Panel Analyses Report

David BYAMUNGU

2021-06-20

Contents

1 Prerequis **5**

2 Introduction **7**

3 Literature **9**

4 The One-way Error Component Regression Model **11**
4.1 INTRODUCTION 11
4.2 THE FIXED EFFECTS MODEL 13
4.3 THE RANDOM EFFECTS MODEL 17

5 The Two-way Error Component Regression Model **23**
5.1 NTRODUCTION 23
5.2 THE FIXED EFFECTS MODEL 24
5.3 THE RANDOM EFFECTS MODEL 25
5.4 Test of Hypotheses with Panel Data 30
5.5 REFERENCES 30

6 Methods **31**

7 Analyses **33**
7.1 Netoyage de la base des données 33
7.2 Analyse descriptive des Vararíales 53

8 Final Words **55**

Chapter 1

Prerequis

Ceci est *une étude* des données de panel avec **Markdown**.

The **bookdown** package can be installed from CRAN or Github:

La structure de ce rapport est que chaque fichier RMD porte un chapitre et un thème bien spécifique de notre analyse

Nous avions utilisé XeLaTeX pour compiler ce document en PDF.

Chapter 2

Introduction

Dans ce document nous cherchons à modéliser les taxes perçues dans différents pays , formant un panel dont la période est 10 ans. Ainsi, nous expliquerons la variable taxe par:

(1) Le poids des Marchandises
(2) La qualité des Marchandises

Le but est d'arbitrer entre:

(1) le modèle *pooling*
(2) le *modele à éffet fixe* et
(3) le *modèle à éffet aléatoire*

et en fin produire un `modèle dynamique` permettant d'expliquer la variation du taxe au cours du temps, avec comme variable dépendante additionnelle le taxe décalé

Nous expliquons Notre méthodologie dans la partie suivante.

Chapter 3

Literature

Chapter 4

The One-way Error Component Regression Model

4.1 INTRODUCTION

A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables, i.e.

$$y_{it} = \alpha + X'_{it}\beta + u_{it}$$

$$i = 1, ..., N; t = 1, ..., T$$

(2.1)

with $ i $ denoting households, individuals, firms, countries, etc. and t denoting time. The i subscript, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension.

$$\alpha$$

is a scalar,

$$\beta$$

is

$$K \times 1$$

and Xi t is the $ it_{th} $ observation on K explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, with

$$u_{it} = \mu_i + v_{it}$$

(2.2)

where i denotes the unobservable individual-specific effect and i t denotes the remainder disturbance. For example, in an earnings equation in labor economics, yi t will measure earnings of the head of the household, whereas

$$X_{it}$$

may contain a set of variables like experience, education, union membership, sex, race, etc. Note that i is time-invariant and it accounts for any individual-specific effect that is not included in the regression. In this case we could think of it as the individual's unobserved ability. The remainder disturbance

$$v_{it}$$

varies with individuals and time and can be thought of as the usual disturbance in the regression. Alternatively, for a production function utilizing data on firms across time,

$$y_{it}$$

will measure output and

$$X_{it}$$

will measure inputs. The unobservable firm-specific effects will be captured by the

$$\mu_i$$

and we can think of these as the unobservable entrepreneurial or managerial skills of the firm's executives. Early applications of error components in economics include Kuh (1959) on investment, Mundlak (1961) and Hoch (1962) on production functions and Balestra and Nerlove (1966) on demand for natural gas. In vector form (2.1) can be written as

$$y = \alpha i_{NT} + X\beta + u = Z\delta + u$$

(2.3)

$$y = \alpha i_{NT} + X\beta + u = Z\delta + u$$

(2.3)

where $ y $ is $ NT \times 1 $, $ X $ is $ NT \times K $, $ Z = [_{NT} , X] $, $ ^{'} = (^{ {'}, }) $ and $\iota_N T$ is a vector of ones of dimension NT. Also, (2.2) can be written as

$$u = Z_\mu \mu + v$$

(2.4)

$$y_{it} = \alpha + X_{it}^{'} + U_{it}$$

, $ i=1,...,N ; t=1,...,T $ with i denoting households, individuals, firms, countries, etc. and t denoting time. The i subscript, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. $ $ is a scalar, $ $ is $ K \times 1 $ and $X_{it} $is the $ it^{th} $ observation on K explanatory variables. disturbances, with it

$$u_{it} = u_i + v_{it}$$

where i denotes the unobservable individual-specific effect and νit denotes the remainder disturbance. For example, in an earnings equation in labor economics, y_{it} will measure earnings of the head of the household, whereas Xi t may contain a set of variables like experience, education, union membership, sex, race, etc. Note that i is time-invariant and it accounts for any individual-specific effect that is not included in the regression. In this case we could think of it as the individual's unobserved ability. The remainder disturbance $ {it} $ varies with individuals and time and can be thought of as the usual disturbance in the regression. Alternatively, for a production function utilizing data on firms across time, $ y_{it} $ will measure output and $ X_{it} $ will measure inputs. The unobservable firm-specific effects will

be captured by the $ _i $ and we can think of these as the unobservable entrepreneurial or managerial skills of the firm's executives. Early applications of error components in economics include Kuh (1959) on investment, Mundlak (1961) and Hoch (1962) on production functions and Balestra and Nerlove (1966) on demand for natural gas. In vector form (2.1) can be written as

$$y = \alpha i_{NT} + X\beta + u = Z\delta + u$$

where y is $ NT × 1$, X is $NT \times K$, $ Z = [i_{NT} , X] $, $ ^{‘}=(^ ’, ^{’})$ and i_{NT} is a vector of ones of dimension NT . Also, (2.2) can be written as

$$u = Z_\mu \mu + v$$

(2.4)

where $ u^{‘} = (u_{11}, . . . , u_{1T} , u_{21}, . . . , u_{2T}, . . . , u_{N1}, . . . , u_{NT}) $ with the observations stacked such that the slower index is over individuals and the faster index is over time.$ Z_ = IN T $ where IN is an identity matrix of dimension N, T is a vector of ones of dimension T and $ $ denotes Kronecker product. $ Z_ $ is a selector matrix of ones and zeros, or simply the matrix of individual dummies that one may include in the regression to estimate the i if they are assumed to be fixed parameters. $ ^{’} = (_1, . . . , _N)$ and $\nu^{'} = (\nu 11, ..., \nu_{1T}, ..., \nu_{N1}, ..., \nu_{NT})$. Note that $Z_ Z^{‘} = I_N J_T $ where J_T is a matrix of ones of dimension T and $ P = Z (Z^{’} Z){−1} Z^{’} $, the projectionmatrix on $ Z_ $, reduces to $ IN J_T $ where $J _T= JT /T $. P is a matrix which averages the observation across time for each individual, and Q = INT − P is a matrix which obtains the deviations from individual means. For example, regressing y on the matrix of dummy variables Z_μ gets the predicted values $ P_y $ which has a typical element

$$\bar{y}_i = \sum_{t=1}^{T} \frac{y_{it}}{T}$$

repeated T times for each individual. The residuals of this regression are given by Qy which has a typical element

$$(y_{it} - \bar{y}_{i.})$$

P and Q are (i) symmetric idempotent matrices, i.e.

$P^{'} = P$ and $ P^2 = P $. This means that $rank(P) = tr(P) = N$ and $rank(Q) = tr(Q) = N(T−1)$. This uses the result that the rank of an idempotent matrix is equal to its trace (see Graybill, 1961,theorem 1.63). Also, (ii) P and Q are orthogonal, i.e. $PQ=0 $ and (iii) they sum to the identity matrix $P + Q = I_{NT}$. In fact, any two of these properties imply the third (see Graybill, 1961, theorem 1.68).

4.2 THE FIXED EFFECTS MODEL

In this case, the i are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with v_{it} independent and identically distributed $IID(0, \sigma_v^2)$. The X_{it} are assumed independent of the v_{it} for all i and t. The fixed effects model is an appropriate specification if we are focusing on a specific set of N firms, say, IBM, GE,Westinghouse, etc. and our inference is restricted to the behavior of these sets of firms. Alternatively, it could be a set of N OECD countries, or N American states. Inference in this case is conditional on the particular N firms, countries or states that are observed. One can substitute the disturbances given by (2.4) into (2.3) to get

$ y= i_{IT} + X + Z_ +v = Z + Z_ +v $ (2.5)

and then perform ordinary least squares (OLS) on (2.5) to get estimates of $, and µ $

Note that Z is $ NT ×(K+1)$ and Z , the matrix of individual dummies, is$NT ×N $ NT × N. If N is large, (2.5) will include too many individual dummies, and the matrix to be inverted by OLS is large and of dimension $(N + K)$. In fact, since and are the parameters of interest, one can obtain the LSDV (least squares dummy variables) estimator from (2.5), by premultiplying the model by Q and performing OLS on the resulting transformed model:

$ Qy=QX + Qv$ (2.6)

This uses the fact that $QZ_\mu = Qi_{NT} = 0$, since $PZ_\mu = Z_\mu$ the Q matrix wipes out the individual effects. This is a regression of $\tilde{y} = QY$ with element $(y_{it} - \bar{y}_{i.})$ on $\tilde{X} = QX$ with typical element

$$\tilde{\beta} = (X'QX)^{-1} X'Qy$$

(2.7) with $\text{var}(\tilde{\beta}) = \sigma_v^2 (X'QX)^{-1} = \sigma_v^2 \left(\widetilde{X}'\tilde{X}\right)^{-1}$. $\tilde{\beta}$ could have been obtained from (2.5) using results on partitioned inverse or the Frisch–Waugh–Lovell theorem discussed in Davidson and MacKinnon (1993, p. 19). This uses the fact that P is the projection matrix on Z_μ and $Q = I_{NT} - P$ (see problem 2.1). In addition, generalized least squares (GLS) on (2.6), using the generalized inverse, will also yield $\tilde{\beta}$ (see problem 2.2).

Note that for the simple regression

$$y_{it} = \beta x_{it} + \mu_i + v_i$$

(2.8)

and averaging over time gives

$$\bar{y}_{i.} = \beta \bar{x}_{i.} + \mu_i + \bar{v}_i$$

(2.9)

Therefore, subtracting (2.9) from (2.8) gives

$$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + (v_{it} - \bar{v}_{i.})$$

(2.10)

Also, averaging across all observations in (2.8) gives

$$\bar{y}_{..} = \alpha + \beta \bar{x}_{..} + \bar{v}_{..}$$

(2.11)i =0where we utilized the restriction that $ *i=1^{n}* i=0 $ This is an arbitrary restriction on the dummy variable coefficients to avoid the dummy variable trap, or perfect multicollinearity; see Suits (1984) for alternative formulations of this restriction. In fact only β and $ (+ *i)$ are estimable from (2.8), and not and i separately, unless a restriction like*

$$\sum_{i=1}^{n} \mu_i = 0$$

is imposed. In this case, $\tilde{\beta}$ is obtained from regression (2.10),

$$\tilde{\alpha} = \bar{y}_{..} - \tilde{\beta} \bar{x}_{..}$$

can be recovered from (2.11) and

$$\tilde{\mu}_i = \bar{y}_{i.} - \tilde{\alpha} - \tilde{\beta} \bar{X}_{i.}$$

from (2.9). For large labor or consumer panels, where N is very large,regressions like (2.5) may not be feasible, since one is including (N − 1) dummies in the regression. This fixed effects (FE) least squares,

also known as least squares dummy variables (LSDV), suffers from a large loss of degrees of freedom. We are estimating (N − 1) extra parameters, and too many dummies may aggravate the problem of multicollinearity among the regressors. In addition, this FE estimator cannot estimate the effect of any time-invariant variable like sex, race, religion, schooling or union participation. These time-invariant variables are wiped out by the Q transformation, the deviations from means transformation (see (2.10)). Alternatively, one can see that these time-invariant variables are spanned by the individual dummies in (2.5) and therefore any regression package attempting (2.5) will fail, signaling perfect multicollinearity. If (2.5) is the true model, LSDV is the best linear unbiased estimator (BLUE) as long as v_{it} is the standard classical disturbance with mean 0 and variance–covariance matrix $ v^{2} I_{NT} $. Note that as $T \to \infty$ the FE estimator is consistent. However, if T is fixed and $N \to \infty$ as is typical in short labor panels, then only the FE estimator of is consistent; the FE estimators of the individual effects $\alpha + \mu_i$ are not consistent since the number of these parameters increases as N increases. This is the incidental parameter problem discussed by Neyman and Scott (1948) and reviewed more recently by Lancaster (2000). Note that when the true model is fixed effects as in (2.5), OLS on (2.1) yields biased and inconsistent estimates of the regression parameters. This is an omission variables bias due to the fact that OLS deletes the individual dummies when in fact they are relevant.

