Methodology

Here is my methodology…

Panel Data

Definition of panel data

Panel data, also called longitudinal data or cross-sectional time-series data include observations on N cross section units (i.e., firms) over T time-periods.

Advantages of panel data :

As panel data analysis uses variation in both these dimensions, it is considered to be one of the most efficient analytical methods for data [@DimitriosAsteriou2006]. It usually contains more degrees of freedom, less collinearity among the variables, more efficiency and more sample variability than one-dimensional method (i.e.cross-sectional data and time series data) giving a more accurate inference of the parameters estimated in the model [@Hsiao2007, @HsiaoChapitrePanelData2014].

Fixed or random effect model

Panel data may have individual (group) effect, time effect, or both, which are analyzed by fixed effect and/or random effect models. *A fixed effect model* examines if intercepts vary across group or time period, whereas a *random effect model* explores differences in error variance components across individual or time period. [@Park2011].

** !! I need to test the fixed-random effect model of my database before moving forward !!**

```
if (!require("plm")) install.packages("plm")
```

```
## Loading required package: plm
```

```
## Loading required package: Formula
```

```
library(plm)

# I dowload my DataBase with read.csv2
DB<-data.frame(read.csv2("Analysis/DataBase/DataBase_010418.csv", sep = ";",stringsAsFactors=FALSE, head

#First Model
M1_ROA<-lm(ROA~EnergyProductivity+CarbonProductivity+WaterProductivity+WasteProductivity+GreenReputatio
```

- @Ng2015 used the two-stage-least-square regressions to estimate its models.

** In case of presence of endogeneyity in an econometric model, OLS is not capable of delivering consistent parameter estimates [@Wooldridge2008].**

Citation from [@Wooldridge2008] :

> The general concept is that of the instrumental variables estimator; a popular form of that estimator, often employed in the context of endogeneity, is known as two-stage least squares (2SLS)

Endogeneity test

Even if panel data have a lot of advantages…

Two issues involved in utilizing panel data, namely heterogeneity bias and selectivity bias [@HsiaoChapitrePanelData2014].

Citation from @HsiaoChapitrePanelData2014:

> It is only by taking proper account of selectivity and heterogeneity biases in the panel data that one can have confidence in the results obtained.

@Dangsearchrobustmethods2015 examine which methods are appropriate for estimating dynamic panel data models in empirical corporate finance,especially in short panels of company data, in the likely presence of (1) unobserved heterogeneity and endogeneity, (2) residual serial correlation, or (3) fractional dependent variables. The bias-corrected fixed-effects estimators, based on an analytical, bootstrap, or indirect inference approach, are found to be the most appropriate and robust methods.

But @MiroshnychenkoGreenpracticesfinancial2017 used the OLS regressions in micro panel using the Huber-White sand wich estimator, to account for the heteroscedasticity problem… **Which method should I use?**

Hausmann test to test the random effects model for both dependant variables?

Econometric Model

The first hypothesis have been tested with T-tests on the impact of each green initiative on green performance.

Both Hypotheses two and three have been tested by regression analysis. Econometric models are based on @Delmas2015 and @MiroshnychenkoGreenpracticesfinancial2017 and started from the general form:

$$Y_{t+1} = \beta_0 + \beta_1(X_{it}) + +\beta_2(C_{it}) + \varepsilon_{it} \tag{1}$$

where Y_{t+1} is the financial performance of firm i in year $t+1$, β is the vector of estimated regression coefficients for each of the explanatory variables X_{it}, C_{it} is a vector of control variables that include financial leverage, firm size and industry sector, $\varepsilon_{it}$is the error term.

More precisely I will test six models :

Model 1 : Green Initiatives on Tobin's Q

$$TobinsQ_{it+1} = \beta_0 + \beta_1(SP_{it}) + \beta_2(ST_{it}) + \beta_3(AS_{it}) + \beta_9(C_{it}) + \varepsilon_{it} \tag{2}$$

Model 2 : Green Initiatives on ROA

$$ROA_{it+1} = \beta_0 + \beta_1(SP_{it}) + \beta_2(ST_{it}) + \beta_3(AS_{it}) + \beta_9(C_{it}) + \varepsilon_{it} \tag{3}$$

Model 3 : Green Performance on Tobin's Q

$$TobinsQ_{it+1} = \beta_0 + \beta_1(EP_{it}) + \beta_2(CP_{it}) + \beta_3(WatP_{it}) + \beta_4(WasP_{it}) + \beta_5(GP_{it}) + \beta_9(C_{it}) + \varepsilon_{it} \tag{4}$$

Model 4 : Green Performance on ROA

$$ROA_{it+1} = \beta_0 + \beta_1(EP_{it}) + \beta_2(CP_{it}) + \beta_3(WatP_{it}) + \beta_4(WasP_{it}) + \beta_5(GP_{it}) + \beta_9(C_{it}) + \varepsilon_{it} \tag{5}$$

Model 5 : Both Green Performance and Green Initiative on Tobin's Q

$$TobinsQ_{it+1} = \beta_0+\beta_1(EP_{it})+\beta_2(CP_{it})+\beta_3(WatP_{it})+\beta_4(WasP_{it})+\beta_5(GP_{it})+\beta_6(SP_{it})+\beta_7(ST_{it})+\beta_8(AS_{it})+\beta_9(C_{it})+\varepsilon_{it} \quad (6)$$

Model 6 : Both Green Performance and Green Initiative on ROA

$$ROA_{it+1} = \beta_0+\beta_1(EP_{it})+\beta_2(CP_{it})+\beta_3(WatP_{it})+\beta_4(WasP_{it})+\beta_5(GP_{it})+\beta_6(SP_{it})+\beta_7(ST_{it})+\beta_8(AS_{it})+(C_{it})+\varepsilon_{it} \quad (7)$$

where :

- $TobinsQ_{it+1}$ = a proxy for a firm's financial performance
- ROA_{it+1} = a proxy for a firm's financial performance
- EP_{it} = a proxy for a firm's energy productivity
- CP_{it} = a proxy for a firm's carbon productivity
- $WatP_{it}$ = a proxy for a firm's water productivity
- $WasP_{it}$ = a proxy for a firm's waste productivity
- GP_{it} = a proxy for a firm's green reputation
- SP_{it} = a proxy for a firm's sustainability pay link
- ST_{it} = a proxy for a firm's sustainability themed commitment
- EP_{it} = a proxy for a firm's audit score
- C_{it} = a vector of control variables that include financial leverage, firm size and industry sector
- ε_{it} = the error term

Dans l'Equation 2, blabla ou dans l'équation 2

Sensitivity Analysis