

Chapitre 3 - Modèle 1 : codes de réplcation et annexes

Table of contents

1	Spécification de base : modèles FE-OLS	2
1.1	Modèle de base (FE-OLS)	2
1.2	Test de non-linéarité (quadratique)	3
1.3	Robustesse informalité : <code>mimic</code> puis <code>dge</code>	4
1.4	Interaction <code>GreenPotSharew</code> \times informalité (MIMIC et DGE)	6
1.5	Régression à seuil avec <code>segmented</code>	9
1.6	Bootstrap pour intervalles de confiance du seuil	14
1.7	Modèle “piecewise” en FE-OLS (seuil fixe à 0,26)	17
1.8	Modèle à spline cubique sur <code>GreenPotSharew</code> (FE-OLS avec spline cubique sur <code>GreenPotSharew</code> (4 df internes))	18
1.9	Vérification des observations extrêmes	20
1.10	Test de Chow à coupure	22
1.11	Test de robustesse en excluant les observations les plus influentes	24
2	Endogénéité et IV	29
2.1	IV externes	29
2.1.1	Préparation des instruments solaires et forestiers	29
2.1.2	Importation et transformation des prix de commodités	30
2.1.3	Chargement et pivot des financements IRENA	31
2.1.4	Hydrologie, littoral et rugosité	32
2.1.5	Fusion de tous les IV dans le panel	36
2.1.6	IV-FE par instruments externes géographiques et solaires	37
2.2	IV-FE par lags internes de <code>GreenPotSharew</code>	51
3	Hétérogénéité Advanced vs Developing	56
3.1	Modèle linéaire	56
3.2	Non-linéarité du verdissement dans chaque sous-échantillon	58
3.3	Effet marginal du verdissement (delta-method + graphiques)	60

4	Panel GMM dynamique (Arellano-Bond diff & System-GMM)	65
5	Hétérogénéité pays-par-pays et slopes	71
5.1	Estimation pays-par-pays (Driscoll-Kraay)	71
5.2	Modèle à effets mixtes	76
6	Difference-in-Differences (seuil de 0,26 et year 2008)	80
7	Tests d'hétérogénéité institutionnelle et géographique	85
7.1	Informalité	85
7.2	Caractéristiques géographiques	95
7.3	Qualité institutionnelle	97

```
efpv_sustainability <- readRDS("/Users/.../Chap3/efpv_sustainability.rds")
df1_mod1 <- efpv_sustainability
```

1 Spécification de base : modèles FE-OLS

1.1 Modèle de base (FE-OLS)

Le code ci-dessous estime un modèle en effets fixes (pays et année), avec erreurs standard clusterisées par pays. Les résultats sont arrondis, annotés selon leur significativité. Les résultats sont présentés tableau A.1.

```
library(dplyr)
library(fixest)

# Estimation en effets fixes (EF) avec clustering par pays
mod1_feols <- feols(
  DeltaY ~
    deltaLnK          + # Variation du capital
    DeltaLnL          + # Variation du travail
    GreenPotSharew    + # Part d'EFPV
    self_emp_pct      + # Part de l'auto-emploi
    ays               + # Âge moyen d'études
    gov_exp_pctgdp    + # Dépenses publiques en % du PIB
    open_pct          + # Ouverture commerciale
    democracy         + # Indice de démocratie
    access_electricity + # Accès à l'électricité
    hhi_va            # Concentration sectorielle
  | iso3c + year,      # Effets fixes pays et année
```

```

cluster = "iso3c",          # Clustering des erreurs par pays
data    = df1_mod1
)

# Affichage du résumé des résultats
summary(mod1_feols)

```

Tableau A.1 – Estimation FE-OLS de ΔY

Variable	Estimateur	Erreur-type	t-stat	p-value
$\Delta \ln K$	0.129	(0.033)	3.864	0.0003***
$\Delta \ln L$	0.048	(0.115)	0.418	0.6776
GreenPotSharew	0.151	(0.173)	0.871	0.3871
self_emp_pct	-0.002	(0.001)	-1.245	0.2175
ays	-0.004	(0.010)	-0.364	0.7170
gov_exp_pctgdp	-0.006	(0.002)	-3.149	0.0025**
open_pct	-0.0001	(0.000)	-0.199	0.8429
democracy	-0.028	(0.014)	-1.920	0.0593.
access_electricity	0.0002	(0.000)	0.475	0.6366
hhi_va	0.473	(0.383)	1.234	0.2217
RMSE	0.035			
Adj. R²	0.692			
Within R²	0.287			

Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.1$

1.2 Test de non-linéarité (quadratique)

Ce code estime le même modèle qu'en 1.1, en ajoutant le terme (GreenPotSharew^2). Les résultats sont présentés tableau A.2.

```

# Estimation du modèle avec terme quadratique
model_quad <- feols(
  DeltaY ~ deltalnK + DeltaLnL
    + GreenPotSharew + I(GreenPotSharew^2)
    + self_emp_pct + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year,
  data    = df1_mod1,

```

```

cluster = "iso3c"
)

summary(model_quad)

```

Tableau A.2 – Estimation FE-OLS de ΔY avec terme quadratique de GreenPotSharew

Variable	Estimateur	Erreur-type	t-stat	p-value
$\Delta \ln K$	0.129	(0.034)	3.848	0.0003***
$\Delta \ln L$	0.048	(0.115)	0.413	0.6813
GreenPotSharew	0.222	(0.360)	0.617	0.5396
I(GreenPotSharew ²)	-0.122	(0.374)	-0.326	0.7456
self_emp_pct	-0.002	(0.001)	-1.252	0.2151
ays	-0.004	(0.010)	-0.369	0.7134
gov_exp_pctgdp	-0.006	(0.002)	-3.190	0.0022**
open_pct	-0.0001	(0.000)	-0.183	0.8557
democracy	-0.028	(0.014)	-1.917	0.0597.
access_electricity	0.0002	(0.000)	0.402	0.6890
hhi_va	0.474	(0.388)	1.223	0.2257
RMSE	0.035			
Adj. R²	0.691			
Within R²	0.287			

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$*

1.3 Robustesse informalité : mimic puis dge

Afin de vérifier que les résultats ne dépendent pas de la mesure de l'informalité, deux indicateurs alternatifs sont introduits :

1- MIMIC : Mesure de l'économie informelle estimée par la méthode MIMIC à partir d'un ensemble d'indicateurs (participation au marché du travail, consommation d'électricité non déclarée, etc.).

2- DGE : Estimation de l'économie informelle à partir d'un modèle de génération de données (Data Generating Engine, DGE) simulant les comportements de déclarations.