(1) *Testing for fixed effects.* One could test the joint significance of these dummies, i.e. H0; $\mu_1 = \mu_2 = ... = \mu_{N-1} = 0$, by performing an F-test. (Testing for individual effects will be treated extensively in Chapter 4.) This is a simple Chow test with the restricted residual sums of squares (RRSS) being that of OLS on the pooled model and the unrestricted residual sums of squares (URSS) being that of the LSDV regression. If N is large, one can perform the Within transformation and use that residual sum of squares as the URSS. In this case

$$F_0 = \frac{\frac{RRSS - URSS}{N-1}}{\frac{URSS}{NT-N-K}} \sim F_{N-1,N(T-1)-K}$$

(2.12)

(2) *Computational warning.* One computational caution for those using theWithin regression given by (2.10). The s^2 f this regression as obtained from a typical regression package divides the residual sums of squares by NT − K since the intercept and the dummies are not included. The proper s^2 , say s^{*2} from the LSDV regression in (2.5), would divide the same residual sums of squares by $N(T-1) - K$. Therefore, one has to adjust the variances obtained from the Within regression (2.10) by multiplying the variance–covariance matrix by

$$\frac{s^2}{s^{*2}}$$

or simply by multiplying by $[NT - K]/[N(T-1) - K]$

(3) *Robust estimates of the standard errors.* For the Within estimator, Arellano (1987) suggests a simple method for obtaining robust estimates of the standard errors that allow for a general variance–covariance matrix on the v_{it} as in White (1980). One would stack the panel as an equation for each individual:

$$y_i = Z_i\delta + \mu i_{it} + v_i$$

(2.13)

where y_i is T × $Z_i =[l_T,X_i] $, X_i is T × K, μ_i is a scalar, $\delta' = (\alpha, \beta')$, i_T is a vector of ones of dimension T and v_i is T x 1 . In general, $E(v_i, v_i') = \Omega_i$ for i = 1, 2, . . . , N, where Ω_i is a positive definite matrix of dimension T . We still assume $E(v_i, v_j') = 0$ for $i \neq j$ T is assumed small and N large as in household or company panels, and the asymptotic results are performed for $N \to \infty$ and T fixed. Performing the Within transformation on this set of equations (2.13) one gets

$$\tilde{y}_i = \widetilde{X}_i\beta + \tilde{v}_i$$

(2.14)

where

$$\tilde{y} = Qy$$

,

$$\widetilde{X} = QX$$

and

$$\tilde{v} = Qv$$

, with

$$\tilde{y} = (\tilde{y}_1', ..., \tilde{y}_N')'$$

and

$$\tilde{y}_i = (I_T - \bar{J}_T)y_i$$

Computing robust least squares on this system, as described by White (1980), under the restriction that each equation has the same β one gets the Within estimator of which has the following asymptotic distribution:

$$N^{\frac{1}{2}}(\tilde{\beta} - \beta) \sim N(0, M^{-1}VM^{-1})$$

(2.15)

where

$$M = \frac{p\lim(\widetilde{X}'\widetilde{X})}{N}$$

Note that

$$\widetilde{X}_i = (I_T - \bar{J}_T)X_i$$

and

$$\widetilde{X}'diag[\Omega_i]Q\widetilde{X}$$

(see problem 2.3). In this case, V is estimated by

$$\frac{\widetilde{V} = \sum_{i=1}^{N} \widetilde{X}_i'\tilde{u}_i\tilde{u}_i'\widetilde{X}_i'}{N}$$

where

$$\tilde{u}_i = \tilde{y}_i - \widetilde{X}_i\tilde{\beta}_i$$

. Therefore, the robust asymptotic variance– covariance matrix of β is estimated by

$$\text{var}(\tilde{\beta}) = \left(\widetilde{X}'\tilde{X}\right)^{-1}\left[\sum_{i=1}^{N}\widetilde{X}_i'\tilde{u}_i\tilde{u}_i'\widetilde{X}_i\right]\left(\widetilde{X}'\tilde{X}\right)^{-1}$$

4.3 THE RANDOM EFFECTS MODEL

There are too many parameters in the fixed effects model and the loss of degrees of freedom can be avoided if the i can be assumed random. In this case $ *i* IID(0, ^2)$, $ _{it} IID(0, _v^2) $ and the μ_i are independent of the v_{it} . In addition, the X_{it} are independent of the μ_i and v_{it},for all i and t. The random effects model is an appropriate specification if we are drawing N individuals randomly from a large population. This is usually the case for household panel studies. Care is taken in the design of the panel to make it "representative" of the population we are trying to make inferences about. In this case, N is usually large and a fixed effects model would lead to an enormous loss of degrees of freedom. The individual effect is characterized as random and inference pertains to the population from which this sample was randomly drawn.

But what is the population in this case? Nerlove and Balestra (1996) emphasize Haavelmo's (1944) view that the population "consists not of an infinity of individuals, in general, but of an infinity of decisions" that each individual might make. This view is consistent with a random effects specification. From (2.4), one can compute the variance–covariance matrix

$$\begin{aligned}\Omega = E(uu') &= Z_\mu E(\mu\mu') Z'_\mu + E(vv') \\ &= \sigma^2_\mu (I_N \otimes J_T) + \sigma^2_v (I_N \otimes I_T)\end{aligned} \tag{4.1}$$

This implies a homoskedastic variance $\text{var}(u_{it}) = \sigma^2_\mu + \sigma^2_v$f for all i and t, and an equicorrelated block-diagonal covariance matrix which exhibits serial correlation over time only between the disturbances of the same individual. In fact,

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma^2_\mu + \sigma^2_v \quad \text{for} \quad i = j, t = s \\ &= \sigma^2_\mu \qquad\qquad \text{for } i = j, t \neq s\end{aligned}$$

and zero otherwise. This also means that the correlation coefficient between μ_{it} and μ_{js} is

$$\begin{aligned}\rho = \text{correl}(u_{it}, u_{js}) &= 1 \qquad\qquad\qquad\qquad\qquad \text{for } i = j, t = s \\ &= \sigma^2_\mu / (\sigma^2_\mu + \sigma^2_v) \quad \text{for } i = j, t \neq s\end{aligned}$$

and zero otherwise. In order to obtain the GLS estimator of the regression coefficients, we need $ ^{-1} $. This is a huge matrix for typical panels and is of dimension $NT \times NT$. No brute force inversion should be attempted even if the researcher's application has a small N and T .1 We will follow a simple trick devised by Wansbeek and Kapteyn (1982b, 1983) that allows the derivation of $ ^{-1} $ ans $\Omega^{-1/2}$ Essentially, one replaces J_T by $T\bar{J}_T$ and I_T by $(E_T + \bar{J}_T)$ where E_T is by definition $(I_T - \bar{J}_T)$. In this case

$$\Omega = T\sigma^2_\mu (I_N \otimes \bar{J}_T) + \sigma^2_v (I_N \otimes E_T) + \sigma^2_v (I_N \otimes \bar{J}_T)$$

Collecting terms with the same matrices, we get

$$\Omega = (T\sigma^2_\mu + \sigma^2_v)(I_N \otimes \bar{J}_T) + \sigma^2_v (I_N \otimes E_T) = \sigma^2_1 P + \sigma^2_v Q \tag{4.2}$$

where $\sigma^2_1 = T\sigma^2_\mu + \sigma^2_v$. (2.18) is the spectral decomposition representation of Ω, with σ^2_1 being the first unique characteristic root ofΩ of multiplicity $N(T-1)$. It is easy to verify, using the properties of P and Q, that

$$\Omega^{-1} = \frac{1}{\sigma^2_1} P + \frac{1}{\sigma^2_v} Q \tag{4.3}$$

and

$$\Omega^{-1/2} = \frac{1}{\sigma_1} P + \frac{1}{\sigma_v} Q \tag{4.4}$$

In fact, $\Omega^r = (\sigma_1^2)^r P + (\sigma_v^2)^r Q$ where r is an arbitrary scalar. Now we can obtain GLS as a weighted least squares. Fuller and Battese (1973, 1974) suggested premultiplying the regression equation given in (2.3) by $\sigma_v \Omega^{-1/2} = Q + (\sigma_v/\sigma_1) P$ and performing OLS on the resulting transformed regression. In this case, $y^* = \sigma_v \Omega^{-1/2} y$ has a typical element $y_{it} - \theta \bar{y}_{i.}$ where $\theta = 1 - (\sigma_v/\sigma_1)$ (see problem 2.4). This transformed regression inverts a matrix of dimension $(K+1)$ and can easily be implemented using any regression package.

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the spectral decomposition of Ω . In fact, $Pu \sim (0, \sigma_1^2 P)$ and $Qu \sim (0, \sigma_v^2 Q)$ and

$$\hat{\sigma}_1^2 = \frac{u' P u}{\text{tr}(P)} = T \sum_{i=1}^{N} \bar{u}_{i.}^2 / N \tag{4.5}$$

and

$$\hat{\sigma}_v^2 = \frac{u' Q u}{\text{tr}(Q)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (u_{it} - \bar{u}_{i.})^2}{N(T-1)} \tag{4.6}$$

provide the BQU estimators of σ_1^2 andσ_v^2 , respectively (see problem 2.5).

These are analyses of variance-type estimators of the variance components and are minimum variance-unbiased under normality of the disturbances (see Graybill, 1961). The true disturbances are not known and therefore (2.21) and (2.22) are not feasible. Wallace and Hussain (1969) suggest substituting OLS residual $\hat{u}_{\text{OLS}}$ instead of the true u. After all, under the random effects model, the OLS estimates are still unbiased and consistent, but no longer efficient. Amemiya (1971) shows that these estimators of the variance components have a different asymptotic distribution from that knowing the true disturbances. He suggests using the LSDV residuals instead of the OLS residuals. In this case $\tilde{u} = y - \tilde{\alpha} \iota_{NT} - X\tilde{\beta}$ where and $\bar{X}'_{..}$ is a $1 \times K$

vector of averages of all regressors. Substituting these $\hat{u}$ for u in (2.21) and (2.22) we get the Amemiya-type estimators of the variance components. The resulting estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}\left(\hat{\sigma}_v^2 - \sigma_v^2\right) \\ \sqrt{N}\left(\hat{\sigma}_\mu^2 - \sigma_\mu^2\right) \end{pmatrix} \sim N\left(0, \begin{pmatrix} 2\sigma_v^4 & 0 \\ 0 & 2\sigma_\mu^4 \end{pmatrix}\right) \tag{4.7}$$

where $\hat{\sigma}_\mu^2 = \left(\hat{\sigma}_1^2 - \hat{\sigma}_v^2\right)/T$.[3]

Swamy and Arora (1972) suggest running two regressions to get estimates of the variance components from the corresponding mean square errors of these regressions. The first regression is the Within regression, given in (2.10), which yields the following s^2:

$$\hat{\sigma}_v^2 = \left[y' Q y - y' Q X \left(X' Q X\right)^{-1} X' Q y\right] / [N(T-1) - K] \tag{4.8}$$

The second regression is the Between regression which runs the regression of averages across time, i.e.

$$\bar{y}_{i.} = \alpha + \bar{X}'_{i.} \beta + \bar{u}_{i.} \quad i = 1, \ldots, N \tag{4.9}$$

(2.25)

This is equivalent to premultiplying the model in (2.5) by P and running OLS. The only caution is that the latter regression has NT observations because it repeats the averages T times for each individual,

while the cross-section regression in (2.25) is based on N observations. To remedy this, one can run the cross-section regression

$$\sqrt{T}\bar{y}_{i.} = \alpha\sqrt{T} + \sqrt{T}\bar{X}'_{i.}\beta + \sqrt{T}\bar{u}_{i.} \tag{4.10}$$

(2.26)

where one can easily verify that var $\left(\sqrt{T}\bar{u}_{i.}\right) = \sigma_1^2$. This regression will yield an s^2 given by

$$\hat{\sigma}_1^2 = \left(y'Py - y'PZ\left(Z'PZ\right)^{-1}Z'Py\right)/(N-K-1) \tag{4.11}$$

(2.27)

Note that stacking the following two transformed regressions we just performed yields

$$\begin{pmatrix} Qy \\ Py \end{pmatrix} = \begin{pmatrix} QZ \\ PZ \end{pmatrix}\delta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} \tag{4.12}$$

(2.28)

and the transformed error has mean 0 and variance–covariance matrix given by

$$\begin{pmatrix} \sigma_v^2 Q & 0 \\ 0 & \sigma_1^2 P \end{pmatrix}$$

Problem 2.7 asks the reader to verify that OLS on this system of 2NT observations yields OLS on the pooled model (2.3). Also, GLS on this system yields GLS on (2.3). Alternatively, one could get rid of the constant by running the following stacked regressions:

$$\begin{pmatrix} Qy \\ \left(P-\bar{J}_{NT}\right)y \end{pmatrix} = \begin{pmatrix} QX \\ \left(P-\bar{J}_{NT}\right)X \end{pmatrix}\beta + \begin{pmatrix} Qu \\ \left(P-\bar{J}_{NT}\right)u \end{pmatrix} \tag{4.13}$$

(2.29)

