{Technical Appendix}

<u>Determine whether the effect is one-way or two-way:</u>

First, I run the general Pooling regression model, which does not take into account the one-way effect and the time effect. Then I run the one-way fixed effect model, which handles the gap between different cross-sections. I can determine if the fixed effect model should be used by looking at the results of both regression models and running a **Restricted F-test**[1]. The null hypothesis of the F-test is "There is no clear difference between the two models" and the alternative hypothesis is "One-way fixed-effect is better".

	Dependent variable:	
	lnGHG	
	(1)	(2)
lnErgpc	0.974***	1.008***
	(0.026)	(0.190)
Cleanerg_rate	-0.010***	-0.015**
	(0.002)	(0.006)
lnGDPpc	0.073***	-0.389***
	(0.010)	(0.111)
Agland_rate	0.006***	0.013***
	(0.001)	(0.003)
Constant		9.056***
		(0.625)
Observations	299	299
R2	0.954	0.195
Adjusted R2	0.952	0.184
F Statistic	1,476.296*** (df = 4; 283)	17.757*** (df = 4; 294)
Note:		*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:	
	lnGHG	
	(1)	(2)
lnErgpc	1.417***	1.008***
	(0.217)	(0.190)
Cleanerg_rate	-0.009	-0.015**
	(0.006)	(0.006)
lnGDPpc	-0.672***	-0.389***
	(0.132)	(0.111)
Agland_rate	0.013***	0.013***
	(0.003)	(0.003)
Constant		9.056***
		(0.625)
Observations	299	299
R2	0.230	0.195
Adjusted R2	0.151	0.184
F Statistic	20.209*** (df = 4; 270)	17.757*** (df = 4; 294)
Note:		*p<0.1; **p<0.05; ***p<0.01

(Figure 1.1, Left is the One-way fixed effect model)

(Figure 1.2 , Left is the Time fixed effect model)

```
> pFtest(plm_pool,plm_1FX)

        F test for individual effects

data:  lnGHG ~ lnErgpc + Cleanerg_rate + lnGDPpc + Agland_rate
F = 26.661, df1 = -11, df2 = 294, p-value = NA
alternative hypothesis: significant effects
```

(Figure2.1)

```
> pFtest(plm_pool,plm_1T)

        F test for individual effects

data:  lnGHG ~ lnErgpc + Cleanerg_rate + lnGDPpc + Agland_rate
F = 0.76356, df1 = -24, df2 = 294, p-value = NA
alternative hypothesis: significant effects
```

(Figure2.2)

[1] F-test Formula: $F = \frac{(R^2_{UR} - (R^2_R)/m}{(1-(R^2_{UR}))/(n-k)}$, Note: UR and R stand for unrestricted and restricted, respectively. Basic-Econometrics-5th-Ed-Gujarati-and-Porter, P598.

As can be seen from the results, both models are overall significant, however the One-way fixed effect model greatly improves the explanatory power of the model, with high R2 and adjR2(Figure1.1). Also, F-test result is rejecting the null hypothesis[2], suggesting that I should use the Fixed effect model (Figure2.1).

Similarly, I also compared the pooling regression model and time effect model and conducted the F-test. The result is shown in the (Figure 1.2) and (Figure2.2)[3]. I concluded that the Time-effect also exists. Therefore, to deal with both effects, I decided to use a **Two-way Fixed Effect Model**.

Determine whether it is a fixed or random effect:

In theory, greenhouse gases are produced in complex ways due to their diverse composition. I am not clear about the endogenous and exogenous variable disturbances.

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + \cdots + \beta_k X_{kit} + (a_i + u_{it})$$

In other word, I am not sure $(a_i + u_{it})$ is not correlated with any of the independent variables. But as we analyzed earlier, the probability of correlation is high. So, I assume the composite error is correlated with the independent variables. Under that assumption the Fixed effect model would be favorable. Also, because my data set is a long panel, the fixed effects model is more commonly used for it.

Testing and Correcting for near multicollinearity:

First, I build a correlation matrix for the independent variables to visually see which independent variables may correlated:

	lnErgpc	Cleanerg_rate	lnGDPpc	Agland_rate
lnErgpc	1.00000000	0.40256931	0.94648188	0.04968012
Cleanerg_rate	0.40256931	1.00000000	0.47458658	-0.05128317
lnGDPpc	0.94648188	0.47458658	1.00000000	0.01740536
Agland_rate	0.04968012	-0.05128317	0.01740536	1.00000000

[2] R shows that P-value= NA, which means that the P-value is very small. In other words, there is significant statistical evidence that the null hypothesis should be rejected.
[3] Ditto

From that we can see the **ln (Energy use-kg of oil equivalent per capita)** highly correlated with **ln(GDP per capita).** This deserves our high attention.