Chaque variante remplace la variable self_emp_pct du modèle de base par l'alternative correspondante. Le même jeu de régresseurs, les effets fixes et le clustering sont conservés.

```

# Modèle avec informalité MIMIC
mod1_mimic <- feols(
  DeltaY ~ deltaLnK + DeltaLnL
    + GreenPotSharew
    + mimic    # nouvelle mesure
    + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year,
  data    = df1_mod1,
  cluster = ~iso3c
)
summary(mod1_mimic)

# Modèle avec informalité DGE
mod1_dge <- feols(
  DeltaY ~ deltaLnK + DeltaLnL
    + GreenPotSharew
    + dge      # nouvelle mesure
    + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year,
  data    = df1_mod1,
  cluster = ~iso3c
)
summary(mod1_dge)

```

Tableau A.3 – Robustesse de la mesure de l’informalité (MIMIC vs DGE)

Variable	Coef. (MIMIC)	Std. Err.	t-stat	p-value	Coef. (DGE)	Std. Err.	t-stat	p-value
$\Delta \ln K$	0.146***	(0.032)	4.529	0.0000	0.146***	(0.032)	4.615	0.0000
$\Delta \ln L$	-0.011	(0.081)	-0.130	0.8970	0.063	(0.083)	0.753	0.4550
GreenPotSharew	0.029	(0.031)	0.930	0.3569	0.040	(0.043)	0.934	0.3549
Informalité	-0.006	(0.005)	-1.227	0.2255	-0.011.	(0.006)	-1.851	0.0698
ays	-0.013	(0.009)	-1.438	0.1563	-0.013	(0.009)	-1.529	0.1324
gov_-	-	(0.001)	-5.212	0.0000	-	(0.001)	-5.406	0.0000
exp_-	0.004***				0.005***			
pctgdp								

Variable	Coef. (MIMIC)	Std. Err.	t-stat	p-value	Coef. (DGE)	Std. Err.	t-stat	p-value
open__- pct	-0.0001	(0.000)	-0.187	0.8526	0.0003	(0.000)	0.949	0.3469
democracy	-0.032.	(0.016)	-1.964	0.0549	-0.028*	(0.013)	-2.063	0.0442
access__- electric- ity	0.001	(0.000)	1.134	0.2620	-0.0004	(0.001)	-0.680	0.4995
hhi_va	0.201	(0.211)	0.955	0.3442	-0.045	(0.201)	-0.224	0.8235
RMSE	0.023				0.022			
Adj. R²	0.811				0.819			
Within R²	0.448				0.472			
Observations	323				323			
Fixed- effects	iso3c: 53, year: 16				iso3c: 53, year: 16			
Std. errors	Clustered (iso3c)				Clustered (iso3c)			

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.1$*

1.4 Interaction GreenPotSharew × informalité (MIMIC et DGE)

Afin d'évaluer si l'effet du verdissement varie selon le niveau d'informalité, on ajoute dans un même modèle deux interactions : 1) GreenPotSharew × mimic + 2) GreenPotSharew × dge :

```
library(fixest)
library(modelsummary)

# 1) Définition des contrôles communs
controls <- "ays + gov_exp_pctgdp + open_pct + democracy + access_electricity
↪ + hhi_va"

# 2) Modèle A : interaction avec MIMIC
fml_A <- as.formula(paste0(
  "DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew + mimic + ",
  "GreenPotSharew:mimic + ", controls,
```

```

" | iso3c + year"
))
modA <- feols(fml_A, data = df1_mod1, cluster = "iso3c")

# 3) Modèle B : interaction avec DGE
fml_B <- as.formula(paste0(
  "DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew + dge + ",
  "GreenPotSharew:dge + ", controls,
  " | iso3c + year"
))
modB <- feols(fml_B, data = df1_mod1, cluster = "iso3c")

# 4) Modèle C : interaction avec Self-emp %
fml_C <- as.formula(paste0(
  "DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew + self_emp_pct + ",
  "GreenPotSharew:self_emp_pct + ", controls,
  " | iso3c + year"
))
modC <- feols(fml_C, data = df1_mod1, cluster = "iso3c")

# 5) Présentation côte-à-côte avec modelsummary
coef_map <- c(
  "deltaLnK"           = "Δ ln K",
  "DeltaLnL"           = "Δ ln L",
  "GreenPotSharew"     = "GreenPotSharew",
  "mimic"              = "MIMIC index",
  "dge"                = "DGE index",
  "self_emp_pct"       = "Self-emp %",
  "GreenPotSharew:mimic" = "GreenPotSharew × MIMIC",
  "GreenPotSharew:dge"  = "GreenPotSharew × DGE",
  "GreenPotSharew:self_emp_pct" = "GreenPotSharew × Self-emp",
  "ays"                = "Avg years schooling",
  "gov_exp_pctgdp"     = "Gov exp (% GDP)",
  "open_pct"           = "Open (% GDP)",
  "democracy"          = "Democracy index",
  "access_electricity" = "Access electricity",
  "hhi_va"             = "HHI VA"
)

modelsummary(
  list("Modèle A\n(interaction MIMIC)" = modA,
        "Modèle B\n(interaction DGE)"   = modB,
        "Modèle C\n(interaction SE)"    = modC),

```

```

statistic = "p.value",
stars      = c('***' = 0.001, '**' = 0.01, '*' = 0.05),
coef_map  = coef_map,
output     = "markdown",
title      = "Effets différenciés de l’informalité sur le verdissement"
)

```

Tableau A.4 – Effets différenciés de l’informalité sur le verdissement

Variable	Modèle A (MIMIC)	Modèle B (DGE)	Modèle C (Self-emp %)
$\Delta \ln K$	0.145*** (p < 0.001)	0.143*** (p < 0.001)	0.133*** (p < 0.001)
$\Delta \ln L$	-0.017 (p = 0.844)	0.039 (p = 0.653)	0.012 (p = 0.929)
GreenPotSharew	-0.032 (p = 0.566)	-0.249* (p = 0.018)	-0.169 (p = 0.301)
MIMIC index	-0.007 (p = 0.170)	—	—
DGE index	—	-0.016** (p = 0.007)	—
Self-emp %	—	—	-0.005 (p = 0.058)
GreenPotSharew × MIMIC	0.003 (p = 0.297)	—	—
GreenPotSharew × DGE	—	0.015* (p = 0.021)	—
GreenPotSharew × Self-emp %	—	—	0.013 (p = 0.223)
Avg years schooling	-0.013 (p = 0.155)	-0.013 (p = 0.144)	-0.006 (p = 0.576)
Gov exp (% GDP)	-0.004*** (p < 0.001)	-0.004*** (p < 0.001)	-0.005*** (p < 0.001)
Open (% GDP)	0.000 (p = 0.858)	0.000 (p = 0.273)	0.000 (p = 0.952)
Democracy index	-0.032 (p = 0.051)	-0.027* (p = 0.038)	-0.031 (p = 0.053)
Access electricity	0.001 (p = 0.298)	-0.001 (p = 0.151)	-0.000 (p = 0.657)
HHI VA	0.191 (p = 0.361)	-0.075 (p = 0.690)	0.553 (p = 0.230)
Observations	323	323	414
Within R²	0.448	0.485	0.314
Adjusted Within R²	0.423	0.462	0.291
RMSE	0.020	0.020	0.030
Std. errors	Clustered (iso3c)	Clustered (iso3c)	Clustered (iso3c)
Fixed-effects	iso3c, year	iso3c, year	iso3c, year

*Signif. codes: *** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1*

#Recherche et validation du seuil critique (piecewise, bootstrap, spline, outliers, tests)