This follows from the fact that Q NT = 0 and (P − J⁻NT) NT = 0. The transformed error has zero mean and variance–covariance matrix

$$\begin{pmatrix} \sigma_v^2 Q & 0 \\ 0 & \sigma_1^2\left(P-\bar{J}_{NT}\right) \end{pmatrix}$$

OLS on this system yields OLS on (2.3) and GLS on (2.29) yields GLS on (2.3). In fact,

$$\begin{aligned} \hat{\beta}_{\text{GLS}} &= \left[(X'QX/\sigma_v^2) + X'\left(P-\bar{J}_{NT}\right)X/\sigma_1^2\right]^{-1}\left[(X'Qy/\sigma_v^2)\right. \\ &\quad \left.+X'\left(P-\bar{J}_{NT}\right)y/\sigma_1^2\right] \\ &= \left[W_{XX}+\phi^2 B_{XX}\right]^{-1}\left[W_{Xy}+\phi^2 B_{Xy}\right] \end{aligned} \tag{4.14}$$

(2.30)

with var $\left(\hat{\beta}_{\text{GLS}}\right) = \sigma_v^2\left[W_{XX}+\phi^2 B_{XX}\right]^{-1}$. Note that $W_{XX} = X'QX, B_{XX} = X'\left(P-\bar{J}_{NT}\right)X$ and $\phi^2 = \sigma_v^2/\sigma_1^2$. Also, the Within estimator of β is $\underline{\beta}_{\text{Within}} = W_{XX}^{-1}W_{Xy}$ and the Between estimator of β is $\hat{\beta}_{\text{Between}} = B_{XX}^{-1}B_{Xy}$. This shows that $\hat{\beta}_{\text{GLS}}$ is a matrix weighted average of $\tilde{\beta}_{\text{Within and}}\ \hat{\beta}_{\text{Between}}$ weighing each estimate by the inverse of its corresponding variance. In fact

$$\hat{\beta}_{\text{GLS}} = W_1\tilde{\beta}_{\text{Within}} + W_2\hat{\beta}_{\text{Between}} \tag{4.15}$$

(2.31) where

$$W_1 = [W_{XX} + \phi^2 B_{XX}]^{-1} W_{XX}$$

and

$$W_2 = [W_{XX} + \phi^2 B_{XX}]^{-1} (\phi^2 B_{XX}) = I - W_1$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = 0$ then $\phi^2 = 1$ and $\hat{\beta}_{\text{GLS}}$ reduces to $\hat{\beta}_{\text{OLs.}}$ (ii) If $T \to \infty$, then $\phi^2 \to 0$ and $\hat{\beta}_{\text{GLS}}$ tends to $\tilde{\beta}_{\text{Within.}}$ Also, if W_{XX} is huge compared to B_{XX} then $\hat{\beta}_{\text{GLS}}$ will be close to $\tilde{\beta}_{\text{Within}}$. However, if B_{XX} dominates W_{XX} then $\hat{\beta}_{\text{GLS}}$ tends to $\hat{\beta}_{\text{Between.}}$ In other words, the Within estimator ignores the Between variation, and the Between estimator ignores the Within variation. The OLS estimator gives equal weight to the Between and Within variations. From (2.30), it is clear that $\text{var}\left(\tilde{\beta}_{\text{Within}}\right) - \text{var}\left(\hat{\beta}_{\text{GLS}}\right)$ is a positive semidefinite matrix, since ϕ^2 is positive. However, as $T \to \infty$ for any fixed $N, \phi^2 \to 0$ and both $\hat{\beta}_{\text{GLS}}$ and $\tilde{\beta}_{\text{Within}}$ have the same asymptotic variance.

Another estimator of the variance components was suggested by Nerlove (1971a). His zuggestion is to estimate σ^2 as $\sum^N \cdot(\hat{u} \cdot - \hat{\pi})^2/(N-1)$ where $\hat{u}.$ are the dummv coefficients estimates from the LSDV regression. σ_v^2 is estimated from the Within residual sums of squares divided by NT without correction for degrees of freedom. [4]

Note that, except for Nerlove's (1971a) method, one has to retrieve $\hat{\sigma}_\mu^2$ as $(\hat{\sigma}_1^2 - \hat{\sigma}_v^2)/T$. In this case, there is no guarantee that the estimate of $\hat{\sigma}_\mu^2$ would be nonnegative. Searle (1971) has an extensive discussion of the problem of negative estimates of the variance components in the biometrics literature. One solution is to replace these negative estimates by zero. This in fact is the suggestion of the Monte Carlo study by Maddala and Mount (1973). This study finds that negative estimates occurred only when the true σ_μ^2 was small and close to zero. In these cases OLS is still a viable estimator. Therefore, replacing negative $\hat{\sigma}_\mu^2$ by zero is not a bad sin after all, and the problem is dismissed as not being serious. [5]

How about the properties of the various feasible GLS estimators of β Under the random effects model, GLS based on the true variance components is BLUE, and all the feasible GLS estimators considered are asymptotically efficient as either N or $N \to \infty$. Maddala and Mount (1973) compared OLS, Within, Between, feasible GLS methods, MINQUE, Henderson's method III, true GLS and maximum likelihood estimation using their Monte Carlo study. They found little to choose among the various feasible GLS estimators in small samples and argued in favor of methods that were easier to compute. MINQUE was dismissed as more difficult to compute and the applied researcher given one shot at the data was warned to compute at least two methods of estimation, like an ANOVA feasible GLS and maximum likelihood to ensure that they do not yield drastically different results. If they do give different results, the authors diagnose misspecification.

Taylor (1980) derived exact finite sample results for the one-way error component model. He compared the Within estimator with the Swamy–Arora feasible GLS estimator. He found the following important results:

(1) Feasible GLS is more efficient than LSDV for all but the fewest degrees of freedom.
(2) The variance of feasible GLS is never more than 17% above the Cramer–Rao lower bound.
(3) More efficient estimators of the variance components do not necessarily yield more efficient feasible GLS estimators.

These finite sample results are confirmed by the Monte Carlo experiments carried out by Maddala and Mount (1973) and Baltagi (1981a).

Bellmann, Breitung and Wagner (1989) consider the bias in estimating the variance components using the Wallace and Hussain (1969) method due to the replacement of the true disturbances by OLS residuals, also the bias in the regression coefficients due to the use of estimated variance components rather than the true variance components. The magnitude of this bias is estimated using bootstrap methods for

two economic applications. The first application relates product innovations, import pressure and factor inputs using a panel at the industry level. The second application estimates the earnings of 936 full-time working German males based on the first and second wave of the German Socio-Economic Panel. Only the first application revealed considerable bias in estimating $ _ ^2 $. However, this did not affect the bias much in the corresponding regression coefficients

4.3.1 Fixed vs Random

Having discussed the fixed effects and the random effects models and the assumptions underlying them, the reader is left with the daunting question, which one to choose? This is not as easy a choice as it might seem. In fact, the fixed versus random effects issue has generated a hot debate in the biometrics and statistics literature which has spilled over into the panel data econometrics literature. Mundlak (1961) and Wallace and Hussain (1969) were early proponents of the fixed effects model and Balestra and Nerlove (1966) were advocates of the random error component model. In Chapter 4, we will study a specification test proposed by Hausman (1978) which is based on the difference between the fixed and random effects estimators. Unfortunately, applied researchers have interpreted a rejection as an adoption of the fixed effects model and nonrejection as an adoption of the random effects model.6 Chamberlain (1984) showed that the fixed effects model imposes testable restrictions on the parameters of the reduced form model and one should check the validity of these restrictions before adopting the fixed effects model (see Chapter 4). Mundlak (1978) argued that the random effects model assumes exogeneity of all the regressors with the random individual effects. In contrast, the fixed effects model allows for endogeneity of all the regressors with these individual effects. So, it is an "all" or "nothing" choice of exogeneity of the regressors and the individual effects, see Chapter 7 for a more formal discussion of this subject.

Hausman and Taylor (1981) allowed for some of the regressors to be correlated with the individual effects, as opposed to the all or nothing choice. These over-identification restrictions are testable using a Hausman-type test (see Chapter 7). For the applied researcher, performing fixed effects and random effects and the associated Hausman test reported in standard packages like Stata, LIMDEP, TSP, etc., the message is clear: Do not stop here. Test the restrictions implied by the fixed effects model derived by Chamberlain (1984) (see Chapter 4) and check whether a Hausman and Taylor (1981) specification might be a viable alternative (see Chapter 7).

Chapter 5

The Two-way Error Component Regression Model

5.1 NTRODUCTION

Wallace and Hussain (1969), Nerlove (1971b) and Amemiya (1971), among others, the regression model given by (2.1), but with two-way error components disturbances:

$$u_{it} = \mu_i + \lambda_t + v_{it} \quad i = 1, ..., N; t = 1, ..., T \tag{5.1}$$

(3.1)

where μ_i denotes the unobservable individual effect discussed in Chapter 2, λ_t denotes the unobservable time effect and v_{it} is the remainder stochastic disturbance term. Note that λ_t is individual-invariant and it accounts for any time-specific effect that is not included in the regression. For example, it could account for strike year effects that disrupt production; oil embargo effects that disrupt the supply of oil and affect its price; Surgeon General reports on the ill-effects of smoking, or government laws restricting smoking in public places, all of which could affect consumption behavior. In vector form, (3.1) can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + v \tag{5.2}$$

(3.2)

where Z_μ, μ and v were defined earlier. $Z_\lambda = i_N \otimes I_T$ is the matrix of time dummies that one may include in the regression to estimate the λ_t if they are fixed parameters, and $\lambda' = (\lambda_1, ..., \lambda_T)$ Note that

$$Z_\lambda Z_\lambda' = J_N \otimes I_T$$

and the projection on Z_λ is

$$Z_\lambda \left(Z_\lambda' Z_\lambda\right)^{-1} Z_\lambda' = \bar{J}_N \otimes I_T$$

This last matrix averages the data over individuals, i.e., if we regress y on Z_λ, the predicted values are given by

$$(\bar{J_N} \otimes I_T) y$$

which has typical element $\bar{y}_{.t} = \sum_{i=1}^{N} y_{it}/N$.

5.2 THE FIXED EFFECTS MODEL

If the μ_i and λ_t are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with $v_{it} \sim \text{IID}\left(0, \sigma_v^2\right)$, then (3.1) represents a two-way fixed effects error component model. The X_{it} are assumed independent of the v_{it} for all i and t. Inference in this case is conditional on the particular N individuals and over the specific time periods observed. Recall that Z_λ, the matrix of time dummies, is $NT \times T$. If N or T is large, there will be too many dummy variables in the regression $\{(N-1)+(T-1)\}$ of them, and this causes an enormous loss in degrees of freedom. In addition, this attenuates the problem of multicollinearity among the regressors. Rather than invert a large $(N+T+K-1)$ matrix, one can obtain the fixed effects estimates of β by performing the following Within transformation given by Wallace and Hussain (1969):

$$Q = E_N \otimes E_T = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \tag{5.3}$$

(3.3)

where $E_N = I_N - \bar{J}_N$ and $E_T = I_T - \bar{J}_T$. This transformation "sweeps" the μ_i and λ_t effects. In fact, $\tilde{y} = Qy$ has a typical element $\tilde{y}_{it} = (y_{it} - \bar{y}_{i.} - \bar{y}_{it} + \bar{y}..)$ where $\bar{y}_{..} = \sum_i \sum_t y_{it}/\ NT$, and one would perform the regression of $\tilde{y} = Qy$ on $\widetilde{X} = QX$ to get the Within estimator $\tilde{\beta} = \left(X'QX\right)^{-1} X'Qy$

Note that by averaging the simple regression given in (2.8) over individuals, we get

$$\bar{y}_{.t} = \alpha + \beta \bar{x}_{.t} + \lambda_t + \bar{v}_{.t} \tag{5.4}$$

(3.4)

where we have utilized the restriction that $\sum_i \mu_i = 0$ to avoid the dummy variable trap. Similarly the averages defined in (2.9) and (2.11) still hold using $\sum_t \lambda_t = 0$ and one can deduce that

$$(y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}) = (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})\,\beta + (v_{it} - \bar{v}_{i.} - \bar{v}_{.t} + \bar{v}_{..}) \tag{5.5}$$

(3.5)

OLS on this model gives $\tilde{\beta}$ the Within estimator for the two-way model. Once again, the Within estimate of the intercept can be deduced from

$$\tilde{\alpha} = \bar{y}_{.} - \tilde{\beta}\bar{x}$$

and those of μ_i and λ_i are given by

$$\tilde{\mu}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \tilde{\beta}\,(\bar{x}_{i.} - \bar{x}_{..}) \tag{5.6}$$

(3.6)

$$\tilde{\lambda}_t = (\bar{y}_{.t} - \bar{y}_{..}) - \tilde{\beta}\,(\bar{x}_{.t} - \bar{x}_{..}) \tag{5.7}$$

(3.7)

Note that the Within estimator cannot estimate the effect of time-invariant and individualinvariant variables because the Q transformation wipes out these variables. If the true model is a two-way fixed effects model as in (3.2), then OLS on (2.1) yields biased and inconsistent estimates of the regression coefficients. OLS ignores both sets of dummy variables, whereas the one-way fixed effects estimator considered in Chapter 2 ignores only the time dummies. If these time dummies are statistically significant, the one-way fixed effects estimator will also suffer from omission bias.