I also used mctest () function in R for NMC test.[4] The results are as follows:

```
Call:
omcdiag(mod = mod, Inter = Inter, detr = detr, red = red, conf = conf,
    theil = theil, cn = cn)


Overall Multicollinearity Diagnostics

                       MC Results detection
Determinant |X'X|:         0.0775         0
Farrar Chi-Square:       756.6909         1
Red Indicator:             0.4634         0
Sum of Lambda Inverse:    23.0190         1
Theil's Method:            1.4842         1
Condition Number:         67.5524         1

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

```
Call:
imcdiag(mod = mod, method = method, corr = FALSE, vif = vif,
    tol = tol, conf = conf, cvif = cvif, ind1 = ind1, ind2 = ind2,
    leamer = leamer, all = all)


 VIF Multicollinearity Diagnostics

                 VIF detection
lnErgpc       9.9517         0
Cleanerg_rate 1.3300         0
lnGDPpc      10.7236         1
Agland_rate   1.0137         0

Multicollinearity may be due to lnGDPpc regressors

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

The independent variable **ln(GDP per capita)** indeed cause the NMC. So, I locked ln(GDP per capita) as the target variable. To study how it correlated to other explanatory variables, I ran an **auxiliary regression**.

ln(GDP per capita) ~ lnErgpc + Cleanerg_rate + Agland_rate

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -3.352481   0.262097 -12.791  < 2e-16 ***
lnErgpc        1.600820   0.034521  46.373  < 2e-16 ***
Cleanerg_rate  0.016080   0.002853   5.636 4.07e-08 ***
Agland_rate   -0.002140   0.001749  -1.223    0.222    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5175 on 295 degrees of freedom
Multiple R-squared:  0.9067,	Adjusted R-squared:  0.9058 
F-statistic: 956.2 on 3 and 295 DF,  p-value: < 2.2e-16
```

(Figure 3.1)

```
lm(formula = lnGDPpc ~ lnErgpc, data = P507_1)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.52197 -0.35933 -0.01968  0.39643  1.32554 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.85728    0.25719  -15.00   <2e-16 ***
lnErgpc      1.67723    0.03319   50.54   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5451 on 297 degrees of freedom
Multiple R-squared:  0.8958,	Adjusted R-squared:  0.8955 
F-statistic:  2554 on 1 and 297 DF,  p-value: < 2.2e-16
```

(Figure 3.2)

The overall F statistic is significant, and the aux regression model’s R^2 and adjR^2 near to 0.9, which means a very serious correlation. (Figure 3.1). From the individual t-test, we can also find that **Alternative and nuclear energy (% of total energy use)** is also correlated with ln (GDP per capita). To figure it out, I ran another auxiliary regression which dropped the lnErgpc.

ln (GDP per capita) ~ Cleanerg_rate + Agland_rate

[4] mctest () cannot be directly applied to the Fixed effect model. My solution is to build an OLS regression model using lm () function and implement mctest on that model.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   8.099380   0.252365  32.094   <2e-16 ***
Cleanerg_rate 0.069816   0.007494   9.317   <2e-16 ***
Agland_rate   0.004101   0.005014   0.818    0.414    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.487 on 296 degrees of freedom
Multiple R-squared:  0.227,	Adjusted R-squared:  0.2218 
F-statistic: 43.46 on 2 and 296 DF,  p-value: < 2.2e-16
```

The overall F-statistic of this auxiliary regression is still significant, and the P-value of the individual t-test of Cleaning_rate is very small. However, the overall explanatory power of the model drops to 22%, which indicates that the Alternative and nuclear energy (% of total energy use) is not seriously correlated to GDP per capita. So we only need to deal with the multicollinearity of the ln(Energy use-kg of oil equivalent per capita) and the ln(GDP per capita).

We can understand the reason why those two variables are correlated. On the one hand, as a country's GDP per capita increases, its people will be more able to pay for transportation like cars and use energy to maintain comfortable temperatures in summer and winter. Also, countries with high GDP per capita are generally highly developed, with more buildings and more lighting in their cities. So, an increase in GDP per capita naturally implies an increase in energy consumption per capita. On the other hand, the increase in energy consumption per capita is bound to mean the increase in GDP per capita. Because most of the energy consumption is used to do production and service activities. This is also confirmed by our auxiliary regression (Figure 3.2), where changes in ln (energy use per capita) have up to 89.6% explanatory power for changes in ln(GDP per capita). Considering such high correlation, my decision is to remove the explanatory variable ln(GDP per capita) from the model.[5]

After the adjustment, I conducted mctest on the new model.