La recherche d'un point de rupture (« threshold ») sur l'effet de GreenPotSharew permet d'identifier à partir de quel niveau de verdissement la significativité de la relation avec la croissance change. Plusieurs approches complémentaires sont mobilisées : 1) Régression à seuil (piecewise / segmented regression), 2) bootstrap de l'estimation du seuil, 3) Spline cubique pour flexibilité non linéaire, 4) Identification et rôle des observations extrêmes (outliers), 5) Test de choix pour valider la discontinuité.

```
mod1_df2 <- df1_mod1
```

1.5 Régression à seuil avec segmented

```
if (!requireNamespace("segmented", quietly = TRUE)) {  
  install.packages("segmented")  
}  
library(segmented)  
  
# 1) Point de départ du seuil : médiane de GreenPotSharew  
psi0 <- median(mod1_df2$GreenPotSharew, na.rm = TRUE)  
  
# 2) Estimation des modèles OLS de base  
bases <- list(  
  Self_emp = lm(  
    DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew +  
      self_emp_pct + ays + gov_exp_pctgdp +  
      open_pct + democracy + access_electricity + hhi_va,  
    data = mod1_df2  
  ),  
  DGE = lm(  
    DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew +  
      dge + ays + gov_exp_pctgdp +  
      open_pct + democracy + access_electricity + hhi_va,  
    data = mod1_df2  
  ),  
  MIMIC = lm(  
    DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew +  
      mimic + ays + gov_exp_pctgdp +  
      open_pct + democracy + access_electricity + hhi_va,  
    data = mod1_df2  
  )  
)
```

```

# 3) Ajustement des modèles segmentés
segs <- lapply(bases, function(m) {
  segmented(m,
    seg.Z = ~GreenPotSharew,
    psi   = list(GreenPotSharew = psi0))
})

# 4) Fonction pour extraire coefficients + p-values
extract_coefs <- function(segobj) {
  tab  <- summary(segobj)$coefficients
  dfR  <- df.residual(segobj)
  tval <- tab[, "Estimate"] / tab[, "Std. Error"]
  pval <- 2 * pt(-abs(tval), dfR)
  res  <- cbind(tab, "p.value" = pval)
  return(as.data.frame(res, stringsAsFactors = FALSE))
}

# 5) Construction du tableau des coefficients pour tous les modèles
coefs_list <- lapply(segs, extract_coefs)

coefs_df <- do.call(rbind, lapply(names(coefs_list), function(m) {
  df <- coefs_list[[m]]
  df$Term <- rownames(df)
  df$Model <- m
  df
})))

coefs_df <- coefs_df[, c("Model", "Term", "Estimate", "Std. Error", "t
↵ value", "p.value")]

# 6) Extraction des statistiques globales de chaque modèle
stats_df <- do.call(rbind, lapply(names(segs), function(m) {
  s  <- summary(segs[[m]])
  n  <- length(segs[[m]]$residuals)
  sigma <- s$sigma
  adjr2 <- s$adj.r.squared
  data.frame(Model = m,
    Observations = n,
    Residual_SE = sigma,
    Adj_R2 = adjr2,
    row.names = NULL)
})))

```

```
# 7) Résultats
cat("=== Statistiques globales ===\n")
print(stats_df)
cat("\n=== Coefficients détaillés ===\n")
print(coefs_df)
```

Tableau A.5 – Régression à seuil (segmented) pour trois mesures de l’informalité

Terme	Self-emp (n=414)	DGE (n=323)	MIMIC (n=323)
(Intercept)	0.019 (0.050) p = 0.707	0.134*** (0.024) p < 0.001	0.134*** (0.024) p < 0.001
$\Delta \ln K$	0.237*** (0.017) p < 0.001	0.248*** (0.017) p < 0.001	0.247*** (0.018) p < 0.001
$\Delta \ln L$	-0.023 (0.100) p = 0.819	0.268* (0.105) p = 0.011	0.268* (0.105) p = 0.011
GreenPotSharew (avant seuil)	-0.234 (0.151) p = 0.123	-0.120 (0.077) p = 0.120	-0.121 (0.077) p = 0.120
Informalité (avant seuil)	0.00119*** (0.00028) p < 0.001	-0.00019 (0.00027) p = 0.487	-0.00015 (0.00024) p = 0.526
Avg years schooling	-0.00384* (0.00169) p = 0.023	-0.00314* (0.00157) p = 0.046	-0.00309* (0.00156) p = 0.048
Gov exp (% GDP)	-0.00076* (0.00030) p = 0.011	-0.00092** (0.00028) p = 0.0016	-0.00092** (0.00028) p = 0.0013
Open (% GDP)	0.00017* (0.00008) p = 0.036	0.00005 (0.00007) p = 0.469	0.00005 (0.00007) p = 0.479
Democracy index	0.00073 (0.00243) p = 0.763	0.00431. (0.00223) p = 0.0536	0.00428. (0.00225) p = 0.0587
Access electricity	0.00017 (0.00025) p = 0.503	-0.00005 (0.00020) p = 0.795	-0.00058 (0.00020) p = 0.776
HHI VA	0.0688 (0.0512) p = 0.180	-0.0744. (0.0393) p = 0.0591	-0.0745. (0.0394) p = 0.0593
Changement de pente (post-seuil)	0.429** (0.158) p = 0.007	0.241. (0.125) p = 0.0538	0.242. (0.125) p = 0.0546
Observations	414	323	323
Residual S.E.	0.0528	0.0403	0.0403
Adj. R²	0.4651	0.5557	0.5556
Std. errors	Clustered (iso3c)	Clustered (iso3c)	Clustered (iso3c)
Fixed-effects	iso3c, year	iso3c, year	iso3c, year

*Signif. codes: *** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1*

Avant d'intégrer ces proxys d'informalité dans nos modèles, il est essentiel de s'assurer de leur qualité et de leur complémentarité : 1. Couverture des données : vérifier le nombre de valeurs manquantes pour chaque proxy afin d'anticiper d'éventuelles pertes d'observations. 2. Statistiques descriptives : examiner la moyenne, l'écart-type, le minimum et le maximum pour comprendre la distribution de chaque proxy. 3. Corrélations : estimer la matrice de corrélation pour détecter d'éventuelles redondances ou multicolinéarité entre les proxys.

```
library(dplyr)
library(fixest)
library(modelsummary)

# 1) Vérification de la couverture (valeurs manquantes)
mod1_df2 %>%
  summarise(across(c(self_emp_pct, mimic, dge),
    ~ sum(is.na(.)),
    .names = "NA_{col}"))

# 2) Statistiques descriptives
mod1_df2 %>%
  summarise(across(c(self_emp_pct, mimic, dge),
    list(
      mean = ~ mean(., na.rm = TRUE),
      sd   = ~ sd(.,   na.rm = TRUE),
      min  = ~ min(.,  na.rm = TRUE),
      max  = ~ max(.,  na.rm = TRUE)
    )))

# 3) Corrélations entre les trois proxys d'informalité (deux méthodes)

# Méthode 1 : dplyr::select explicite
df_inf <- mod1_df2 %>%
  dplyr::select(self_emp_pct, mimic, dge)

cor(df_inf, use = "pairwise.complete.obs")

# Méthode 2 : base R
cor(
  mod1_df2[, c("self_emp_pct", "mimic", "dge")],
  use = "pairwise.complete.obs"
)
```