5.2.1 Testing for Fixed Effects

As in the one-way error component model case, one can test for joint significance of the dummy variables:

$$H_0 : \mu_1 = ... = \mu_{N-1} = 0 \quad \text{and} \quad \lambda_1 = ... = \lambda_{T-1} = 0$$

The restricted residual sums of squares (RRSS) is that of pooled OLS and the unrestricted residual sums of squares (URSS) is that from the Within regression in (3.5). In this case,

$$F_1 = \frac{(\text{RRSS} - \text{URSS})/(N+T-2)}{\text{URSS}/(N-1)(T-1)-K} \overset{H_0}{\sim} F_{(N+T-2),(N-1)(T-1)-K} \tag{5.8}$$

(3.8)

Next, one can test for the existence of individual effects allowing for time effects, i.e. $H_2 : \mu_1 == u_N = 0$ allowing $\lambda \neq 0$ for $t = 1 ... \quad T-1$

The URSS is still the Within residual sum of squares. However, the RRSS is the regression with time-series dummies only, or the regression based upon

$$(y_{it} - \bar{y}_{.t}) = (x_{it} - \bar{x}_{.t})\beta + (u_{it} - \bar{u}_{.t}) \tag{5.9}$$

(3.9)

In this case the resulting F -statistic is $F_2 \overset{H_0}{\sim} F_{(N-1),(N-1)(T-1)-K}$. Note that F_2 differs from F_0 in (2.12) in testing for $\mu_i = 0$. The latter tests $H_0 : \mu_i = 0$ assuming that $\lambda_t = 0$, whereas the former tests $H_2 : \mu_i = 0$ allowing $\lambda_t \neq 0$ for $t = 1, ..., T-1$. Similarly, one can test for the existence of time effects allowing for individual effects, i.e.

$$H_3 : \lambda_1 = ... = \lambda_{T-1} = 0 \quad \text{allowing} \quad \mu_i \neq 0; i = 1, ..., (N-1) \tag{5.10}$$

The RRSS is given by the regression in (2.10), while the URSS is obtained from the regression (3.5). In this case, the resulting F-statistic is $F_3 \sim^{H_0} F_{(T-1),(N-1)(T-1)-K}$.

Computational Warning

As in the one-way model, s^2 from the regression in (3.5) as obtained from any standard regression package has to be adjusted for loss of degrees of freedom. In this case, one divides by$ (N − 1)(T − 1) − K $ and multiplies by $(NT − K) $to get the proper variance–covariance matrix of the Within estimator.

5.3 THE RANDOM EFFECTS MODEL

If $\mu_i \sim \text{IID}\left(0, \sigma_\mu^2\right), \lambda_t \sim \text{IID}\left(0, \sigma_\lambda^2\right)$ and $v_{it} \sim \text{IID}\left(0, \sigma_v^2\right)$ independent of each other, then this is the two-way random effects model. In addition, X_{it} is independent of μ_i, λ_t and v_{it} for all i and t. Inference in this case pertains to the large population from which this sample was randomly drawn. From (3.2), one can compute the variance-covariance matrix

$$\begin{aligned} \Omega = E\left(uu'\right) &= Z_\mu E\left(\mu\mu'\right) Z_\mu' + Z_\lambda E\left(\lambda\lambda'\right) Z_\lambda' + \sigma_v^2 I_{NT} \\ &= \sigma_\mu^2 \left(I_N \otimes J_T\right) + \sigma_\lambda^2 \left(J_N \otimes I_T\right) + \sigma_v^2 \left(I_N \otimes I_T\right) \end{aligned} \tag{5.11}$$

(3.10)

The disturbances are homoskedastic with $\text{var}\left(u_{it}\right) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2$ for all i and t,

$$\begin{aligned} \operatorname{cov}\left(u_{it}, u_{js}\right) &= \sigma_\mu^2 \quad i = j, t \neq s \\ &= \sigma_\lambda^2 \quad i \neq j, t = s \end{aligned} \tag{5.12}$$

(3.11)

and zero otherwise. This means that the correlation coefficient

$$\begin{aligned} \operatorname{correl}\left(u_{it}, u_{js}\right) &= \sigma_\mu^2 / \left(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2\right) \quad i = j, t \neq s \\ &= \sigma_\lambda^2 / \left(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2\right) \quad i \neq j, t = s \\ &= 1 \quad i = j, t = s \\ &= 0 \quad i \neq j, t \neq s \end{aligned} \tag{5.13}$$

(3.12)

In order to get Ω^{-1}, we replace J_N by $N\bar{J}_N, I_N$ by $E_N + \bar{J}_N, J_T$ by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$ and collect terms with the same matrices. This gives

$$\Omega = \sum_{i=1}^{4} \lambda_i Q_i \tag{5.14}$$

(3.13)

where $\lambda_1 = \sigma_v^2, \lambda_2 = T\sigma_\mu^2 + \sigma_v^2, \lambda_3 = N\sigma_\lambda^2 + \sigma_v^2$ and $\lambda_4 = T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_v^2$. Correspond- ingly, $Q_1 = E_N \otimes E_T, Q_2 = E_N \otimes \bar{J}_T, Q_3 = \bar{J}_N \otimes E_T$ and $Q_4 = \bar{J}_N \otimes \bar{J}_T$, respectively. The λ_i are the distinct characteristic roots of Ω and the Q_i are the corresponding matrices of eigenprojectors. λ_1 is of multiplicity $(N-1)(T-1), \lambda_2$ is of multiplicity $(N-1), \lambda_3$ is of multiplicity $(T-1)$ and λ_4 is of multiplicity 1.[1] Each Q_i is symmetric and idempotent with its rank equal to its trace. Moreover, the Q_i are pairwise orthogonal and sum to the identity matrix. The advantages of this spectral decomposition are that

$$\Omega^r = \sum_{i=1}^{4} \lambda_i^r Q_i \tag{5.15}$$

(3.14)

where r is an arbitrary scalar so that

$$\sigma_v \Omega^{-1/2} = \sum_{i=1}^{4} \left(\sigma_v / \lambda_i^{1/2}\right) Q_i \tag{5.16}$$

(3.15)

and the typical element of $y^* = \sigma_\nu \Omega^{-1/2} y$ is given by

$$y_{it}^* = y_{it} - \theta_1 \bar{y}_{i.} - \theta_2 \bar{y}_{.t} + \theta_3 \bar{y}_{..} \tag{5.17}$$

(3.16)

where $\theta_1 = 1 - \left(\sigma_v / \lambda_2^{1/2}\right), \theta_2 = 1 - \left(\sigma_v / \lambda_3^{1/2}\right)$ and $\theta_3 = \theta_1 + \theta_2 + \left(\sigma_v / \lambda_4^{1/2}\right) - 1$. As a result, GLS can be obtained as OLS of y^* on Z^*, where $Z^* = \sigma_\nu \Omega^{-1/2} Z$. This transformation was first derived by Fuller and Battese (1974), see also Baltagi (1993).

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the fact that $Q_i u \sim (0, \lambda_i Q_i)$. Hence,

$$\hat{\lambda}_i = u' Q_i u / \operatorname{tr}(Q_i) \tag{5.18}$$

(3.17)

is the BQU estimator of λ_i for $i = 1, 2, 3$. These ANOVA estimators are minimum variance unbiased (MVU) under normality of the disturbances (see Graybill, 1961). As in the one-way error component model, one can obtain feasible estimates of the variance components by replacing the true disturbances by OLS residuals (see Wallace and Hussain, 1969). OLS is still an unbiased and consistent estimator under the random effects model, but it is inefficient and results in biased standard errors and t -statistics. Alternatively, one could substitute the Within residuals with $\tilde{u} = y - \tilde{\alpha}\iota_{NT} - X\tilde{\beta}$, where $\tilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\tilde{\beta}$ and $\tilde{\beta}$ is obtained by the regression in (3.5). This is the method proposed by Amemiya (1971). In fact, Amemiya (1971) shows that the Wallace and Hussain (1969) estimates of the variance components have a different asymptotic distribution from that knowing the true disturbances, while the Amemiya (1971) estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}\left(\hat{\sigma}_v^2 - \sigma_v^2\right) \\ \sqrt{N}\left(\hat{\sigma}_\mu^2 - \sigma_\mu^2\right) \\ \sqrt{T}\left(\hat{\sigma}^2 - \sigma_\mu^2\right) \end{pmatrix} \sim N\left(0, \begin{pmatrix} 2\sigma_v^4 & 0 & 0 \\ 0 & 2\sigma_\mu^4 & 0 \\ 0 & 0 & 2\sigma^4 \end{pmatrix}\right) \tag{5.19}$$

(3.18)

SubstitutingOLSorWithin residuals instead of the true disturbances in (3.17) introduces bias in the corresponding estimates of the variance components. The degrees of freedom corrections that make these estimates unbiased depend upon traces of matrices that involve the matrix of regressors X. These corrections are given in Wallace and Hussain (1969) and Amemiya (1971), respectively. Alternatively, one can infer these correction terms from the more general unbalanced error component model considered in Chapter 9. Swamy and Arora (1972) suggest running three least squares regressions and estimating the variance components from the corresponding mean square errors of these regressions. The first regression corresponds to the Within regression which transforms the original model by $Q_1 = E_N \otimes E_T$. This is equivalent to the regression in (3.5), and yields the following estimate of σ_v^2 :

$$\hat{\lambda}_1 = \hat{\sigma}_v^2 = \left[y' Q_1 y - y' Q_1 X \left(X' Q_1 X\right)^{-1} X' Q_1 y\right] / [(N-1)(T-1) - K] \tag{5.20}$$

(3.19)
(à arranger)

The second regression is the Between individuals regression which transforms the original model by

$$Q_2 = E_N \otimes \bar{J}_T$$

This is equivalent to the regression of

$$(\bar{y}_{i.} - \bar{y}_{i..})$$

on

$$(\bar{X}_{i.} - \bar{X}_{i..})$$

and yields the following estimate of \$ _2=T _v^2 + _v^2 \$:

$$\hat{\lambda}_2 = \left[y' Q_2 y - y' Q_2 X \left(X' Q_2 X\right)^{-1} X' Q_2 y\right] / [(N-1) - K] \tag{5.21}$$

(3.20)

from which one obtains $\hat{\sigma}_\mu^2 = \left(\hat{\lambda}_2 - \hat{\sigma}_v^2\right)/T$. The third regression is the Between time-periods regression which transforms the original model by $Q_3 = \bar{J}_N \otimes E_T$. This is equivalent to the regression of $(\bar{y}_{.t} - \bar{y}_{..})$ on $\left(\bar{X}_{.t} - \bar{X}_{..}\right)$ and yields the following estimate of $\lambda_3 = N\sigma_\lambda^2 + \sigma_v^2$

$$\hat{\lambda}_3 = \left[y'Q_3y - y'Q_3X\left(X'Q_3X\right)^{-1}X'Q_3y\right]/[(T-1)-K] \tag{5.22}$$

(3.21)

from which one obtains

$$(\widehat{\widehat{\sigma_\lambda^2}} = \hat{\hat{\lambda}}_3 - \hat{\hat{\lambda}}_v)/N$$

Stacking the three transformed regressions just performed yields

$$\begin{pmatrix} Q_1y \\ Q_2y \\ Q_3y \end{pmatrix} = \begin{pmatrix} Q_1X \\ Q_2X \\ Q_3X \end{pmatrix}\beta + \begin{pmatrix} Q_1u \\ Q_2u \\ Q_3u \end{pmatrix} \tag{5.23}$$

(3.22)

since $Q_i\iota_{NT} = 0$ for $i = 1, 2, 3$, and the transformed error has mean 0 and variance-covariance matrix given by diag $[\lambda_i Q_i]$ with $i = 1, 2, 3$. Problem 3.4 asks the reader to show that OLS on this system of $3NT$ observations yields the same estimator of β as OLS on the pooled model (2.3). Also, GLS on this system of equations (3.22) yields the same estimator of β as GLS on (2.3). In fact,

$$\begin{aligned}\hat{\beta}_{\text{GLS}} &= \left[\left(X'Q_1X\right)/\sigma_v^2 + \left(X'Q_2X\right)/\lambda_2 + \left(X'Q_3X\right)/\lambda_3\right]^{-1} \\ &\quad \times \left[\left(X'Q_1y\right)/\sigma_v^2 + \left(X'Q_2y\right)/\lambda_2 + \left(X'Q_3y\right)/\lambda_3\right] \\ &= \left[W_{XX} + \phi_2^2B_{XX} + \phi_3^2C_{XX}\right]^{-1}\left[W_{Xy} + \phi_2^2B_{Xy} + \phi_3^2C_{Xy}\right]\end{aligned} \tag{5.24}$$

