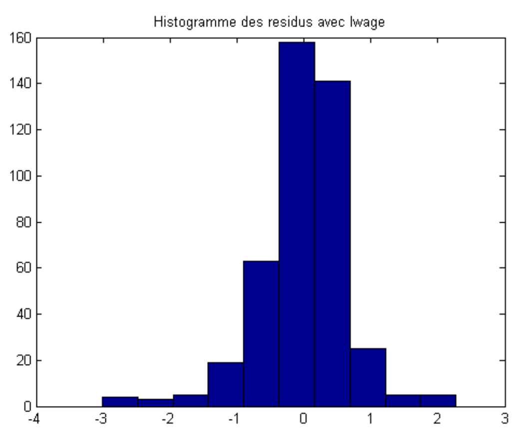


Figure 6 – Histogramme des résidus de la regression (logwage)

```
y = positive_wage(:,21);
```

```
[n,k] = size(y);
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
[n,k] = size(X);
beta = inv(X'*X)*X'*y;
u=y-X*beta;
figure();
hist(u);
title('Histogramme des residus avec lwage');
```

L'histogramme est centré en 0. Le bruit présente le profil d'une gaussienne centrée réduite. Son espérance est égale donc à 0.

1.9 Question 9 : Tester l'hypothèse de non significativité de exper avec un seuil de significativité de 1%, 5% et 10% (test alternatif des deux côtés). Commentez les p-values.

```
sig2 = u'*u/(n-k);
std = sqrt(diag(sig2*inv(X'*X)));
t = beta ./ std;
t_exp = t(4)
% 1%
if t_exp <= tdis_inv(1-0.005,n-k)
disp('on accepte a 1%')
else disp('on rejette a 1%')
end
% 5%
if t_exp <= tdis_inv(1-0.025,n-k)
disp('on accepte a 5%')
else disp('on rejette a 5%')
end
% 10%
if t_exp <= tdis_inv(1-0.05,n-k)
disp('on accepte a 10%')
else disp('on rejette a 10%')
end
```

On rejette l'hypothèse à 1%, 5% et 10%.

1.10 Question 10 : Tester l'hypothèse que le coefficient associé à educ est égal à 10% avec un seuil de significativité de 5% (test à alternatif des deux côtés.)

```
% Question 10 :
```

```
t_exp = tdis_inv(0.025,n-k)
t_educ = (beta(3)-0.1) ./ std(3)
p = tdis_prb(t_educ ,n - k )
```

On obtient que p = 0.5284. On ne peut pas rejeter l'hypothèse que le coefficient associé à educ est égal à 10% avec un seuil de significativité de 5%.

1.11 Question 11 : Tester l'hypothèse jointe que le rendement de l'éducation est de 10% et que celui de l'expérience professionnelle est de 5%.

```
% Question 11
% On estime le modèle non contraint et on calcule SSR0
y = positive_wage(:,21);
[n,k] = size(y);
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
[n,k]=size(X)
beta0=inv(X'*X)*X'*y
u0=y-X*beta0
SSR0=u0'*u0
%On estime maintenant le modèle contraint, c'est-à-dire avec b_educ=0.1 et b_exper=0.05. On calcule SSR1.
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
beta1=inv(X'*X)*X'*y
beta1(2) = beta1(2) - 0.1
beta1(3) = beta1(3) - 0.05
u1=y-X*beta1
SSR1=u1'*u1
% On calcule la statistique F du test de Fisher
k0=6
F=((SSR1-SSR0)/SSR0)*((n-k0)/2)
p=fdis_prb(F,2,n-k0)
```

On trouve une p-value de 0.711. On ne peut pas rejeter l'hypothèse.

1.12 Question 12 : De combien augmente le salaire en pourcentage avec 10 années d'expérience ?.

```
% Question 12
augment_salaire_10 = 10 * beta(4)
```

On trouve augmentsalaire10 = 0.1549. Sur 10 ans, on observe une augmentation d'environ 15.5%.

1.13 Question 13 : Tester l'égalité des coefficients associés aux variables kidsgt6 et kidslt6. Interprétez.

```
% Question 13
tmp = positive_wage(:,3) + positive_wage(:,4);
X = [ones(n,1),positive_wage(:,[18,6,19,20,3]),tmp];
beta = inv(X'*X)*X'*y;
u=y-X*beta;
sig2 = u'*u/(n-k);
std = sqrt(diag(sig2*inv(X'*X)));
t = beta ./ std
p = tdis_prb(t(7) ,n - k)
```

p = 0.6620. On ne peut pas rejeter l'hypothèse d'égalité des deux coefficients.