[5] When this variable is removed, the overall explanatory power of the model decrease only slightly.

```
Call:
imcdiag(mod = mod, method = method, corr = FALSE, vif = vif,
    tol = tol, conf = conf, cvif = cvif, ind1 = ind1, ind2 = ind2,
    leamer = leamer, all = all)


 VIF Multicollinearity Diagnostics

                 VIF detection
lnErgpc       1.2005         0
Cleanerg_rate 1.2007         0
Agland_rate   1.0086         0

NOTE:  VIF Method Failed to detect multicollinearity


0 --> COLLINEARITY is not detected by the test
```

(Figure 4.1)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.153746   0.147833  48.391  < 2e-16 ***
Cleanerg_rate 0.033568   0.004390   7.647  2.9e-13 ***
Agland_rate   0.003899   0.002937   1.327    0.185
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8713 on 296 degrees of freedom
Multiple R-squared:  0.167,     Adjusted R-squared:  0.1614
F-statistic: 29.68 on 2 and 296 DF,  p-value: 1.794e-12
```

(Figure 4.2)[6]

It shows that no multicollinearity is detected. (Figure 4.1) And I take **ln (Energy use-kg of oil equivalent per capita)** as the dependent variable and conduct an auxiliary regression with other independent variables. The regression shows it is correlated with **Alternative and nuclear energy (% of total energy use)**, but the explanatory power of the overall model is only 16% (Figure 4.2), which is an acceptable correlation. At this point, the near multicollinearity issue in the model is solved.

<u>Detecting and Correcting for Autocorrelation:</u>

Considering that we use panel data and adopt a Fixed-Effect regression model. We are best suited to examine the issue of autocorrelation using Wooldridge's test.[7]

```
        Wooldridge's test for serial correlation in FE panels

data:  plm.model
F = 1147.3, df1 = 1, df2 = 272, p-value < 2.2e-16
alternative hypothesis: serial correlation
```

The result is that Autocorrelation is indeed present in our fixed effects model.

I also run the Durbin-Watson Test[8] on my regression to detect the Autocorrelation, and the result shows that:

---

[6] ln (Energy use-kg of oil equivalent per capita) ~ Cleanerg_rate + Agland_rate

[7] An autocorrelation detection method specifically for Panel Data. In R, the pwartest() function under the plm package can help us do that test. Its null hypothesis is "there is no Autocorrelation." And its alternative hypothesis is “there is Autocorrelation”. For the mechanism and proof of the test please see Wooldridge JM (2002). Econometric Analysis of Cross-Section and Panel Data. MIT Press.

[8] Because my data are long-Panel data, it contains 25 years of observations. So, it is appropriate to use Durbin-Watson test method. (Strictly speaking, this test should not be used with panel data.) Its null hypothesis is "there is no Autocorrelation." And its alternative hypothesis is “there is Autocorrelation”.

```
lag Autocorrelation D-W Statistic p-value
  1       0.9492827    0.08987088       0
Alternative hypothesis: rho != 0
```

obs=299, K*=3, D_L=1.643, D_H=1.704

Since 0.0898 < 1.643, we can reject the H_0, and conclude that there is Autocorrelation. To solve that issue, I use the **Cochrane-Orcutt Iterative Process** to estimate the autocorrelation coefficient $\boldsymbol{\rho}$. Then I use it to transform all the variables in my original model and use those transformation variables to run the final Two-way Fixed effect model. The results are as follows:

```
Durbin-Watson statistic (original): 0.08987
Durbin-Watson statistic (transformed): 1.966
```

The autocorrelation is resolved.

For the presentation and interpretation of the final model results, please refer to Regression Results.

References:

Damodar N. Gujarati, Dawn C. Porter. (2009) Basic Econometrics Fifth Edition Published by McGraw-Hill/Irwin

IPCC (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Stocker, T.F., D. Qin, G.K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press

IPCC (2007) Climate Change 2007: AR4 Climate Change 2007: Mitigation of Climate Change

Jonathan Kropko, Robert Kubinec. 2020 Interpretation and identification of within-unit and cross-sectional variation in panel data models

Methane as a Greenhouse Gas: Why the EPA Should Regulate Emissions from Animal Feeding Operations and Concentrated Animal Feeding Operations Under the Clean Air Act Author(s): JOHN VERHEUL Source: Natural Resources Journal , Spring 2011, Vol. 51, No. 1 (Spring 2011), pp. 163- 187

Smith, P., D. Martino, Z. Cai, D. Gwary, H. Janzen, P. Kumar, B. McCarl, S. Ogle, F. O'Mara, C. Rice, B. Scholes, O. Sirotenko, 2007: Agriculture. In Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Wooldridge JM (2002). Econometric Analysis of Cross-Section and Panel Data. MIT Press.

Yves Croissant, Giovanni Millo.(2008) Panel Data Econometrics in R: The plm Package. Journal of Statistical Software