(3.23)

with $\text{var}\left(\hat{\beta}_{\text{GLS}}\right) = \sigma_v^2\left[W_{XX} + \phi_2^2B_{XX} + \phi_3^2C_{XX}\right]^{-1}$. Note that $W_{XX} = X'Q_1X, B_{XX} = X'Q_2X$ and $C_{XX} = X'Q_3X$ with $\phi_2^2 = \sigma_v^2/\lambda_2, \phi_3^2 = \sigma_v^2/\lambda_3$. Also, the Within estimator of β is $\tilde{\beta}_W = W_{XX}^{-1}W_{Xy}$, the Between individuals estimator of β is $\hat{\beta}_B = B_{XX}^{-1}B_{Xy}$ and the Between timeperiods estimator of β is $\hat{\beta}_C = C_{XX}^{-1}C_{Xy}$. This shows that $\hat{\beta}_{\text{GLS}}$ is a matrix-weighted average of $\tilde{\beta}_W, \hat{\beta}_B$ and $\hat{\beta}_C$. In fact,

$$\hat{\beta}_{\text{GLS}} = W_1\tilde{\beta}_W + W_2\hat{\beta}_B + W_3\hat{\beta}_C \tag{5.25}$$

(3.24)

where

$$\begin{aligned}W_1 &= \left[W_{XX} + \phi_2^2B_{XX} + \phi_3^2C_{XX}\right]^{-1}W_{XX} \\ W_2 &= \left[W_{XX} + \phi_2^2B_{XX} + \phi_3^2C_{XX}\right]^{-1}\left(\phi_2^2B_{XX}\right) \\ W_3 &= \left[W_{XX} + \phi_2^2B_{XX} + \phi_3^2C_{XX}\right]^{-1}\left(\phi_3^2C_{XX}\right)\end{aligned} \tag{5.26}$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = \sigma_\lambda^2 = 0, \phi_2^2 = \phi_3^2 = 1$ and $\hat{\beta}_{\text{GLS}}$ reduces to $\hat{\beta}_{\text{OLs}}$; (ii) as T and $N \to \infty, \phi_2^2$ and $\phi_3^2 \to 0$ and $\hat{\beta}_{\text{GLS}}$ tends to $\tilde{\beta}_W$; (iii) if $\phi_2^2 \to \infty$ with ϕ_3^2 finite, then $\hat{\beta}_{\text{GLS}}$ tends to $\hat{\beta}_B$; (iv) if $\phi_3^2 \to \infty$ with ϕ_2^2 finite, then $\hat{\beta}_{\text{GLS}}$ tends to $\hat{\beta}_C$. Wallace and Hussain (1969) compare $\hat{\beta}_{\text{GLS}}$ and $\tilde{\beta}_{\text{Within}}$ in the case of nonstochastic (repetitive) X and find that both are (i) asymptotically normal, (ii) consistent and unbiased and that

(iii) $\hat{\beta}_{\text{GLS}}$ has a smaller generalized variance (i.e. more efficient) in finite samples. In the case of non-stochastic (nonrepetitive) X they find that both $\hat{\beta}_{\text{GLS}}$ and $\tilde{\beta}_{\text{Within}}$ are consistent, asymptotically unbiased and have equivalent asymptotic variance-covariance matrices, as both N and $T \to \infty$. The last statement can be proved as follows: the limiting variance of the GLS estimator is

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \left(X'\Omega^{-1}X/NT\right)^{-1} = \frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \left[\sum_{i=1}^{3} \frac{1}{\lambda_i} \left(X'Q_iX/NT\right)\right]^{-1} \tag{5.27}$$

(3.25)

but the limit of the inverse is the inverse of the limit, and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{X'Q_iX}{NT} \quad \text{for } i = 1, 2, 3 \tag{5.28}$$

(3.26)

all exist and are positive semidefinite, since

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \left(X'X/NT\right)$$

is assumed finite and positive definite. Hence

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{\left(N\sigma_\lambda^2 + \sigma_v^2\right)} \left(\frac{X'Q_3X}{NT}\right) = 0$$

and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{\left(T\sigma_\mu^2 + \sigma_v^2\right)} \left(\frac{X'Q_2X}{NT}\right) = 0$$

Therefore the limiting variance of the GLS estimator becomes

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \sigma_v^2 \left(\frac{X'Q_1X}{NT}\right)^{-1}$$

which is the limiting variance of the Within estimator. One can extend Nerlove's (1971a) method for the one-way model, by estimating σ_μ^2 as $\sum_{i=1}^{N} \left(\hat{\mu}_i - \hat{\mu}\right)^2/(N-1)$ and σ_λ^2 as $\sum_{t=1}^{T} \left(\hat{\lambda}_t - \bar{\lambda}\right)^2/(T-1)$ where the $\hat{\mu}_i$ and $\hat{\lambda}_t$ are obtained as coefficients from the least squares dummy variables regression (LSDV). σ_v^2 is estimated from the Within residual sums of squares divided by NT. Baltagi (1995, appendix 3) develops two other methods of estimating the variance components. The first is Rao's (1970) minimum norm quadratic unbiased estimation (MINQUE) and the second is Henderson's method III as described by Fuller and Battese (1973). These methods require more notation and development and may be skipped in a brief course on this subject. Chapter 9 studies these estimation methods in the context of an unbalanced error component model.

Baltagi (1981a) performed a Monte Carlo study on a simple regression equation with twoway error component disturbances and studied the properties of the following estimators: OLS, the Within estimator and six feasible GLS estimators denoted by WALHUS, AMEMIYA, SWAR, MINQUE, FUBA and NERLOVE corresponding to the methods developed by Wallace and Hussain (1969), Amemiya (1971), Swamy and Arora (1972), Rao (1972), Fuller and Battese (1974) and Nerlove (1971a), respectively. The mean square error of these estimators was computed relative to that of true GLS, i.e. GLS knowing the true variance components. To review some of the properties of these estimators: OLS is unbiased, but asymptotically

inefficient, and its standard errors are biased; see Moulton (1986) for the extent of this bias in empirical applications. In contrast, the Within estimator is unbiased whether or not prior

information about the variance components is available. It is also asymptotically equivalent to the GLS estimator in case of weakly nonstochastic exogenous variables. Early in the literature, Wallace and Hussain (1969) recommended the Within estimator for the practical researcher, based on theoretical considerations but more importantly for its ease of computation. In Wallace and Hussain's (1969, p. 66) words the "covariance estimators come off with a surprisingly clear bill of health". True GLS is BLUE, but the variance components are usually not known and have to be estimated. All of the feasible GLS estimators considered are asymptotically efficient. In fact, Prucha (1984) showed that as long as the estimate of σ_v^2 is consistent, and the probability limits of the estimates σ_μ^2 and σ_λ^2 are finite, the corresponding feasible GLS estimator is asymptotically efficient. Also, Swamy and Arora (1972) proved the existence of a family of asymptotically efficient two-stage feasible GLS estimators of the regression coefficients. Therefore, based on asymptotics only, one cannot differentiate among these twostage GLS estimators. This leaves undecided the question of which estimator is the best to use. Some analytical results were obtained by Swamy (1971) and Swamy and Arora (1972). These studies derived the relative efficiencies of (i) SWAR with respect to OLS, (ii) SWAR with respect to Within and (iii) Within with respect to OLS. Then, for various values of N, T, the variance components, the Between groups, Between time-periods and Within groups sums of squares of the independent variable, they tabulated these relative efficiency values (see Swamy, 1971 , chapters II and III; Swamy and Arora, 1972, p. 272). Among their basic findings is the fact that, for small samples, SWAR is less efficient than OLS if σ_μ^2 and σ_λ^2 are small. Also, SWAR is less efficient than Within if σ_μ^2 and σ_λ^2 are large. The latter result is disconcerting, since Within, which uses only a part of the available data, is more efficient than SWAR, a feasible GLS estimator, which uses all of the available data.

5.4 Test of Hypotheses with Panel Data

5.5 REFERENCES

Chapter 6

Methods

We describe our methods in this chapter.

Les données de panel, ou données longitudinales possèdent les deux dimensions précédentes (individuelle et temporelle). En effet, il est souvent intéressant d'identifier l'effet associé à chaque individu (un effet qui ne varie pas dans le temps, mais qui varie d'un individu à un autre). Cet effet peut être fixe ou aléatoire.

Par conséquent, le modèle en données de panel s'écrit comme un modèle à double indice qui prend la forme suivante :

$$Y_{it} = \alpha_i \sum_k \beta_{ki} x_{ki} + \epsilon_{it}$$

avec

$$i : 1 \to N$$

et

$$t : 1 \to T$$

La double dimension qu'offrent les données de panel est un atout majeur. En effet, si les données en séries temporelles permettent d'étudier l'évolution des relations dans le temps, elles ne permettent pas de contrôler l'hétérogénéité entre les individus. A l'inverse, les données en coupes transversales permettent d'analyser l'hétérogénéité entre les individus mais elles ne peuvent pas tenir compte des comportements dynamiques, puisque la dimension temporelle est exclue du champ d'analyse.

Ainsi, en utilisant des données de panel, on pourra exploiter les deux sources de variation de l'information statistique : - Temporelle où variabilité intra-individuelle (within) - et individuelle ou variabilité inter-individuelle (Between).

Chapter 7

Analyses

Nous faisons *application* des méthodes présentées dans le chapitre précédant pour l'analyse des données de pannel

Avant de passefr à la modélisation, nous ferons une description de nos variables d'interet d'une manière statique : nos prédicteurs et la variables réponses

7.1 Netoyage de la base des données

Apperçue globale des données

Voici la structure de la base des données

```
## Rows: 3,310
## Columns: 6
## $ `N°`                   <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1~
## $ `Country/destination` <chr> "AFRIQUE DU SUD", "AFRIQUE DU SUD", "AFRIQUE DU ~
## $ Year                   <dbl> 2011, 2011, 2011, 2013, 2013, 2013, 2013, 2013, ~
## $ Goods                  <fct> "GRUES SUR PNEUMATIQUE", "CAMION FAMIL", "CAMION~
## $ Weight                 <dbl> 13500, 12000, 24000, 183, 19520, 19520, 19520, 1~
## $ Taxe                   <dbl> 0, 0, 0, 0, 264771, 272817, 283220, 264142, 0, 0~
```

Table 7.1: Echantillon de la base des données

N°	Country/destination	Year	Goods	Weight	Taxe
1	AFRIQUE DU SUD	2011	GRUES SUR PNEUMATIQUE	13500	0
2	AFRIQUE DU SUD	2011	CAMION FAMIL	12000	0
3	AFRIQUE DU SUD	2011	CAMION SOMUL	24000	0
4	AFRIQUE DU SUD	2013	Café vert arabica k4	183	0
5	AFRIQUE DU SUD	2013	Café vert arabica k4	19520	264771
6	AFRIQUE DU SUD	2013	Café vert arabica k4	19520	272817
7	AFRIQUE DU SUD	2013	Café vert arabica k4	19520	283220
8	AFRIQUE DU SUD	2013	Café vert arabica k4	19520	264142
9	AFRIQUE DU SUD	2013	CAMION	24000	0
10	AFRIQUE DU SUD	2017	Instruments et appareils du n°90.15	654	0

Voici les modalités de la variabme `Goods` qui signifie `Marchandises`

Table 7.2: Modalités de la variable Goods à l'importation des donnees

x
0
3Café vert arabica, en feve K3
Abats comestibles,congeles,de chevaux,anes,mulets,ovins ou caprins
ABATS COMESTIBLES;CONGELES;DE CHEVAUX;ANES;MULETS;OVINS
Accessoires de radio diffusion
Accessoires de vehicules
Accumulateurs electriques
Acide acetique
ages de 5 ans ou moins
ages de plus de 5 ans
Agés de plus de 5 ans ou moins
Alcaloides du quinquina et leurs derives;
ALCOOL ETYLIQUE NON DENATURE
ambulance d'une cylindree excedant 2500 cm3
Antennes
Antennes et reflecteurs d'antennes
antennes et refleteurs
Appareils d'eclairage electriques
Appareils d'eclairage non electriques
Appareils d'eclairages electriques
Appareils du n°84.14
Appareils electrothermiques pour la
appareils pour la reception,la conversion et la transmission
Art et materiel d'athletisme
Articles confectionnes en textiles
Articles d'economie domestique,en
Articles de bureau
ARTICLES DE BUREAU
Articles de bureau ou de la papeterie
Articles de friperie
ARTICLES DE FRIPERIE
Articles et materiel d'athletisme
Ashok Layland
ASPIRATEUR ET ACCESSOIRES
Autes bois sciés
AUTRE MACHINE ET APPAREIL A IMPRIMER
AUTRE MINERAIS DE TITANE (Coltant)
AUTRE PARTIE DE PLANTE
AUTRE PEAUX
AUTRE PREP ALIMENTAIRE
Autre vehicules automobiles a usages speciaux
Autres

AUTRES
Autres bois scies
Autres abats comestibles frais ou refrigérés de chevaux,anes,mulets,ovins,caprins