Tableau A.6 – Nombre de valeurs manquantes pour chaque proxy d'informalité

Variable	Nombre de NA
self_emp_pct	15
mimic	238
dge	235

Tableau A.7 – Statistiques descriptives des proxys d’informalité

Variable	Moyenne	Écart-type	Minimum	Maximum
self_emp_pct	39.111	22.474	3.489	92.776
mimic	31.295	13.810	8.070	64.165
dge	28.258	12.510	7.910	62.398

Tableau A.8 – Matrice de corrélation (pairwise.complete.obs)

	self_emp_pct	mimic	dge
self_emp_pct	1.000	0.564	0.524
mimic	0.564	1.000	0.975
dge	0.524	0.975	1.000

Les diagnostics réalisés sur les proxys d’informalité mettent en évidence une couverture inégale des données, des caractéristiques statistiques différenciées et des corrélations variables. Le taux d’auto-emploi (**self_emp_pct**) présente seulement quinze valeurs manquantes, tandis que les indices MIMIC et DGE totalisent respectivement deux cent trente-huit et deux cent trente-cinq NA, conduisant à la suppression d’environ cent quinze observations supplémentaires et à une perte de puissance statistique. Sur le plan descriptif, la moyenne de **self_emp_pct** est proche de 39 % avec une dispersion importante (écart-type de 22 %), traduisant une hétérogénéité marquée entre pays, alors que MIMIC (moyenne 31 %, écart-type 14 %) et DGE (moyenne 28 %, écart-type 12 %) offrent des estimations plus concentrées et stables.

La corrélation quasi parfaite entre MIMIC et DGE ($r = 0,975$) suggère qu’ils mesurent essentiellement la même dimension de l’économie informelle, tandis que leurs liens plus modérés avec **self_emp_pct** ($r = 0,56$ et $r = 0,52$) révèlent que ce dernier capte une facette distincte, centrée sur l’auto-emploi. Ces constats soulignent que le choix du proxy affecte à la fois la taille de l’échantillon exploitable et la robustesse des résultats : MIMIC et DGE, interchangeables en termes de dimension informelle, présentent l’inconvénient d’une couverture réduite, alors que **self_emp_pct** assure une couverture quasi intégrale et apporte une mesure complémentaire de l’informalité.

1.6 Bootstrap pour intervalles de confiance du seuil

L'estimation procède en deux temps :

1. Une première passe avec `segmented`, initialisée à la médiane de `GreenPotSharew`, pour obtenir $((\hat{\psi}))$ et les coefficients $(_1)$ (avant seuil) et $(_2)$ (après seuil).
2. Un bootstrap par pays ($B=500$) répétitions, tirage avec remise sur `iso3c` pour construire un intervalle de confiance à 95 % de $((\hat{\psi}))$, garantissant la robustesse de l'estimation du threshold. La démarche tire un échantillon bootstrap par pays, estime le modèle linéaire de base, applique `segmented` à partir de la médiane initiale de `GreenPotSharew`, extrait le seuil $((\hat{\psi}))$, et répète pour construire la distribution empirique de $((\hat{\psi}))$.

```
# 0) Charger les packages
if (!requireNamespace("segmented", quietly = TRUE)) {
  install.packages("segmented")
}
library(segmented)
library(dplyr)
library(modelsummary)

# 1) Point de départ global
psi0_global <- median(mod1_df2$GreenPotSharew, na.rm = TRUE)

# 2) Estimation des OLS de base
bases <- list(
  Self_emp = lm(DeltaY ~ deltaInK + DeltaLnL + GreenPotSharew +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va,
    data = mod1_df2),
  DGE      = lm(DeltaY ~ deltaInK + DeltaLnL + GreenPotSharew +
    dge      + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va,
    data = mod1_df2),
  MIMIC    = lm(DeltaY ~ deltaInK + DeltaLnL + GreenPotSharew +
    mimic    + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va,
    data = mod1_df2)
)

# 3) Affichage des résultats OLS de base
modelsummary(
  bases,
```

```

stars      = c('***' = 0.001, '**' = 0.01, '*' = 0.05, '.' = 0.1),
coef_map = c(
  "(Intercept)"      = "Constante",
  "deltalnK"          = "Δ ln K",
  "DeltaLnL"          = "Δ ln L",
  "GreenPotSharew"    = "GreenPotSharew",
  "self_emp_pct"      = "Self-emp %",
  "dge"               = "DGE index",
  "mimic"             = "MIMIC index",
  "ays"               = "Avg years schooling",
  "gov_exp_pctgdp"    = "Gov exp (% PIB)",
  "open_pct"          = "Open (% PIB)",
  "democracy"         = "Democracy index",
  "access_electricity" = "Access electricity",
  "hhi_va"            = "HHI VA"
),
title      = "Tableau A.9 - Estimations OLS de base avant segmentation",
output     = "markdown"
)

# 4) Ajustement des segmentés initiaux (comme précédemment)
segs <- lapply(bases, function(m) {
  segmented(m, seg.Z = ~GreenPotSharew, psi = list(GreenPotSharew =
    ↪ psi0_global))
})

# 5) Tableau A.10 - Modèles segmentés avec Std.Errors et p-values
modelsummary(
  segs,
  statistic = c("std.error", "p.value"),          # deux lignes sous chaque
  ↪ coef
  stars      = c('***' = 0.001, '**' = 0.01, '*' = 0.05, '.' = 0.1),
  coef_map = c(
    "(Intercept)"      = "Constante",
    "GreenPotSharew"    = "Pente avant seuil",
    "U1.GreenPotSharew" = "Variation de pente après seuil",
    "deltalnK"          = "Δ ln K",
    "DeltaLnL"          = "Δ ln L",
    "self_emp_pct"      = "Self-emp %",
    "dge"               = "DGE index",
    "mimic"             = "MIMIC index",
    "ays"               = "Avg years schooling",
    "gov_exp_pctgdp"    = "Gov exp (% PIB)",

```

```

    "open_pct"          = "Open (% PIB)",
    "democracy"         = "Democracy index",
    "access_electricity" = "Access electricity",
    "hhi_va"            = "HHI VA"
  ),
  title      = "Tableau A.10 - Modèles segmentés (pente avant et variation
    ↪ après seuil)",
  output     = "markdown"
)

# 6) Extraction des seuils initiaux
proxies <- c(Self_emp="self_emp_pct", DGE="dge", MIMIC="mimic")
psi0_list <- sapply(names(proxies), function(key) {
  segs[[key]]$psi["psi1.GreenPotSharew", "Est."]
}, simplify = FALSE)

# 7) Fonction bootstrap
bootstrap_seuil <- function(var, psi0) {
  replicate(500, {
    samp <- mod1_df2 %>% group_by(iso3c) %>% sample_frac(1, replace = TRUE)
    ↪ %>% ungroup()
    fml <- as.formula(paste0(
      "DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew + ", var,
      " + ays + gov_exp_pctgdp + open_pct + democracy + access_electricity +
      ↪ hhi_va"
    ))
    m <- lm(fml, data = samp)
    sg <- try(segmented(m, seg.Z = ~GreenPotSharew, psi =
    ↪ list(GreenPotSharew = psi0)), silent = TRUE)
    if (inherits(sg, "try-error")) return(NA_real_)
    sg$psi["psi1.GreenPotSharew", "Est."]
  })
}

# 8) Bootstrap et IC
set.seed(123)
ci_table <- do.call(rbind, lapply(names(proxies), function(key) {
  var <- proxies[[key]]
  psi0 <- psi0_list[[key]]
  boot <- bootstrap_seuil(var, psi0)
  ci <- quantile(boot, c(0.025, 0.975), na.rm = TRUE)
  data.frame(