1.14 Question 14 : En utilisant le modèle de la question 7, faire le test d'hétéroscédasticité de forme linéaire. Corriger le problème par rapport à la variable la plus importante en utilisant la méthode des MCG. Comparer les écarts-types des coefficients estimés avec ceux obtenus à la question 7. Commenter.

```
y = positive_wage(:,7);
[n,k] = size(y);
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
[n,k] = size(X);
beta = inv(X'*X)*X'*y;
u=y-X*beta;
sig2=u'*u/(n-k)
u2=u.^2;
y=u2;
% modele non contraint
beta=inv(X'*X)*X'*y;
u=y-X*beta;
SSR0=u'*u
% modele contraint (tous les coefficients sauf l'intercept
    sont nuls)
X=[ones(n,1)];
beta=inv(X'*X)*X'*y
u=y-X*beta;
SSR1=u'*u
 %Test de Fisher
```

```
F = ((SSR1-SSR0)/SSR0)*(n-k)/6
p = fdis_prb(F,6,n-k)
% correction MCG :
X = [ones(n,1),positive_wage(:,[18,6,19,20,3,4])];
denum = sqrt( X(:,find(max(beta))) );
y = y ./ denum;
for i =1:k
        X(:,i)= X(:,i)./sqrt(denum)
end
[n,k]=size(X)
beta=inv(X'*X)*X'*y;
u=y-X*beta;
y = u.^2;
% modele non contraint
beta = inv(X'*X)*X'*y;
u = y-X*beta;
SSR0 = u'*u
% modele contraint (tous les coefficients sauf l'intercept sont nuls)
X=[ones(n,1)];
beta=inv(X'*X)*X'*y
u=y-X*beta;
SSR1=u'*u;
% Test de Fisher
F = ((SSR1-SSR0)/SSR0)*(n-k)/6;
p = fdis_prb(F,6,n-k)
```

On a $p = 0.1477 \geq 0.025$ et $F = 1.5926$. On ne rejette pas l'hypothèse d'homoscédasticité. Après correction, $p = 0.0633$. On ne rejette pas non plus l'hypothèse d'homoscédasticité.

1.15 Question 15 : Tester le changement de structure de la question 8 entre les femmes qui ont plus de 43 ans et les autres : test sur l'ensemble des coefficients. Refaire le test avec 3 groupes (mutuellement exclusifs) : les femmes de moins de 30 ans, entre 30 et 43 ans, plus de 43 ans.

```
positive_wage = MROZ(find(MROZ(:,21) > 0),:);
y = positive_wage(:,21);
y_moins43 = positive_wage(find(positive_wage(:,5) <= 43),21);
y_43 = positive_wage(find(positive_wage(:,5) > 43),21);
y_plus43 = positive_wage(find(positive_wage(:,5) < 2),21);
female_obs_43 = positive_wage(find(positive_wage(:,5) <= 43)
    ,[18,6,19,20,3,4]);
```

```
female_obs_plus43 = positive_wage(find(positive_wage(:,5) >
    43),[18,6,19,20,3,4]);
female_obs = positive_wage(:,[18,6,19,20,3,4]);
X = female_obs;
[n,k] = size(X);
X_moins43 = female_obs_43;
[n_moins43,p_moins43] = size(X_moins43);
X_43 = female_obs_43;
[n_43,p_43] = size(X_43);
X_plus43 = female_obs_plus43;
[n_plus43,p_plus43] = size(X_plus43);
X = [ones(n,1),X];
X_moins43 = [ones(n_moins43,1), X_moins43];
X_43 = [ones(n_43,1), X_43];
X_43 = [ones(n_plus43,1), X_plus43];
beta = inv(X'*X)*X'*y
u = y-X*beta;
SSR = u'*u
beta = inv(X_moins43'*X_moins43)*X_moins43'*y_moins43
u = y_moins43 -X_moins43*beta;
SSR0=u'*u
beta = inv(X_43'*X_43)*X_43'*y_43
u = y_43 -X_43*beta;
SSR1=u'*u
[n,k] = size(X);
t = ((SSR-(SSR1+SSR0))/SSR) * ((n-2*k)/k)
p = fdis_prb(t,k,n-2*k)
```

p = 0.8577. Le sous groupe étudié n'a pas la même structure de salaire par rapport au reste des observations. On ne rejette pas cette hypothèse.