Autres abats comestibles,congeles,de chevaux,anes,mulets,ovins ou caprins
Autres accessoires de tuyauterie en fonte
Autres accumulateurs electriques
Autres appareils elevateurs, a action continue pour marchandises
Autres armes

Autres art de bureau ou de papeterie en papier
Autres Art de menage
Autres articles d'economie domestique, en
autres articles de bureau
AUTRES ARTICLES DE BUREAU OU D PAPETERIE EN PAPIER OU CARTON

Autres articles de bureau ou de papeterie en
Autres articles de campement
Autres articles de friperie
Autres articles de menage ou d'économie en aluminium
Autres articles de transport ou d'emballage

Autres articles en toles emailles
Autres babeurre,lait et creme
Autres babeurre,lait et creme caillés,yoghourt
autres baies fraiches
Autres bois

Autres bois plaques ou stratifies
AUTRES BOIS ROPICAUX SCIES LONG
Autres bois sciées
Autres bois scies
AUTRES BOIS SCIES

autres bois sciés
Autres bois sciés
Autres bois tropicaux, a scies long, sciages d'1 avives d'1 epaisseur superieure a 50mm
Autres bois tropicaux, scies lon, sciages avives d'1 epaisseur superieure a 50mm
Autres bois tropicaux,scies long

AUTRES BOIS TROPICAUX,SCIES LONG-SCIAGES AVIVESD'1 EPAISSEEUR SUPERIEURE A 50MM
Autres bois tropicaux,scies long,sciages avivées
Autres bois tropicaux,scies long,sciages avives
Autres bois tropicaux,sciés long…sciages avives
Autres bois tropicauxmscies long

Autres Caoutchouc
Autres cassiterites
Autres chariots y.c. chariots tracteurs,sans
Autres chaussures
Autres coltan

Autres coltans
Autres constructions
Autres couvertures non chauffantes
Autres dechets et debris
Autres dechets et debris d'aciers alliés

Autres déchets et débris d'aciers alliés
Autres dechets et debris d aciers
autres eaux, y compris l'eau douce
Autres engins du n84 29
AUTRES EQUIPEMENTS DE PROTECTION

Autres et debris de bois
Autres etiquettes de tous genres, en papier
Autres fours industriels ou de laboratoires
Autres fours, cuisinieres, rechauds, grils
Autres fours,cuisinières,rechauds

AUTRES GROUPES ELECT
Autres groupes electrogenes
Autres haut-parleurs
Autres huiles de pal;iste ou babassu
Autres instrument et appareils du n°90,18

Autres instruments de musique a vent
Autres instruments et appareils
Autres instruments et appareils du n 90.18
Autres jus de tout autres fruit ou legume
Autres legumes ¿ cosse secs, ¿coss¿s, m¿me

Autres legumes, frais ou refrigeres
Autres lunettes
autres machines
Autres machines
Autres machines electriques a laver la

Autres machines et appareils
Autres machines et appareils de manutention du n,8428
Autres machines et appareils du n 8515
Autres medicaments
AUTRES MEDICAMENTS

Autres mélanges d'épices
Autres meubles et leurs parties
AUTRES MINERAIS DE TANTALE
AUTRES MINERALS DE TANTALE (Coltan)
Autres montres (y.c. compteur de temps)

Autres moteurs diésel
Autres moteurs diésels
Autres nattes en matières vegetales
autres outils
Autres outils agricoles

Autres ouvrages en plomb
Autres papiers non denommes ni compris ailleurs
AUTRES PAPIERS NON DENOMMES NI COMPRIS AILLEURS
Autres papiers non denommés ni compris ailleurs
Autres parties d'avions ou d'helicopteres

Autres parties de lampes portatives
AUTRES PARTIES DE PLANTES EN MEDECINE
AUTRES PARTIES DE PLANTES UTILISEES PRINCIPALEMENT EN MEDECINE

Autres parties de plantes utilisés en medecine
Autres parties des appareils des n 88.01 ou

Autres parties des plantes utilisées principalement en medecine
Autres parties et accesoires des vehicules
Autres parties et accessoires de
AUTRES PARTIES ET ACCESSOIRES DE VEHICULES DES n87.01 a 87.05
AUTRES PARTIES ET ACCESSOIRES DES VEHICULES

autres parties et accessoires des vehicules des 8701 a 8705
Autres parties et accessoires des vehicules des 8701 a 8705
AUTRES PARTIES ET ACCESSOIRES DES VEHICULES DES N 87.01 A 87.05
Autres parties et accessoires des vehicules des n 8701 à 8705
Autres parties et accessoires des vehicules n 8701 à 8705

AUTRES PEAUX
AUTRES PEAUX DES BOVINS
Autres pierres
autres piles et batteries
Autres plantes

Autres plantes, parties de plantes,
AUTRES PLAQUES
Autres plaques, feuilles,bandes en polymeres
Autres pneumat autres qu'a crpn,a chvn en autr ctc synt
Autres pneumatiques neuf en autres mat que les mat synth de type pour voiture tourisme

AUTRES POMPES A DISPOSITIF MESUREUR OU CONCUES POUR EN COMPORTER UN
Autres préparations alimentaires
Autres preparations alimentaires n d c a
Autres preparations homogenisees
AUTRES PRODUITS DE BEAUTE

Autres produits de beaute,de
Autres produits du no 33.07,contenant de
Autres recepteurs fixes de radiodiffusion
AUTRES RECERVOIRS
Autres refrigerateurs menagers

Autres sacs,sachets,pochettes
Autres salmonides entiers,
Autres savons
AUTRES TENTES
Autres tissus de coton imprimé,200 g/m2 et

Autres tracteurs
Autres unites de machines automatiques de
Autres vehicules autom a usages speciaux
Autres vehicules automobiles a usages
Autres vehicules automobiles a usages speciaux

Autres vetements de dessous en coton, pour
Autres vetements en bonneterie d'autres
Autres, en couleurs
Babeurre,lait et creme cailles,yoghourt ou avec fruits ou cacao
BACHES

Baches et stores d'exterieur de fibres

Baches et stores d'exterieur de fibres synth
Baies fraiches
bières de malt titrant moins de 6°
BISCUIT

Biscuits additionnes d'edulcorants
BITUMES
Bois ciés
BOIS RABOTES OU PONCES
Bois scies

BOIS SCIES
Bois sciés
Bois scies longitudinalement
Bois topicaux, scie long..sciages avives d'1 epaisseur supérieur à 50mm
Bois tropicaux

Bois tropicaux scies
Bois tropicaux scies long,sciages avives d'un epaisseur superieur à 50 mm
Bois tropicaux,scies
Bois tropicaux,scies long
Bois tropicaux,scies long,sciages avives

Bois tropicaux,scies long,sciages vives d'un epaisseur supérieure à 50 mm
Boites et cartonnages,pliants,en papier ou carton non ondulé
Boites et caisses en papier ou carton ondule
Boites et caisses en papier ou carton ondulé
Boites et cartonnaages,pliants en papier ou carton non ondule

Boites et cartonnages pliants en papier ou carton non ondule
Boites et cartonnages, en papier ou carton non ondule
Boites et cartonnages, pliants, en papier ou
Boites et cartonnages, pliants, en papier ou caton non ondule
Boites et cartonnages,pliants

Boites et cartonnages,pliants , en papier ou carton
Boites et cartonnages,pliants en papier ou carton non ondule
BOITES ET CARTONNAGES,PLIANTS EN PAPIER OU CARTON NON ONDULE
Boites et cartonnages,pliants,en papier
Boites et cartonnages,pliants,en papier ou carton non ondule

Boites et cartonnages,pliants,en papier ou carton non ondulé
Boites et cartonnages,pliants,en papier ou carton non ondules
Boites et cartonnagesmpliantsmen papier ou carton non ondule
Boites et cartonnagres
Boites et catonnages, pliants, en papier ou carton non ondule

Boites, caisses et similaires en matieres plastiques
Boites,caisses,casiers et similaires en matières plastiques
Boites,caisses,casiers et similaires plastiques
Brosses a dents y compris les brosses a
Cables coaxiaux

Cadres et conteneurs
Cadres et conteneurs multimodaux y.c.conteneurs citernes et rervoirs
CAF Vert arabica, en feves K4, Est
Caf? vert arabica, en feves K4, Est

Caf¿ vert arabica, en feves K3, Est

CAFE
Cafe arabica
CAFE ARABICA
Café arabica
Café Arabica

Café arabica k4
Café arabica k7
Café arabica, en feve K4
Café arabica,en feves k3,est
Cafe arabica,en feves k4

Café arabica,en feves k4,est
Cafe arabica,en feves k7
Café cert, en feves k4 (kivu4)
Cafe non torrefie, non decaféiné
Café non torrifié

Café ver arabica k4
CAFE VERT
Café vert
Café vert arabica
CAFE VERT AFRICA;EN FEVES K4

CAFE VERT ARAB K4
Cafe vert arabica
CAFE VERT ARABICA
Café vert arabica
Café vert Arabica

CAFE VERT ARABICA ;EN FEVES
Cafe vert arabica en feves k4
café vert arabica en feve k4 Est
Cafe vert arabica en feves
café vert arabica en feves

Cafe vert arabica en fèves
Café vert arabica en fèves
CAFE VERT ARABICA EN FEVES DE K4
Cafe vert arabica en feves k3
CAFE vert arabica en feves k3

Café vert arabica en feves k3
Café vert arabica en fèves K3
café vert arabica en feves k3 est
Cafe vert arabica en fèves K3(kivu)
cafe vert arabica en feves k4

Cafe vert arabica en feves k4
CAFE VERT ARABICA EN FEVES K4
Café vert arabica en feves k4
Café vert arabica en fèves K4
Cafe vert arabica en feves k4 est

café vert arabica en feves k4 est
Cafe vert arabica en fèves K4(kivu4)

Cafe vert arabica en fèves k4,Est
Cafe vert arabica en feves k7
Café vert arabica en feves,k3

Café vert arabica en feves,k4
Cafe vert arabica en fevesk4
Cafe vert arabica en fevres
Cafe vert arabica feves k4
café vert arabica k

Café vert arabica k'
CAFE VERT ARABICA K3
café vert arabica k3
Café vert arabica k3
Cafe vert arabica k4

CAFE VERT ARABICA k4
CAFE VERT ARABICA K4
café vert arabica k4
Café vert arabica k4
Café vert Arabica K4

Cafe vert arabica k7
CAFE VERT ARABICA K7
café vert arabica k7
Café vert arabica k7
Café vert arabica k8

Café vert arabica, en faves k4
Cafe vert arabica, en faveurs k4
Cafe vert arabica, en feve K3
Café vert arabica, en feve K3
Café vert arabica, en feve k3, est

Cafe vert arabica, en feve k4
café vert arabica, en feve K4
Café vert arabica, en feve k4
Café vert arabica, en feve K4
Cafe vert arabica, en feves k3

Café vert arabica, en feves k3
Café vert arabica, en feves K3 (KIVU 3)
Cafe vert arabica, en feves k3 (kivu 3)
Cafe vert arabica, en feves k3 (kivu3)
café vert arabica, en feves k3, Est

Café vert arabica, en feves k3, Est
Café vert arabica, en fèves K3, EST
Cafe vert arabica, en feves k4
Café vert arabica, en feves k4
Cafe vert arabica, en feves k4 (kivu 4)

Café vert arabica, en feves k4 (kivu 4)
Café vert arabica, en feves K4 (KIVU 4)
Cafe vert arabica, en feves k4 (kivu4)
Café vert arabica, en feves K4 (KIVU4)
Café vert arabica, en feves k4 est

Cafe vert arabica, en feves K4, Est
Café vert arabica, en feves k4, Est
Café vert arabica, en feves K4, Est
Café vert arabica, en fèves K4, EST
Cafe vert arabica, en feves k5 (kivu 5)

Cafe vert arabica, en feves k6 (kivu 6)
Cafe vert arabica, en feves k7 (kivu 7)
Café vert arabica, en feves K7 (KIVU7)
Café vert arabica, en feves k7, Est
Café vert arabica,en feve k4 est

Café vert arabica,en feve k7 est
Café vert arabica,en feves
café vert arabica,en feves 4,est
cafe vert arabica,en feves k3
Cafe vert arabica,en feves k3

Cafe vert arabica,en feves K3
café vert arabica,en feves k3
Café vert arabica,en feves k3
café vert arabica,en fèves k3
CAFE VERT ARABICA,EN FEVES K3(kivu)

Café vert arabica,en feves K3, Est
café vert arabica,en feves k3,Est
Café vert arabica,en feves k3,Est
Café vert arabica,en feves K3,Est
cafe vert arabica,en feves k4

Cafe vert arabica,en feves k4
Café vert arabica,en feves k4
café vert arabica,en fèves k4
café vert arabica,en feves k4 est
Café vert arabica,en feves k4 est