```



```

    Proxy    = key,
    Seuil     = round(psi0, 3),
    IC_low    = round(ci[1], 3),
    IC_high   = round(ci[2], 3),
    stringsAsFactors = FALSE
  )
}))

# 9) Affichage des résultats bootstrap
print(ci_table)

```

1.7 Modèle “piecewise” en FE-OLS (seuil fixe à 0,26)

L’objectif est de comparer directement les pentes avant/après rupture dans un cadre à effets fixes, et voir si l’ajout de cette rupture améliore AIC/BIC par rapport à la spline et au quadratique.

```

# 1) Construction des variables splitées
th <- 0.26
mod1_df2 <- mod1_df2 %>%
  mutate(
    gps_below = pmin(GreenPotSharew, th),
    gps_above = pmax(GreenPotSharew - th, 0)
  )

# 2) Estimation FE-OLS à seuil ("piecewise")
mod_piecewise <- feols(
  DeltaY ~ deltaLnK + DeltaLnL
    + gps_below + gps_above
    + self_emp_pct + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year,
  data      = mod1_df2,
  cluster   = ~iso3c
)

summary(mod_piecewise)

```

Tableau A.11 – Estimation FE-OLS « piecewise » de ΔY (seuil à 0,26)

Variable	Estimate	Std. Error	p-value
$\Delta \ln K$	0.129	(0.033)	< 0.001***
$\Delta \ln L$	0.047	(0.115)	0.686
gps_below	0.083	(0.191)	0.668
gps_above	0.194	(0.185)	0.298
self_emp_pct	-0.002	(0.001)	0.227
ays	-0.004	(0.010)	0.719
gov_exp_pctgdp	-0.006	(0.002)	0.003**
open_pct	-0.0001	(0.000)	0.826
democracy	-0.027	(0.014)	0.060.
access_electricity	0.0003	(0.000)	0.592
hhi_va	0.468	(0.378)	0.220
Observations	414		
RMSE	0.035		
Adj. R²	0.691		
Within R²	0.287		

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Remarque : gps_below et gps_above correspondent aux effets estimés de GreenPotSharew avant et après le seuil.*

1- Effet de GreenPotSharew : Pour les pays dont GreenPotSharew $\leq 0,26$ (« gps_below »), la pente est estimée à 0,083 ($p = 0,668$), ce qui n'est pas statistiquement significatif. Au-delà du seuil de 0,26 (« gps_above »), la pente additionnelle est estimée à 0,194 ($p = 0,298$) ; là encore, cette différence n'est pas significative. Aucune preuve solide d'un changement de pente avant ou après le seuil de verdissement à 0,26 : l'effet marginal reste indistinct dans les deux segments. 2- Contrôles et comparaisons : L'élasticité produit-capital ($\Delta \ln K$) demeure robuste et fortement significative ($0,13$, $p < 0,001$). L'élasticité produit-travail ($\Delta \ln L$) n'est pas significative. Les dépenses publiques (% PIB) continuent de réduire significativement la croissance ($-0,0065$, $p < 0,01$). L'indice de démocratie montre une tendance négative marginale ($p = 0,06$). Les autres variables de contrôle n'apparaissent pas significatives.

En somme, le modèle segmenté au seuil estimé (0,26) n'offre pas de différence détectable dans l'effet de GreenPotSharew avant et après ce point de rupture. Il paraît donc plus judicieux de compléter par une approche spline ou de vérifier la sensibilité à des observations extrêmes avant de conclure à l'existence ou l'absence d'un effet non linéaire robuste.

1.8 Modèle à spline cubique sur GreenPotSharew (FE-OLS avec spline cubique sur GreenPotSharew (4 df internes))

Pour assouplir davantage la forme fonctionnelle de l'effet de verdissement, une régression FE-OLS est estimée en incluant une spline cubique (4 nœuds internes) sur la part de potentiel

vert. Cette spécification permet de capter des non-linéarités plus fines qu'un simple modèle « piecewise ».

```
library(dplyr)
library(fixest)
library(splines)
library(broom)      # broom tidiers pour fixest
library(stringr)

# 1) Estimation du modèle
mod_spline <- feols(
  DeltaY ~ deltaLnK
  + DeltaLnL
  + ns(GreenPotSharew, df = 4)      # spline cubique (4 df internes)
  + self_emp_pct
  + ays
  + gov_exp_pctgdp
  + open_pct
  + democracy
  + access_electricity
  + hhi_va
  | iso3c + year,
  data      = mod1_df2,
  cluster   = ~iso3c
)

# 2) Extraction via summary.fixest()
sm <- summary(mod_spline)
spline_coefs_matrix <- sm$coefstable[grepl("ns\\(GreenPotSharew",
  ↪ rownames(sm$coefstable)), ]
print(spline_coefs_matrix)

# 3) Extraction via broom::tidy() and dplyr
tidy_smod <- broom::tidy(mod_spline,
  vcov      = "cluster",
  conf.int  = FALSE)
spline_coefs_tibble <- tidy_smod %>%
  filter(str_detect(term, "ns\\(GreenPotSharew"))
print(spline_coefs_tibble)
```

Les coefficients des quatre segments de la spline cubique sur GreenPotSharew sont reportés ci-dessous. Aucune des bases spline n'est statistiquement significative (tous $p > 0.29$), ce qui

indique qu’aucune flexibilité supplémentaire n’améliore sensiblement le modèle par rapport à une spécification plus simple.

Tableau A.12 – Estimation FE-OLS avec spline cubique sur GreenPotSharew (4 df internes)

Terme	Estimateur	Erreur-type	t-stat	p-value
ns(GreenPotSharew, df = 4)1	0.003	(0.050)	0.068	0.9459
ns(GreenPotSharew, df = 4)2	0.051	(0.050)	1.009	0.3168
ns(GreenPotSharew, df = 4)3	-0.157	(0.180)	-0.872	0.3866
ns(GreenPotSharew, df = 4)4	0.080	(0.076)	1.052	0.2969

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Les quatre coefficients des bases de la spline cubique ne diffèrent pas statistiquement de zéro. Autrement dit, après avoir contrôlé pour les effets fixes pays et année ainsi que l’ensemble des covariables, il n’existe pas de non-linéarité significative de plus haut ordre que celle déjà capturée par les termes de degré deux (cf. Modèle quadratique). La flexibilité ajoutée par une spline à quatre degrés de liberté sur GreenPotSharew n’apporte ainsi pas de gain explicatif notable.