1.16 Question 16 : A partir de la variable kidslt6, créer un ensemble de variables binaires pour le nombre d'enfants de moins de 6 ans. Refaire la question 8 avec ces variables et en utilisant comme référence les femmes qui ont des enfants de plus de 6 ans. Ces catégories sontelles mutuellement exclusives ? Interprétez les paramètres associés aux variables binaires. Faire le test de non significativité de l'ensemble des variables binaires.

```
function oneHotLabels = Encoding(labels)
valueLabels = unique(labels);
nLabels = length(valueLabels);
```

```
[nSamples,dummy] = size(labels);
oneHotLabels = zeros(nSamples, nLabels);
for i = 1:nLabels
    oneHotLabels(:,i) = (labels == valueLabels(i));
end
```

```
binaire = Encoding(positive_wage(:,3));
positive_wage_bin = [positive_wage,binaire]
female_obs = positive_wage_bin(find(positive_wage_bin(:,4) > 0),[18,6,19,20,3,4,23,24,25]);
[n,k] = size(female_obs);
X = [ones(n,1),female_obs];
y = positive_wage_bin(find(positive_wage_bin(:,4) > 0),21);
beta = inv(X'*X)*X'*y
beta
u=y-X*beta;
sig2=u'*u/(n-k)
u2=u.^2;
y=u2;
%modele non contraint
beta=inv(X'*X)*X'*y;
u=y-X*beta;
SSR0=u'*u
SSR0 = u'*u
%modele contraint (tous les coefficients sauf l'intercept sont nuls)
X=[ones(n,1)];
beta=inv(X'*X)*X'*y
u=y-X*beta;
SSR1=u'*u;
% Test de Fisher
F = ((SSR1-SSR0)/SSR0)*(n-k)/2;
p = fdis_prb(F,2,n-k)
```

L'hypothèse de non significativité de l'ensemble des variables binaires est rejetée. p = 6.5980e-05.

Les catégories ne sont pas mutuellement exclusives.

beta = 0.0000 -0.0000 -0.0000 -0.0000 -0.0000 -3.5849 -0.0000 -0.5313 -0.0489 0.0000.

Les femmes avec 0 enfant de moins de 6 gagnent 53% plus que les femmes avec au moins un enfant de plus de 6 ans. Les femmes avec 1 enfant de moins de 6 ans gagnent 4% de plus que les femmes avec au moins un enfant de plus de 6 ans.

1.17 Question 17 : Créer une variable binaire pour les femmes qui ont des enfants de moins de 2 ans. Créer un terme d'interaction entre le nombre d'enfants de moins deux ans et l'éducation. Faire la régression de lwage sur une constante, city, educ, exper, nwifeinc, et ces deux dernières variables. Faire le test de significativité du terme d'interaction. Interpréter.

```
% Question 17 :
% Création des labels binaires
binaire = Encoding(positive_wage(:,3));
positive_wage_bin = [positive_wage,binaire]
female_obs = positive_wage_bin(find(positive_wage_bin(:,4) > 0),[18,6,19,20,3,4,23,24,25]);
% Création de la variable d'interaction
inter = positive_wage_bin(:,3) .* positive_wage_bin(:,6)
y = positive_wage_bin(:,21)
[n,k]=size(y)
X=[ones(n,1), positive_wage_bin(:,[18,6,19,2, 25]),inter];
[n,k]=size(X)
%y = positive_wage_bin(:,21)
beta=inv(X'*X)*X'*y
u=y-X*beta;
sig2=u'*u/(n-k)
u2=u.^2;
y=u2;
%modele non contraint
beta=inv(X'*X)*X'*y;
u=y-X*beta;
SSR0=u'*u
%modele contraint (tous les coefficients sauf l'intercept sont nuls)
X=[ones(n,1)];
beta=inv(X'*X)*X'*y
u=y-X*beta;
SSR1=u'*u
%Test de Fisher
F = ((SSR1-SSR0)/SSR0)*(n-k)/6
p = fdis_prb(F,6,n-k)
```

On trouve que p = 2.0386e-04.

