Methodology

Panel Data : a theoretical background

This study uses the panel data methodology. Panel data is a common approach to address the CFP-CEP nexus [@Albertini2013]. It is considered to be one of the most efficient analytical methods for data analysis [@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]. @Roberts2013 also argued that using panel data offers a partial solution to the problem of omitted variables in the econometric model, namely the most common causes of endogeneity in empirical corporate finance. Panel data takes the following econometric form :

$$Y_{it} = \alpha + \beta X_{it} + u_{it} \tag{1}$$

Panel data, also called longitudinal data, includes observations on $i = 1, ..., n$ cross- section units (e.g. firms) over $t = 1, ..., T$ time-periods [@Hsiao2007]. Here, Y_{it} is the dependent variable, X_{it} represents a K-dimensional row vectors of independent variables, α is the intercept, β is a K-dimensional column vectors of parameters and u_{it} is the random disturbance term of mean equals zero. The latter can be decomposed as $u_{it} = \mu_i + \epsilon_{it}$. The first term, μ_i, represents the individual error component and is time-invariant. It can be considered as the unobserved effect model. The second term, ϵ_{it}, is the idiosyncratic error which is assumed well-behaved and independent of X_{it} and μ_i.

The starting point of all panel data is to determine if μ_i is correlated with X_{it}. In presence of correlation, then μ_i is considered as the *Fixed Effect* (i.e. FE) and the initial equation 1 becomes equation 2. Else, μ_i is considered as the *Random Effect* (i.e. RE) and the equation 1 becomes equation 3.

$$Y_{it} = (\alpha + \mu i) + \beta X_{it} + \epsilon_{it} \tag{2}$$

$$Y_{it} = \alpha + \beta X_{it} + (\epsilon_{it} + \mu i) \tag{3}$$

Fixed (i.e. Equation 2) and Random (i.e. Equation 3) Effect Model imply that the Ordinary Least Square (i.e. OLS) estimators of β are inconsistent. Five assumptions are required to produce consistent estimators with OLS : (i) a random sample of observations on y and $(x_1, ..., x_n)$, (ii) a random sample of n observations, (iii) no linear relationship among the explanatory variables, (iv) an error term that is uncorrelated with each explanatory variables and (v) an error term with zero mean conditional on the explanatory variables. FE Model violates the fourth assumption while RE model implies that the common error component over individuals induces correlation across the composite error terms making the third assumption violated [@Croissant2008].

The R package *plm* provides pertinent estimation methods to estimate panel data model. (i) *The pooled OLS estimation* ignores the panel structure of the data and applies the same coefficient to each individual [@Schmidheiny2015]. (ii) *The random effects estimation* is the feasible Generalized Least Squares estimator. (iii) *The fixed effects estimation* also called *within estimation*, transforms the original equation 1 in subtracting the time average from every variable, such as :

$$Y_{it} - \bar{Y_i} = \beta(X_{itk} - \bar{X}_{ik}) + (u_{it} - \bar{u}_i) \tag{4}$$

The presence of RE model in panel data is tested using the Breusch-Pagan Lagrange Multiplier (i.e. BPLM) test [@Breusch1980] which is represented by the *plmtest* function in *R*. It examines if time and/or individual specific variance components equal zero [@Park2011]. If Ho is verified, there is no RE model in the panel data. The presence of FE model is tested by an F test (i.e. the function *pFtest* in *R*). The latter tests the individual and/or time effects based on the comparison of the within and the pooling model [@Croissant2008]. If Ho is verified, there is no FE model in the panel data.

In case of the absence of both RE and FE model, namely $\mu_i = 0$, pooled OLS estimation is the most efficient estimator [@Croissant2008]. Under FE model, the random effects estimators are biased and inconsistent given that μ_i is omitted and potentially correlated with other regressors. Therefore, the fixed effects estimation need to be used. Under RE model, FE and RE estimators are unbiased and consistent. According to @Schmidheiny2015, scholars should prefer the RE estimator only and only if $E[\mu_i, X_i] = 0$. This precondition is tested by the Hausman test [@Hausman1981]. If Ho is verified, scholars should use RE estimator.

Econometric Model

This study uses equation 5 to study the link between outcome-based and process-based CEP and equation 6 to test their effect on CFP (short-term and long-term).

$$Y_{it} = \alpha + \beta_1 SPL_{it} + \beta_2 STC_{it} + \beta_3 A_{it} + d_t + u_{it} \tag{5}$$

where Y_{it} is a proxy of outcome-based CEP measured as carbon productivity, water productivity and waste productivity, SPL_{it} is a proxy for a firm's sustainability pay link, STC_{it} is a proxy for a firm's sustainability themed commitment, A_{it} is a proxy for a firm's audit score, d_t represents time effect and u_{it} is the error term.

$$\begin{aligned} Y_{it+1} = \quad & \alpha + \beta_1 SPL_{it} + \beta_2 STC_{it} + \beta_3 A_{it} + \beta_4 CaP_{it} \\ & + \beta_5 WatP_{it} + \beta_6 WastP_{it} + Controls_{it} + d_t + u_{it} \end{aligned} \tag{6}$$

where Y_{it+1} is a proxy of CFP measured as ROA or Tobin's Q, SPL_{it} is a proxy for a firm's sustainability pay link, STC_{it} is a proxy for a firm's sustainability themed commitment, A_{it} is a proxy for a firm's audit score, CP_{it} is a proxy for a firm's carbon productivity, $WatP_{it}$ is a proxy for a firm's water productivity, $WasP_{it}$ is a proxy for a firm's waste productivity, $Controls_{it}$ is a vector of control variables that includes firm size, industry sector, financial leverage and growth, d_t represents time effect and u_{it} is the error term.

Recent meta-analysis provided evidence of the bidirectional causality of the CFP-CEP nexus [@Orlitzky2001; @Orlitzky2003; @Wu2006; @Albertini2013; @Dixon-Fowler2013; @EndrikatMakingsenseconflicting2014;

@Ludecadedebatenexus2014, @WangMetaAnalyticReviewCorporate2016; @Busch2018]. This could cause simultaneous causality between the dependent and independent variables and lead to endogeneity concern [@Sanchez-Ballesta2007; @Biorn2008; @Roberts2013]. To address this issue, I lag observations in independent and control variables one year behind the dependent variable. This increases the confidence of the direction of the relationship [@Hart1996; @Delmas2015; @MiroshnychenkoGreenpracticesfinancial2017] and *in fine* reduces the potential simultaneity bias.