1.9 Vérification des observations extrêmes

Pour s’assurer qu’aucune observation atypique ne pèse de façon excessive sur les résultats, on calcule plusieurs diagnostics d’influence à partir du modèle quadratique (model_quad) et du modèle à seuil (“piecewise”). On retiendra notamment la distance de Cook, les valeurs de levier et les DFBETAs.

```
library(fixest)
library(broom)
library(dplyr)

# 1) lm "nu" pour diagnostics (sans FE)
lm_diag <- lm(
  DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew + I(GreenPotSharew^2) +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va,
  data = mod1_df2
)

# 2) broom::augment() fournit .cooksd, .hat, et .rownames
```

```

diag_aug <- augment(lm_diag)

# 3) Recréer iso3c à partir de .rownames
diag_aug <- diag_aug %>%
  mutate(
    obs_index = as.integer(.rownames),
    iso3c      = mod1_df2$iso3c[obs_index]
  )

# 4) Calcul des dfbetas
dfb_mat <- dfbetas(lm_diag)[, c("deltalnK", "I(GreenPotSharew^2)")]
colnames(dfb_mat) <- c("dfbeta_K", "dfbeta_GPS2")

# 5) Assemblage final
diag_df <- diag_aug %>%
  transmute(
    iso3c,
    cooks_d = .cooks_d,
    leverage = .hat
  ) %>%
  bind_cols(as_tibble(dfb_mat))

# 6-7) Seuils et filtre des points influents
n      <- nrow(diag_df)
p      <- length(coef(lm_diag))
th_c   <- 4 / n
th_lev <- 2 * p / n
th_df  <- 2 / sqrt(n)

extremes <- diag_df %>%
  filter(
    cooks_d > th_c |
    leverage > th_lev |
    abs(dfbeta_K) > th_df |
    abs(dfbeta_GPS2) > th_df
  ) %>%
  arrange(desc(cooks_d))

# 8) Affichage des 10 observations les plus influentes
extremes %>%
  dplyr::select(iso3c, cooks_d, leverage, dfbeta_K, dfbeta_GPS2) %>%
  slice_head(n = 10)

```

Les diagnostics d'influence mettent en évidence les dix observations (pays) les plus influentes selon la distance de Cook, le levier et les dfbetas pour $\Delta \ln K$ et $I(\text{GreenPotSharew}^2)$.

Tableau A.13 - Observations extrêmes

iso3c	cooks_d	leverage	dfbeta_K	dfbeta_GPS2
ETH	0.9493	0.1761	1.6733	1.3113
TLS	0.7084	0.0783	-1.1635	0.1179
TLS	0.1126	0.0874	0.5665	-0.3750
IND	0.0867	0.0394	0.4244	0.3294
CZE	0.0841	0.0495	0.7453	0.0403
BDI	0.0582	0.1350	0.2079	-0.5119
MDV	0.0509	0.0840	0.0981	0.0630
MNG	0.0453	0.0598	-0.5577	0.1335
NAM	0.0361	0.0696	-0.3560	-0.2512
TZA	0.0313	0.0403	0.2284	0.0130

Éthiopie (ETH) présente la plus forte distance de Cook (0.95) et un levier élevé (0.18), ainsi que des dfbetas particulièrement importants pour $\Delta \ln K$ (1.67) et $I(\text{GreenPotSharew}^2)$ (1.31) : ce pays exerce une influence disproportionnée sur l'estimation conjointe de ces deux variables. Timor-Leste (TLS) apparaît deux fois : la première observation (cook's D = 0.71) indique un impact majeur de $\Delta \ln K$ (dfbeta_K = -1.16), la seconde un effet modéré sur la courbure du verdissement (dfbeta_GPS2 = -0.38). Inde (IND) et République tchèque (CZE) ont des dfbetas élevés sur $\Delta \ln K$ (> 0.4), mais des distances de Cook plus faibles, suggérant qu'elles influencent surtout la pente linéaire du capital. Burundi (BDI) se distingue par un dfbeta_GPS2 négatif élevé (-0.51), signalant qu'il infléchit notablement la composante quadratique du verdissement.

Ces résultats justifient de tester la robustesse des estimations en excluant temporairement ces observations influentes et d'examiner de près les données pour ETH, TLS et BDI afin de vérifier qu'il ne s'agit pas d'erreurs de mesure ou d'anomalies structurelles. Dans la suite de l'analyse, il conviendra de reré-estimer le modèle après avoir retiré (ou contrôlé) ces observations pour s'assurer que les effets estimés de GreenPotSharew et de $\Delta \ln K$ restent stables.

1.10 Test de Chow à coupure

L'application du test de Chow à coupure permet de vérifier formellement si l'on observe une rupture structurelle dans la relation entre le verdissement et la croissance avant et après le seuil estimé ($\hat{\psi}$). En définissant un indicateur $D_{\text{high}} = \mathbf{1}\{\text{GreenPotSharew} > \hat{\psi}\}$, ce test évalue l'hypothèse nulle suivante : $H_0: \beta_j^{\text{below}} = \beta_j^{\text{above}} \quad \forall j$, c'est-à-dire l'égalité simultanée de toutes les pentes (coefficients) avant et après la coupure.

Contrairement à une simple comparaison visuelle des coefficients segmentés, le test de Chow procède via une statistique de type Wald, **cluster-robuste**, qui prend en compte la dépendance intra-pays. Il offre un critère rigoureux pour déterminer si l'introduction d'un point de rupture améliore significativement l'ajustement du modèle et justifie l'usage d'une spécification en **piecewise** plutôt qu'un modèle continûment paramétrique (quadratique, spline, etc.).

```
library(dplyr)
library(fixest)

# 1) Construction du dummy "haut/bas" selon le seuil estimé
threshold <- 0.25956
mod1_df2 <- mod1_df2 %>%
  mutate(D_high = as.integer(GreenPotSharew > threshold))

# 2) Estimation FE-OLS avec interactions
mod_chow <- feols(
  DeltaY ~
    (deltalnK + DeltaLnL + GreenPotSharew + I(GreenPotSharew^2) +
     self_emp_pct + ays + gov_exp_pctgdp +
     open_pct + democracy + access_electricity + hhi_va)
    * D_high |
  iso3c + year,
  data = mod1_df2,
  cluster = ~iso3c
)

# 3) Construction du vecteur des interactions à tester
rhs_vars <- c("deltalnK", "DeltaLnL", "GreenPotSharew", "I(GreenPotSharew^2)",
              "self_emp_pct", "ays", "gov_exp_pctgdp",
              "open_pct", "democracy", "access_electricity", "hhi_va")
# on veut tester: var:D_high = 0 pour chaque var
test_terms <- paste0(rhs_vars, ":D_high")

# 4) Wald test cluster-robuste
chow_wald <- wald(
  mod_chow,
  test_terms,
  cluster = "iso3c",
  tidy = TRUE
)

# 5) Résultats
```

```
print(mod_chow)
print(chow_wald)
```

Tableau A.12 – Estimation FE-OLS avec interactions et test de Chow (coupure à $\hat{\psi} = 0,2596$)