2 Partie 2 : Séries temporelles 1

2.1 Préface avant les réponses - Fonctions récurrentes utilisées :

```
function [t, p_value] = test_auto_correlation(residu)
[n, ~] = size(residu);
u_ = [residu(2:n)];
u_lag = [residu(1: n -1)];
y= u_;
X =[u_lag];
[n , k] = size(X);
rho = inv(X'* X )*X'* y;
u = y-X*rho;
sig2 =u'* u/(n -k);
std = sqrt(diag(sig2*inv(X'*X)));
t=rho ./std;
% test de student: on teste rho =0.
p_value = tdis_prb(t, n-k);
end

% Fonction qui régresse X par rapport à y, calcule les résidus et la RMSE.
function [beta, rmse, u, SSR] = regression(X, y)
beta = inv(X'*X)*X'*y;
u = y - X*beta;
rmse = sqrt(u'*u);
SSR = u'*u;
end

% Fonction qui régression x(t) par rapport à x(t-1) et x(t-2) et calcule la
% RMSE et effectue des prédictions à t+1, t+2 et t+3.
function [pred, rmse] = compute_prediction_rmse(x)
[n, ~] = size(x);
y_ = x(3:n);
x_lag = x(2:n-1);
x_lag2 = x(1:n-2);
[n, ~] = size(x_lag);
X = [ones(n, 1) x_lag x_lag2];
beta = inv(X'*X)*X'*y_;
u = y_-X*beta;
```

```
pred = X*beta;
rmse = sqrt(u'*u);
end

% Fonction qui détermine la statistique t d'une série temporelle donnée.
function t = determine_t_statistics(x)
[n, ~] = size(x);
y_ = x(2:n);
y_lag = x(1:n-1);
delta_y = y_ - y_lag;
[n, ~] = size(delta_y);
X = [ones(n, 1) y_lag];
[n, k] = size(X);
beta = inv(X'*X)* X'*delta_y;
u = delta_y - X*beta;
sig2 = u'*u/(n-k);
std = sqrt(diag(sig2*inv(X'*X)));
t_vec = beta./std;
t = t_vec(2);
end
```

2.2 Question 1 : Ouvrir le fichier volat.raw; importer les données sur matlab en recodant les valeurs manquantes en -9999.

Import des données via la fonction Home/Import data avec remplacement des valeurs manquantes (.) par -999.

2.3 Question 2 : Pourquoi doit-on stationnariser les séries ?

Stationnariser les séries pour pouvoir les modéliser à l'aide de la méthode des MCO (modéle ARMA). Pour une série non-stationnaire : indépendance des observations n'est plus valable, théoréme central limite ne s'applique plus, distribution des paramètres de la régression ne suit plus une loi de Student. Une série temporelle Y_t (t=1,2...) est dite stationnaire (au sens faible) si ses propriétés statistiques ne varient pas dans le temps (espérance, variance, auto-corrélation).

2.4 Question 3 : Faire le test de racine unitaire de log(sp500), de pcsp de div.

Hypothèse nulle H0 : série a été générée par un processus présentant une racine unitaire $\mapsto$ non stationnaire.

```
load('VOLAT.mat');  % valeurs manquantes fixées à 9999.
```

```
% logsp500
logsp500 = log(VOLAT(:, 2));
t_logsp500 = determine_t_statistics(logsp500);
fprintf('Valeur de t pour logsp500: ');
disp(t_logsp500)

% pcsp
pcsp = VOLAT(:, 6);
t_pcsp = determine_t_statistics(pcsp);
fprintf('Valeur de t pour pcsp: ');
disp(t_pcsp)

% div
div = VOLAT(:, 3);
t_div = determine_t_statistics(div);
fprintf('Valeur de t pour div: ');
disp(t_div)
```

```
Valeur de t pour logsp500:    -0.4900

Valeur de t pour pcsp:  -249.2764

Valeur de t pour div:    -1.5655
```

Pour logsp500 : on ne peut pas rejeter l'hypothèse H0 à 10% (-2.57).
Pour pcsp : on rejette l'hypothèse H0 à 1% ($\leq$ -3.43).
Pour div : on ne peut pas rejeter l'hypthèse H0 à 10% (-2.57).

2.5 Question 4 : Interpréter l'autocorrélogramme et l'autocorrélogrammes partiels de pcsp et de div. Quelle est la différence entre ces deux graphiques ?

Autocorrelogramme et l'autocorrelogramme partiel permettent d'identifier le type de séries auxquelles on a affaire.