Café vert arabica,en feves K4 EST
CAFE VERT ARABICA,EN FEVES K4(kivi)
CAFE VERT ARABICA,EN FEVES K4(kivu4)
Café vert arabica,en feves k4(kivu4)
Cafe vert arabica,en feves k4,est

café vert arabica,en feves k4,Est
café vert arabica,en feves K4,est
Café vert arabica,en feves k4,est
Café vert arabica,en feves K4,Est
cafe vert arabica,en feves k5

Cafe vert arabica,en feves k7
Café vert arabica,en feves k7,est
Café vert arabica,en feves k7,Est
cafe vert arabira,en feves k3
CAFE VERT EN FEVE

Cafe vert en fèves k4
Cafe vert en feves,k3
CAFE VERT K4

CAFE VERTS
Caféver arabica,en feves K4,Est

cameras
CAMION
Camion Ben d'occasion
Camion Citerne d'occasion
CAMION FAMIL

Camion renault kerax
CAMION SOMUL
CAOUTCHOU
CAOUTCHOUC NATUREL
CAOUTHOUX NATUREL

Cartons vide
Cartons vides
CARTONS VIDES USAGES POUR RE-EMPLOIES
CARTONS VIDES USAGES POUR RE-EMPLOIS
casseroles en aluminium

Casseterite
Cassitérie
Cassiteriet oxyde d'etain, 60-64% de Sn
cassiterite
Cassiterite

CASSITERITE
Cassitérite
CASSITERITE DE 55-65 % DE SNO2
Cassiterite (oxyde d'etain) de 55-65% de
Cassiterite de 55-65%

Cassiterite de 55% - 65 %
Cassiterite oxyde
Cassiterite oxyde d'etain
Cassiterite oxyde d'etain 55-65%
Cassiterite oxyde d'étain de 55-59% de sn

Cassitérite oxyde d'étain de 55-59% de Sn
Cassiterite oxyde d'etain de 65-69
Cassiterite oxyde d'etain de 66-69% de Sn
Cassitérite oxyde d étain de 55-59% de sn
Cassiterite oxyde d étain de 65-69% de sn

cassiterites
Cassiterites
Cassitérites oxyde d'etain de 55-59% de sn
Cassiterites oxyde d'etain de 65-69% de Sn
Cassiterites oxyde d etain de 55-59

CASSITERTE
CASSITTERITTES
CEFE VERT ARABICA
Cerceuil
Cereales autre que semences

Cereales autres que semences

Chargeurs autopropulses a chargement frontal
Chars et automobiles blindees de combat et
Chassis d'autres vehicules automobiles des n 8701 à 8705 avec moteur
CHAUDIERE ET SES ACCESSOIRES

Chaudières
Chocolat et préparations alimentaires du cacao
Cigares(meme a bouts coupes) et
CIMENT
CITERNES EN ACIER

coltan
Coltan
COLTAN
Coltan-scorie stanique 20-25% Ta2O5
Coltan & scorie stanique (concentres de

Coltan et scorie stanique de 26-30%
COLTANS
COLTN
Compacteurs autopropulses
COMPACTEURS, NIVELEUSES

compresseurs
Conditionneur d'air de mur ...puis. < ou =
Conserves de sprats et esprots entiers ou en morceaux non haches
construction prefabriquees
Constructions et parties de constructions en aluminium autre que 761010

Constructions prefabriquées
Constructions prefabriques
CONTAINER
CUISINIERES
Dechets et brisures de cafe vert arabica

DECHETS ET BRISURES DE CAFE VERT ROBUSTA
Dechets et brisures de café vert robusta
Dechets et debris d'aciers allies
Dechets et debris de bois
Dechets et debris de fer ou d'acier

Dechets et debris de zirconium et ouvrages
Demareurs,meme fonctionnant comme generatrice
Desinfectant
Désinfectants
Dispositifs de fermeture

Eaux conditionnées pour la table
Eaux de vie de vin ou de marc de raisin
EAUX MINERALES
EAUX MINERALES ET EAUX GAZEIFIEES
Eaux minerales et eaux gazeifiées

ECAUSSINE ET AUTRES PIERRES CALCAIRES DE TAILLE OU CONSTRUCTION,ALBATRE
ECHANTILLONS DES MATERIAUX GEOLOGIQUES
ECORCE QUINQUINA
ECORCE QUINQUINNA

ECORCES DE QUINQUINA

ECORCES QUINKINA
ECORCES QUINQUINA
EFFETS PERSONNELS
Engins du n°84.29
Engrais du chapitre31 en tablettes ou emballages n'excedant pas 10 kg brut

equipement concus pour la protection individuelles
Equipements de sport
ESCALIERS MECANIQUES ET TROTTOIRS ROULANTS
Etuis,ecrins et similaires à surface exterieure en plastique ou textile
Extincteurs meme charges

Farine de froment (ble) ou de meteil
Farine de froment ou de meteil
FARINE DE MAIS
Farine de moutarde et moutarde préparée
Farine de moutarde preparée

FARINES DE MAIS
FILS D'ACETATE
FILS D'ACETATE DE CELLULOSE
FOURS,CUISINIERES,RECHAUDS,GRILS ET ROTISSOIRES ELECTRIQUES
Goupes electrogene (diesel, semi- disiel)

Groupes electrogene
Groupes electrogene(diesel,semi-diesel)de puissance n'excedant pas 75kva
Groupes electrogenes(diesel,semi-diesel) de
GRUES SUR PNEUMATIQUE
HARICOTS

Haricots communs:de semences
HUILE DE GRAISSAGE
Huile de palme brute
HUILES VEGETALES
Huiles,graissages et fractions,non chimiquement modifiées

Insecticides
Instruments et appareils de photographie
Instruments et appareils du n°90.15
IVECO TRAKKER D OCCAS
LAIT EN POUDRE

Lampes -reclames
LAND CRUISER
LAND CRUISER PRADO
LAND ROVER
Land rover defender 90

Levures vivantes
limes,rapes et outils similaires
Liqueurs titrant moins de 25?
MACHINE A MELANGER
MACHINE A RABOTER ET ACCESSOIRES

MACHINE A SCIER ET ACCESSOIRES
Machine et appareil pour la fabrication du tabac

MACHINE POUR PREPARATION DU TABAC
MACHINE POUR PREPARER DU TABAC
MACHINE POUR TRANS DE TABAC

Machines à meuler,poncer ou polir les matières du n°84.65
MACHINES A PERCER OU MORTAISER LES MATIRES DU N° 84.65
Machines et appareil
Machines et appareils du n° 84.30
MACHINES ET APPAREILS POUR LA PREP OU TRANSFORMATION DU TABAC

Machines et appareils pour le brochage ou la
Machines geneatrices à courant alternatif de 75 kva
Machines generatrices a courant alternatif
Machines pr la transformation du tabac
Machines qui assurent au moins deux des fonctions suivantes

Machines qui assurent au moins deux des fonctions suivantes impression,copie ou transm
Maghony,scies longitudinalement,sciages avives
Mahagany, scies lingutudinalement, sciage avices d'une epaisseur superieur a 50mm
Mahagany, scies longitudinalement, sciage avives d'une epaisseur inferieur à 50mm
Mahogaany, scies lingitudinalement, sciagaes avives d'une epaisseur superieur a 50mm

Mahogany, scies lingitudianlement, sciages avives d'une epaisseur superieure a 50mm
Mahogany,scies lingitudinalement,sciages avives
Mahogany,scies lingitudinalement,sciages avives d une epaisseur superieure a 50 mm
Mahogany,scies longitudinalement
Mahogany,sciés longitudinalement

Mahogany,scies longitudinalement,sciages avives
Mahogany,scies longitudinalement,sciages avives d'une epaisseur
Mahogany,scies longitudinalement,sciages avives d'une épaisseur inf à 50 mm
Mahogany,scies longitudinalement,sciages avives d'une epaisseur inférieure à 50 mm
Mahogany,scies longitudinalement,sciages avives d une epaisseur inferieure à 50 mm

Malles,valises,mallettes à surface exter en plastique
Margarine, a l'exclusion de la margarine
Margarine,a l exclusion de la margarine liquide
marteaux et masses
Matériel d'échafaudage,de coffrage ou d'étayage

Materiel d'echafaudage,de coffrage ou d'etayge en fonte,fer ou acier
MATERIEL POUR LA GYMNASTIQUE
Materiels d'echafaudage,de coffrage ou d'etayage en fonte,fer ou acier
MATIERES COLORANTES
Matières colorantes

MEDICAMENTS
Melanges d'epices
MEUBLES ET LEURS PARTIES
MIERAIS DE COLTAN ET SCORIE STANIQUE DE 26-30%TA05
Minerais autres que ceux des n 26171000 a 26179091

MINERAIS CASSITERITES
Minerais d'etain et leurs concentré
Minerais d'etain et leurs concentré d'une teneur de 55 à 65 % en etain
Minerais d'etain et leurs concentres
minerais d'etain et leurs concentres d'une teneur de 55 à 65

Minerais d'étain et leurs concentrés d une teneur de 55 à 65% en etain/cassiterite
Minerais d étain et leurs concentrés d une teneur de 55 à 65% en étain
Minerais d étain et leurs concentrés dune teneur de 55% à 65%
MINERAIS DE COLTA § SCORIE STANIQUE DE + DE 35
Minerais de coltan

MINERAIS DE COLTAN
Minérais de coltan
Minerais de coltan & scorie de 26-30 % Tao5
Minerais de coltan & scorie stanique
Minerais de coltan & scorie stanique de

Minerais de Coltan & scorie stanique de
Minerais de coltan & scorie stanique de 26-30 % Ta05
MINERAIS DE COLTAN & SCORIE STANIQUE DE 26-30%
Minerais de coltan & scorie stanique de 26-30% Ta 05
Minerais de coltan & scorie stanique de 26-30% Ta05

Minerais de coltan & scorie stanique de 26-30% Tao5
Minerais de coltan & scorie stanique de 26-30% TaO5
Minerais de Coltan & scorie stanique de 26-30% TaO5
Minerais de coltan & scorie stanique de 26-30%Ta05
Minerais de coltan & scorie stanique de 4-14% Ta05

Minerais de coltan & scories stanique de 26-30%
MINERAIS DE COLTAN &SCORIE STANIQUE DE 26-30 %Ta05
MINERAIS DE COLTAN &SCORIE STANIQUE DE 26-30%
Minerais de coltan &scorie stanique de 26-30%TaO5
Minerais de coltan et scorie

Minerais de coltan et scorie stanique
MINERAIS DE COLTAN ET SCORIE STANIQUE
Minerais de coltan et scorie stanique de 26-30%
Minerais de coltan et scorie stanique de 26-30% Ta05
Minerais de coltan et scorie stanique de 26-30%Ta05

Minerais de coltan et scorie stanique de 4-14% ta05
Minérais de coltan et scories
MINERAIS DE COLTAN&SCORIE STANIQUE
Minerais de coltant & scorie stanique de 26-30% Ta05
MINERAIS DE TANTALE

Minérais de tantale
Minerais de tantale et leurs concentres
Minerais de tantale teneur 26-30%
Minerais de tantale teneur 26-30% tantale et 40-59% oxyde niobium ou colombite
MINERAIS DE TANTALE TENEUR 26-30% TANTALE ET 40-59% OXYDE NIOBIUM OUCOLOM

Minerais de tantale teneur 26-30%tentale et 40-59% oxyde niobium ou colombite
Minerais de tantale teneure 26-30% et 40-59% oxyde niobium ou colombite
Minerais et leurs concentres
Minerais et leurs concentrés d une teneur de 55 à 65% en etain/cassiterite
Mineraus de coltant & scorie stanique de 26-30% Ta05

Mitraille de fer
Mitrailles de fer
Mitralles de fer

Mohogany, scies longitudinalement, sciages avives d'une epaisseur inferieur à 50mm
Moteur Aviation

Moteurs à explosion pour vehicule du chapitre 87
Moteurs universels de puissance excedant 37,5w
MOTOCYCLES
Motos et lubrifiant
Navets,betteraves,celeris-raves et similaires frais refrigeres

NISSAN VANETTE
Niveleuses autopropulsées
Non decafeine
OBJET DE DEMEN
OBJET DE DEMENAGEMENT

Œufs
Or a usages non monetaires d'exploitation
OR ARTISANAL
Oxygene
Palettes simples, palettes-caisses et autres

PAPIER A CIGARETTE EN ROULEAUX D'UNE LARGEUR N'EXCEDANT PAS 5 CM
Papier a cigarettes en rouleaux d'une
Papiers ndnca
Papiers non denommes ni compris ailleurs
Parfuns et eaux de toillette titrant - de 50

Parties de machines
PARTIES DES MACHINES ET APPAREILS DU N 84.74
Parties des machines et appareils du n° 84.53
PARTIES ET ACC VEHICULE
Parties et accessoires

PARTIES ET ACCESSOIRES DES MACHINES
Parties et accessoires des vehicules des 87.01 à 87,05
Parties et accessoires des vehicules des n° 87.11 à 87.13
Parties et accessoires des vehicules des n°87.01 à 87.05
PARTIES RECONNAISSABLES DES MACHINES DES N°85.01 OU 85.02