Variable	Estimateur	Erreur-type	t-stat	p-value
deltalnK	0.254	(0.085)	2.993	0.0039**
DeltaLnL	-0.085	(0.200)	-0.427	0.6709
GreenPotSharew	-0.982	(2.509)	-0.391	0.6968
I(GreenPotSharew ²)	2.256	(6.810)	0.331	0.7415
self_emp_pct	-0.003	(0.004)	-0.738	0.4631
ays	-0.001	(0.020)	-0.058	0.9538
gov_exp_pctgdp	-0.010	(0.003)	-3.888	0.0002***
open_pct	-0.000	(0.001)	-0.328	0.7440
democracy	0.000	(0.002)	—	—
access_electricity	0.000	(0.000)	—	—
hhi_va	0.000	(0.000)	—	—
Observations	414			
RMSE	0.031			
Adj. R²	0.756			
Within R²	0.459			

Test de Chow :

Wald-stat = 3.128 (df =10, df =309), p = 0.0008***

*Signif. codes: *** p < 0.001; ** p < 0.01; * p < 0.05*

Le test de Chow à coupure rejette l'hypothèse nulle ($p = 0,0008$), ce qui signifie qu'il existe au moins une différence significative entre les pentes estimées avant et après le seuil ($\hat{\psi}$). En particulier, l'effet du capital ($\Delta \ln K$) reste positif et hautement significatif ($p < 0,01$) dans les deux régimes, tandis que le coefficient associé aux dépenses publiques (gov_exp_pctgdp) demeure négatif et significatif ($p < 0,001$). Les autres variables explicatives n'affichent pas de variation de pente significative autour de la coupure, ce qui suggère que leur influence sur la croissance ne change pas de manière substantielle avant et après ce seuil de verdissement.

1.11 Test de robustesse en excluant les observations les plus influentes

Avant d'évaluer l'impact des valeurs extrêmes, nous ré-estimons la spécification quadratique sur l'échantillon complet puis en excluant trois pays très influents (Ethiopie, Timor-Leste, Burundi), afin de vérifier la stabilité des coefficients et des critères d'ajustement.


```

library(dplyr)
library(fixest)
library(modelsummary)

# 1. Définir la liste des pays influents à exclure
influents <- c("ETH", "TLS", "BDI")

# 2. Préparer deux jeux de données : full et sans influents
df_full      <- mod1_df2
df_trimmed   <- mod1_df2 %>% filter(!(iso3c %in% influents))

# 3. Ré-estimer le modèle quadratique sur chacun
mod_full <- feols(
  DeltaY ~ deltaInK + DeltaLnL +
    GreenPotSharew + I(GreenPotSharew^2) +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year,
  data      = df_full,
  cluster   = ~iso3c
)

mod_trim <- feols(
  DeltaY ~ deltaInK + DeltaLnL +
    GreenPotSharew + I(GreenPotSharew^2) +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year,
  data      = df_trimmed,
  cluster   = ~iso3c
)

# 4. Comparer les deux estimations côte à côte
modelsummary(
  list(`Jeu Complet` = mod_full, `Sans influents` = mod_trim),
  stars   = c('*' = .05, '**' = .01, '***' = .001),
  coef_map = c(
    "deltaInK"           = "Δ ln K",
    "DeltaLnL"           = "Δ ln L",
    "GreenPotSharew"      = "GreenPotSharew",
    "I(GreenPotSharew^2)" = "GreenPotSharew2",
    "self_emp_pct"        = "Self-emp %",
    "ays"                 = "Avg years schooling",

```

```

"gov_exp_pctgdp"      = "Gov exp (%GDP)",
"open_pct"           = "Open (%GDP)",
"democracy"          = "Democracy index",
"access_electricity" = "Access electricity",
"hhi_va"             = "HHI VA"
),
title = "Robustesse : exclusion des pays influents"
)

```

Variable	Jeu Complet	Sans influents
$\Delta \ln K$	0.129*** (0.034)	0.129*** (0.033)
$\Delta \ln L$	0.048 (0.115)	0.127 (0.090)
GreenPotSharew	0.222 (0.360)	-0.204 (0.188)
GreenPotSharew ²	-0.122 (0.374)	0.367 (0.323)
Self-emp %	-0.002 (0.001)	-0.002 (0.001)
Avg years schooling	-0.004 (0.010)	-0.004 (0.010)
Gov exp (% GDP)	-0.006** (0.002)	-0.004*** (0.001)
Open (% GDP)	-0.000 (0.000)	-0.000 (0.000)
Democracy index	-0.028 (0.014)	-0.025 (0.014)
Access electricity	0.000 (0.000)	0.000 (0.001)
HHI VA	0.474 (0.388)	0.200 (0.214)
Observations	414	407
R²	0.760	0.783
R² ajusté	0.691	0.722
Within R²	0.287	0.313
Within R² ajusté	0.263	0.289
AIC	-1406.6	-1572.1
BIC	-1032.2	-1211.3
RMSE	0.040	0.030
Std. errors	Clustered (iso3c)	Clustered (iso3c)
Effets fixes	iso3c, year	iso3c, year

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Cette comparaison montre que l'exclusion des trois pays les plus influents n'affecte que marginalement les coefficients clés ($\Delta \ln K$ reste stable) et améliore même légèrement les critères d'ajustement (R^2 Within augmente de 0,287 à 0,313, AIC et BIC diminuent). Les signes et significations des dépenses publiques et des autres variables demeurent cohérents, ce qui renforce la robustesse de la spécification quadratique.

Nous procédons alors à l'utilisation de B-splines cubiques sur la variable GreenPotSharew, ce qui permet de libérer l'estimation des contraintes d'une forme fonctionnelle pré-spécifiée

(quadratique, linéaire, piecewise fixe, etc.). En insérant plusieurs segments lisses aux quartiles de GreenPotSharew, on capte d'éventuelles non-linéarités plus fines et des points de retournement de la pente qui ne seraient pas visibles dans un modèle paramétrique simple. Ce test est donc particulièrement utile pour vérifier si l'effet marginal du verdissement sur la croissance varie de manière complexe selon le niveau de « potentiel vert » et pour éviter des artefacts liés à un choix arbitraire de spécification.

```
library(fixest)
library(splines)
library(dplyr)
library(ggplot2)
library(stringr)
library(knitr)
library(tibble)    # pour rownames_to_column()

# 1) Choix des nœuds internes aux quartiles
knots_gps <- quantile(mod1_df2$GreenPotSharew,
                      probs = c(0.25, 0.5, 0.75),
                      na.rm = TRUE)

# 2) Estimation FE-OLS avec B-splines cubiques (degree = 3)
mod_spline_flex <- feols(
  DeltaY ~ deltaLnK
  + DeltaLnL
  + bs(GreenPotSharew, knots = knots_gps, degree = 3)
  + self_emp_pct
  + ays
  + gov_exp_pctgdp
  + open_pct
  + democracy
  + access_electricity
  + hhi_va
  | iso3c + year,
  data = mod1_df2,
  cluster = ~iso3c
)

# 3) Obtenir la table cluster-robuste
sm <- summary(mod_spline_flex, cluster = "iso3c")
coef_table <- sm$coeftable

# 4) Sélectionner les termes de spline
spline_rows <- rownames(coef_table) %>% str_subset("^bs\\((GreenPotSharew)")
```

```

coef_spline <- coef_table[spline_rows, , drop = FALSE]

# 5) Transformer et afficher en Markdown
coef_spline %>%
  as.data.frame() %>%
  tibble::rownames_to_column("Terme spline") %>%
  dplyr::select(
    `Terme spline`,
    Estimate,
    `Std. Error`,
    `t value`,
    `Pr(>|t|)`
  ) %>%
  knitr::kable(
    digits      = 3,
    col.names = c("Terme spline", "Estimate", "Std. Error", "t value",
      ↪ "Pr(> t)"),
    caption    = "Coefficients des B-splines cubiques sur GreenPotSharew"
  )