PCSP : ACF ne décroit pas de manière exponentielle, PACF ne décroit à 0.
$\mapsto$ Processus non AR.
DIV : ACF décroit de manière exponentielle, PACF décroit de 0 a 2.
$\mapsto$ Processus AR(2).

```
subplot(2,3,1);
plot(pcsp);
```

```
title('pcsp');

subplot(2,3,2);
sacf(pcsp, 20);
title('acf de pcsp');

subplot(2,3,3);
spacf(pcsp, 20);
title('pacf de pcsp');

subplot(2,3,4);
plot(div);
title('div');

subplot(2,3,5);
sacf(div, 20);
title('acf de div');

subplot(2,3,6);
spacf(div, 20);
title('pacf de div');
```

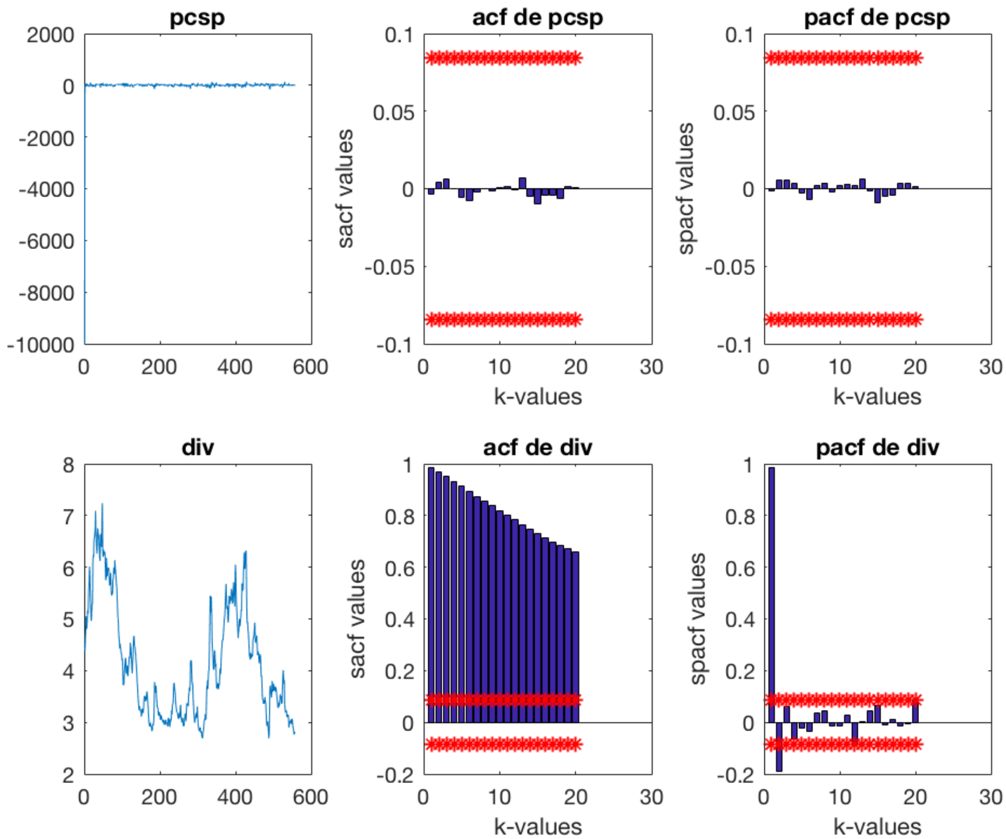


Figure 7 – Autocorrélogramme et l'autocorrélogramme partiel de pcsp et de div.

2.6 Question 5 : Proposer une modélisation AR(p) d'une série de votre choix, en utilisant tous les outils vus au cours.

Modélisation AR d'une série de notre choix (série div) :
Pour choisir l'ordre p de notre modèle auto-régressif, nous allons utiliser le critère AIC. On choisi le modèle AR qui a le AIC le plus petit.

```
% Pour un processus AR(1)
[n, ~] = size(div);
y_ = div(2: n);
y_lag = div(1: n-1);
X = y_lag;
[n , k] = size(X);
beta = inv(X'*X)*X'*y_;
u=y_-X* beta;
sig2 =u'*u/(n-k);
AIC1 = log(sig2)+(2.*(1))./n;
```

```
% Pour un processus AR(2)
y_ = div(3:n);
y_lag = div(2:n-1);
y_lag2 = div(1:n-2);
X =[y_lag y_lag2];
[n, k] = size(X);
beta = inv(X'*X)*X'*y_;
u=y_-X* beta;
sig2 =u'* u/(n-k);
AIC2 = log(sig2)+(2.*(2))./n;

fprintf('AIC1: ');
disp(AIC1);
fprintf('AIC2: ');
disp(AIC2);

% AIC2 < AIC1, donc on retient l'ordre 2 pour div.
```