PC
PEAUX
PEAUX DE BETTES
Peaux de bovins
PEAUX DES BEBES

Peaux des betes
PEAUX DES BETTES
Petrol lampant sans biodiesel
photocopie
Pieres synthétiques ou reconstituees

Pieres synthetiques ou reconstituées
Pierres de couleur
Pierres de taille ou construction
Pierres synth ou reconstituées
PIOCHES,PICS,HOUES,BINETTES,RATEAUX ET RACLOIRS

Plantes médecinales

Plantes medicales
Plaques,feuilles,bandes,rubans,pellicules
Pneumatiques
Pneumatiques rechapes pour auto-bus ou camion

Poivre non broye ni pulverise
Pompe à vide
pompes a carburant
Pompes à injection pour moteurs à explosion
PREPARATIONS ALIMENTAIRES NDCA

Préparations alimentaires ndca
Preparations pour enfants,conditionnees pour
Préparations soufflées ou grillées a base de cereales
Préparations,pour sauces,sauces,condiments et assaisonnements
PRODUIT DE BEAUTE

Produits de beaute
Produits de beaute,de maquillage,solaires ou pour la peau
propulseurs a reaction autres que les turboreacteurs
quinquina
Quinquina

QUINQUINA
Recipients pour gaz comprimes,liquefiés,en fonte,fer ou acier
REFRIGERATEUR
Reservoirs, futs, tambours, bidons, d'une
Reservoirs,foudres,cuves et autres(+300)en matières plastiques

Riz semi-blanchi ou blanchi, meme poli ou
Sacs et sachets d'e mballage en autres synthétiques ou articles
Sacs et sachets d'emballage en autres sythetiques ou artificiels
Sacs et sachets en bonneterie de jute ou autres fibres
Sacs,sachets,pochettes,cornets en polyethylene

Salmonides entiers,congeles
SCORIE
Scorpio 4x4 Mihindra
Sel
SEL IODE

Slips et caleçons pour hommes ou garçonnets en bonneterie de coton
Sucres de canne
Sucs et extraits végétaux de houblon
Tabac partiellement ecotes
TABACS

TABACS ECOTES
TABACS PARTIELLEMENT ECOTES
The noir (fermente...), en emballage
TILAPIAS
Tissus Imprimés

Tours et pylones en fontes
Toyota ambulance
TOYOTA DINA Agés de plus de 5 ans
Toyota Hilux

Toyota hilux surf

TOYOTA ID
Toyota Land cruiser
TOYOTA LAND CRUISER
Toyota pick up
Toyota Prado tx

TOYOTA RAV4 D'OCCASION
Treuils cabestans
Treuils cabestants
TROTTINETTES(CHUKUDU)
Tungstene

vaisselle et art de tables et de cuisine en mat plastiques
Vaisselle et art de tables et de cuisine en matières plast
Véhicule
VEHICULE AUTO
Vehicule automobile

VEHICULE KIA SORENTO
VEHICULE LAND ROVER
Vehicules automobiles a usages speciaux
Vetements neuf
vetements usage

Vis et boulons filetes en fonte,fer ou acier
WOLFRAM
WOLFRAMI
Wolframite
WOLFRAMITE

La variables **Goods** a 740 modalités

Faison la cactérisation des niveaux des marchandises dont l'encodage fait défaut

```
class(taxe_df$Goods)
```

```
## [1] "character"
```

Usage de tm et Stringr

```
## Warning: Unknown levels in `f`: equipements protection
```

```
##  [1] "Autres Marchandises"
##  [2] "Bois"
##  [3] "Machines et appareils domestique"
##  [4] "Médicaments et plantes médécinales"
##  [5] "Poissaons, viande et oeufs"
##  [6] "Matériels de construction"
##  [7] "Materiel Informatique et Electroniques"
##  [8] "Véhicules,camions,Motos et acc"
##  [9] "Vetements,tissus et acc et chaussure"
## [10] "boissons, bières et limonades"
## [11] "Machine us Ingsutriel"
## [12] "Article Menange et Campement"
## [13] "sacs, sachetsn emballages"
```

```
## [14] "Papiers et fournitures de bureaux"
## [15] "Produits alimentaires,prep et huiles"
## [16] "caféarabica"
## [17] "Minérais et dérivés"
## [18] "engins et tracteurs"
## [19] "Cigarette et papier cigarettes et tabac"
## [20] "constructionprefabriquees"
## [21] "cadreset conteneurs"
## [22] "Pièces de Réchange appareils"
## [23] "Générateurs,baterie et piles"
## [24] "etuis en plastique ou textile"
## [25] "Pétrole et dérivées et huile de graissage"
## [26] "boissons, bières,liqueurs et limonades"
## [27] "produits beaute"
## [28] "peauxdes betes"
```

```
##                                          n    % val%
## Autres Marchandises                    160  4.8  4.8
## Bois                                   140  4.2  4.2
## Machines et appareils domestique        30  0.9  0.9
## Médicaments et plantes médécinales      34  1.0  1.0
## Poissaons, viande et oeufs              12  0.4  0.4
## Matériels de construction               27  0.8  0.8
## Materiel Informatique et Electroniques  32  1.0  1.0
## Véhicules,camions,Motos et acc         113  3.4  3.4
## Vetements,tissus et acc et chaussure   100  3.0  3.0
## boissons, bières et limonades           15  0.5  0.5
## Machine us Ingsutriel                   54  1.6  1.6
## Article Menange et Campement            37  1.1  1.1
## sacs, sachetsn emballages                6  0.2  0.2
## Papiers et fournitures de bureaux       24  0.7  0.7
## Produits alimentaires,prep et huiles    68  2.1  2.1
## caféarabica                           1303 39.4 39.5
## Minérais et dérivés                   1053 31.8 31.9
## engins et tracteurs                     18  0.5  0.5
## Cigarette et papier cigarettes et tabac 14  0.4  0.4
## constructionprefabriquees                7  0.2  0.2
## cadreset conteneurs                      1  0.0  0.0
## Pièces de Réchange appareils             6  0.2  0.2
## Générateurs,baterie et piles            15  0.5  0.5
## etuis en plastique ou textile            1  0.0  0.0
## Pétrole et dérivées et huile de graissage 4  0.1  0.1
## boissons, bières,liqueurs et limonades   2  0.1  0.1
## produits beaute                         10  0.3  0.3
## peauxdes betes                          14  0.4  0.4
## NA                                      10  0.3   NA
```

netoyyage de la variable country_desti qui est un facteur dans le quel nous retrouvons les niveaux rédondants (sur l'identifiant des pays)

```
##  [1] "AFRIQUE DU SUD"
##  [2] "ALGERIE"
##  [3] "ALLEMAGNE"
```

```
##  [4] "AMERIQUE LATINE"
##  [5] "ANGLETERRE"
##  [6] "ANGOLA"
##  [7] "ARABIE"
##  [8] "ASIE"
##  [9] "AUSTRALIE"
## [10] "BELGIQUE"
## [11] "BURUNDI"
## [12] "CANADA"
## [13] "CHINE"
## [14] "CHYPRE"
## [15] "CONGO BRAZA"
## [16] "CZECH REP"
## [17] "DOMBASI SIMBA"
## [18] "EMIRATES ARABES UNIES"
## [19] "ESPAGNE"
## [20] "FRANCE"
## [21] "GABON"
## [22] "GRANDE BRATAGNE"
## [23] "GRECE"
## [24] "HONG KONG"
## [25] "ILE MAURICE"
## [26] "INDE"
## [27] "ITALIE"
## [28] "J WOLFF"
## [29] "JAPON"
## [30] "KENYA"
## [31] "KP - Corée, République Populaire démocra"
## [32] "LIBAN"
## [33] "LUXEMBOURG"
## [34] "MADRID"
## [35] "MALAISIE"
## [36] "MAROC"
## [37] "NERLAND"
## [38] "NERTHERLAND"
## [39] "NIGERIA"
## [40] "NOUVELLE ZELANDE"
## [41] "OUGANDA"
## [42] "PANAMA"
## [43] "PAYS BAS"
## [44] "PHILLIPINE"
## [45] "POLOGNE"
## [46] "PORTUGAL"
## [47] "R-U"
## [48] "RDC"
## [49] "RDC/BELGIQUE"
## [50] "RDC/BUNIA"
## [51] "RDC/CHINE"
## [52] "RDC/ETATS UNIS"
## [53] "RDC/FRANCE"
## [54] "RDC/MALAISIE"
## [55] "RDC/OUGANDA"
```

```
## [56] "RDC/R-U"
## [57] "RDC/RWANDA"
## [58] "RDC/SINGAPOUR"
## [59] "RDC/SUISSE"
## [60] "REP TCHEQUE"
## [61] "ROYAUME UNI"
## [62] "RWANDA"
## [63] "SENEGAL"
## [64] "SINGAPOUR"
## [65] "SKN"
## [66] "SOMALIE"
## [67] "SOUDAN"
## [68] "SUCAFINA"
## [69] "SUD SOUDAN"
## [70] "SUEDE"
## [71] "SUISSE"
## [72] "SUITZERLAND"
## [73] "Swaziland"
## [74] "SWEDEN"
## [75] "SWITZERLAND"
## [76] "TANZANIE"
## [77] "TCHAD"
## [78] "THAILANDE"
## [79] "TWIN TRADING"
## [80] "TZ"
## [81] "UAE"
## [82] "UNION EUROPEENNE"
## [83] "USA"
## [84] "WALTER MATTER"
## [85] "ZAMBIE"
```

```
##  [1] "AFRIQUE DU SUD"        "ALGERIE"               "ALLEMAGNE"
##  [4] "AMERIQUE LATINE"       "GRANDE BRATAGNE"       "ANGOLA"
##  [7] "ARABIE"                "ASIE"                  "AUSTRALIE"
## [10] "BELGIQUE"              "BURUNDI"               "CANADA"
## [13] "CHINE"                 "CHYPRE"                "CONGO BRAZA"
## [16] "REP TCHEQUE"           "NA"                    "EMIRATES ARABES UNIES"
## [19] "ESPAGNE"               "FRANCE"                "GABON"
## [22] "GRECE"                 "HONG KONG"             "ILE MAURICE"
## [25] "INDE"                  "ITALIE"                "JAPON"
## [28] "KENYA"                 "KP - Corée"            "LIBAN"
## [31] "LUXEMBOURG"            "MALAISIE"              "MAROC"
## [34] "NERLAND"               "PAYS BAS"              "NIGERIA"
## [37] "NOUVELLE ZELANDE"      "OUGANDA"               "PANAMA"
## [40] "PHILLIPINE"            "POLOGNE"               "PORTUGAL"
## [43] "ROYAUME UNI"           "RDC"                   "USA"
## [46] "RWANDA"                "SINGAPOUR"             "SUISSE"
## [49] "SENEGAL"               "SOMALIE"               "SOUDAN"
## [52] "SUD SOUDAN"            "SUEDE"                 "Swaziland"
## [55] "TANZANIE"              "TCHAD"                 "THAILANDE"
## [58] "UNION EUROPEENNE"      "ZAMBIE"
```

Dans la base des données il y a des entreprises que l'on a enregistré à la place des pays. ces genre des

Table 7.3: Table de corrélation entre les variables quantitatives

var1	var2	coef_corr
Weight	Year	-0.1727414
Taxe	Year	-0.1965648
Year	Weight	-0.1727414
Taxe	Weight	0.6699457
Year	Taxe	-0.1965648
Weight	Taxe	0.6699457

cas ont été traité par remplacement avec le *NA* pour **Not Available** et ces dernier on été élargués de la base des données, car nous avions jugé qu' aucune méthode d'imputation n'est applicable pour ce genre de situation. Nous avions fait la même chose pour les variables tels que **Les marchandises**.

7.1.1 Nouvelle base de données pour les analyses

Regroupement des variables pour la synthèse pour rendre la base des données simple à exploiter, éliminer les NA dans les observations telsque les pays et les valeurs pour les marchandises et les taxes.

```
DBase <- taxe_df  %>%
  select(Year,Country_dest,Goods,Weight,Taxe) %>%
  group_by(Year,Country_dest,Goods) %>%
  summarise(Weight=sum(Weight),Taxe=sum(Taxe),.groups = "drop") %>% drop_na()

correlate(DBase) %>% kable(caption = "Table de corrélation entre les variables quantitatives")
```

```
#plot_correlate(DBase)
```

```
df <- pdata.frame(DBase,index = c("Year","Country_dest"))
```

```
## Warning in pdata.frame(DBase, index = c("Year", "Country_dest")): duplicate couples (id-time) in r
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
DF <- df %>% pivot_wider(names_from = Goods,values_from = c(Taxe,Weight))
```

```
DB <- pdata.frame(DBase,index=c("Year","Goods"))
```

```
## Warning in pdata.frame(DBase, index = c("Year", "Goods")): duplicate couples (id-time) in resultin
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

7.2 Analyse descriptive des Vararíales

Conversion des données en modèle des panels des données

Chapter 8

Final Words

We have finished a nice book.