Résultats :

```
AIC1:    -3.6054

AIC2:    -3.6420
```

2.7 Question 6 : Faire des prévisions pour T+1, T+2, T+3 en utilisant le modèle AR(p) de question 5. Calculer les RMSE.

Prévision à t + 1, t + 2, t + 3 en utilisant div et calcul de RMSE :

```
[pred, rmse] = compute_prediction_rmse(div);
fprintf('RMSE = ');
disp(rmse);
```

```
RMSE =     3.7808
```

3 Partie 3 : Séries temporelles 2

3.1 Question 1 : Ouvrir le fichier Faire une régression de i3 sur inf_1, inf_2, inf_3 et def_1, def_2, def_3 où x_i représente la variable x laggée de i périodes. Calculer l'impact à 1, 2 et 3 périodes de inf et def sur i3. Calculer les impacts à long terme; commentez.

Régression de i3 sur inf_1, inf_2, inf_3 et def_1, def_2, def_3 :

```
load('intdef.mat')

i3 = intdef1(:, 2);
[n, ~] = size(i3);
i3 = i3(4:n);

inf = intdef1(:, 3);
[n, ~] = size(inf);
inf_1 = inf(3:n-1);
inf_2 = inf(2:n-2);
inf_3 = inf(1:n-3);
X1 = [inf_1 inf_2 inf_3];

def = intdef1(:, 6);
[n, ~] = size(def);
def_1 = def(3:n-1);
def_2 = def(2:n-2);
def_3 = def(1:n-3);
X2 = [def_1 def_2 def_3];

% Impact à 1, 2, 3 périodes, impact à long terme

[beta1, rmse1, residus1, SSR1] = regression(X1, i3);
fprintf('i3 | inf_1, inf_2, inf_3')
disp(beta1);
disp('Impact à long terme de inf');
disp(sum(beta1))

[beta2, rmse2, residus2, SSR2] = regression(X2, i3);
disp('Beta2')
disp(beta2);
disp('Impact à long terme de def');
disp(sum(beta2))
```

```
i3 | inf_1, inf_2, inf_3    0.7369
    0.0001
    0.4106

Impact � long terme de inf
    1.1476

Beta2
    0.7627
    0.3865
    0.7494

Impact � long terme de def
    1.8986
```

L'hypothèse H0 (les résidus sont dé-corrélés) est rejette très fortement pour la première et la deuxième régressions (p_value très faible et inférieure à la statistique de test t).

3.2 Question 2 : Tester l'auto-corrélation des erreurs

```
[t1, p_value1] = test_auto_correlation(residus1);
disp('t1');
disp(t1)
fprintf('p_value1');
disp(p_value1)


[t2, p_value2] = test_auto_correlation(residus2);
fprintf('t2');
disp(t2)
fprintf('p_value2');
disp(p_value2)
```

Résultats :

```
t1
    5.5429

p_value1   1.0565e-06

t2   12.2528

p_value2     0
```

L'hypothèse H0 (les résidus sont dé-corrélés) est rejetée très fortement pour la première et la deuxième régression (p_value très faible et inférieure à la statistique de test t).

3.3 Question 3 : Tester l'hypothèse que les valeurs passées de *inf* n'ont pas d'impact sur i3.

Les valeurs passées de inf n'ont pas d'impact sur i3.
Pour cela, il faut tester que les bêtas issus de la régression de *inf* sur i3 sont tous égaux à 0.
H0 : beta_1 = beta_2 = beta_3 = 0.
On utilisera la statistique de Fisher pour rejeter ou non H0.

```
% Modèle contraint (uniquement beta_0 qui reste)
[n, ~] = size(i3);
X = ones(n,1);
[beta3, rmse3, residus3, SSR3] = regression(X, i3);

% Test de Fisher
F = ((SSR3-SSR1)/SSR1)*(n-3)/3;
disp('Statistique de Fisher')
disp(F)
p=fdis_prb(F, 3, n-3);
disp('p_value de la stat. de Fisher')
disp(p)
```

Résultats du test :

```
Statistique de Fisher
   10.8547

p_value de la stat. de Fisher
   1.3262e-05
```

$\boxed{\text{On rejette l'hypothèse H0 car sa p\_valeur est négligeable.}}$