```

Tableau A.14 – Coefficients des B-splines cubiques sur GreenPotSharew (4 nœuds internes aux quartiles)

Terme spline	Estimateur	Erreur-type	t-stat	p-value
bs1	0.095	(0.345)	0.275	0.784
bs2	-0.125	(0.216)	-0.577	0.566
bs3	0.110	(0.185)	0.593	0.555
bs4	0.031	(0.182)	0.172	0.864
bs5	0.129	(0.185)	0.697	0.489
bs6	0.021	(0.179)	0.120	0.905

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Les six coefficients des bases B-spline (bs1 à bs6) ne sont pas significativement différents de zéro (tous p-values $> 0,3$), ce qui suggère qu'aucune des composantes non linéaires introduites par ces nœuds internes n'apporte d'amélioration notable à l'ajustement du modèle. En d'autres termes, la relation entre GreenPotSharew et ΔY apparaît globalement linéaire ou, du moins, sans cassure locale détectable aux quartiles considérés. Pour confirmer visuellement cette absence de non-linéarité forte, on peut tracer la courbe estimée de l'effet marginal de GreenPotSharew (via `predict()`), qui devrait alors se rapprocher d'une droite.

2 Endogénéité et IV

Lorsque l'on suspecte que GreenPotSharew (et/ou d'autres variables de contrôle) est endogène (p. ex. corrélation avec l'erreur via des chocs simultanés ou de la mesure), on peut recourir à une approche en deux étapes (2SLS) dans l'espace des moindres carrés à effets fixes. On peut instrumenter la part potentielle de verdissement par sa valeur décalée d'une période (validité à condition d'absence d'auto-corrélation de l'erreur), ou par une variable exogène.

2.1 IV externes

2.1.1 Préparation des instruments solaires et forestiers

Nous allons utiliser les IV suivants :

1- Irradiation solaire : Les pays plus « ensoleillés » ont mécaniquement plus de potentiel de solaire photovoltaïque, ce qui pousse la part verte de l'emploi ; cette insolation est plausiblement exogène à votre ΔY annuel. On récupère les données dans “Global Photovoltaic Power Potential by Country” (<https://datacatalog.worldbank.org/search/dataset/0038379>). On a tout téléchargé et constitué le fichier “solargis_pvpotential.xlsx”. a. “atp_GHI” : Average theoretical potential (GHI, kWh/m²/day), long-term. Mesure purement climatique, stable dans le temps, exogène aux chocs de croissance, très corrélée à notre part d'emplois solaires donc potentiellement à GreenPotSharew. b. “pb_equ_ar” : Photovoltaic equivalent area (% of total area), long-term. Désigne la capacité foncière structurelle pour le solaire, également time-invariant et exogène. Complète le GHI en captant les contraintes de disponibilité du sol.

2- “forest_area” : disponible dans les données World Bank “Forest area (% of land area).” AG.LND.FRST.ZS”. Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.

```
library(readxl); library(dplyr)

# 1) Solargis PV potential (solaire)
solargis_df <- read_excel(
  "/Users/.../Chap3/solargis_pvpotential.xlsx"
) %>%
  transmute(
    iso3c      = iso3c,
    atp_GHI    = as.numeric(atp_GHI),      # kWh/m2/j
    pb_equ_ar  = as.numeric(pb_equ_ar)     # % surface équivalente PV
```

```

) %>%
distinct(iso3c, .keep_all = TRUE)

# 2) Forest area (WB WDI)
if (!requireNamespace("WDI", quietly = TRUE)) install.packages("WDI")
library(WDI)

forest_area_df <- WDI(
  country    = "all",
  indicator  = "AG.LND.FRST.ZS",
  start      = 1990,
  end        = 2023,
  extra      = TRUE
) %>%
filter(!is.na(iso3c), region != "Aggregates") %>%
transmute(
  iso3c      = iso3c,
  year       = year,
  forest_area = AG.LND.FRST.ZS
) %>%
arrange(iso3c, year)

```

2.1.2 Importation et transformation des prix de commodités

On importe ici “**brent**” : issu du *World Bank Commodity Price Data* (The Pink Sheet), variable “*Crude oil, Brent (\$/bbl)*”. On le prend en *Annual Prices (Real)*, plus précisément en dollars constants de 2010 (USD 2010). Référence « internationale » par excellence, très peu corrélée aux chocs domestiques de court terme, et hautement liquide.

Puis on le met en log-différences :

$$\Delta \ln(\text{Brent}_t) = \ln(\text{Brent}_t) - \ln(\text{Brent}_{t-1})$$

L’intérêt de passer en log-différences est de stationnariser. Beaucoup de séries de prix sont non-stationnaires (trend, volatilité), leurs premières différences de log sont souvent plus stables. C’est à peu près le taux de variation du prix d’une année sur l’autre. Également, on vise l’homogénéité d’unité : on compare des élasticités (effet en % de variation du prix sur % de variation de GreenPotSharew), ce qui est plus robuste que d’utiliser des niveaux en USD/baril.

Ici on a un instrument purement temporel (identique pour tous les pays) qui va se fixer sur l’année et s’associe à chaque pays-année de notre panel.

Nous gardons également des IV secondaires qui sont des matières premières critiques pour le déploiement massif des technologies “vertes” : 1) des métaux de réseaux et d’électronique : “copper” pour la conductivité électrique, câblage des réseaux, turbines éoliennes et “aluminium” pour les structure des panneaux solaires, transport léger ; 2) métaux de batteries comme “nickel” et “lead” plomb-acid batteries pour stockages, ou “tin” pour soudures ; 3) fertilisants comme “phosphate rock”, “dap”, “tsp”, “urea” pour soutenir l’agriculture durable, utiles à la résilience des filières bio ; 4) gaz naturel comme “natural_gas_index” comme combustible de transition moins carboné que le charbon.

```
library(readxl); library(dplyr)

commod_raw <- read_excel(
  "/Users/.../Chap3/commodities.xlsx", sheet = "data"
)
commod_dln <- commod_raw %>%
  mutate(across(-year, as.numeric)) %>%
  arrange(year) %>%
  mutate(across(
    c(brent, natural_gas_index,
      phosphate_rock, dap, tsp, urea,
      aluminium, copper, lead, tin, nickel),
    ~ log(.x) - log(lag(.x)),
    .names = "dln_{.col}"
  )) %>%
  filter(!is.na(dln_brent))
```

2.1.3 Chargement et pivot des financements IRENA

Nous ajoutons comme IV également l’évolution des financements dans des technologies vertes à savoir : “On-grid Solar photovoltaic”, “Off-grid Solar photovoltaic”, “Solar thermal energy”, “Concentrated solar power”, “Onshore wind energy”, “Offshore wind energy”, “Renewable hydropower”, “Pumped storage”, “Marine energy”, “Other primary solid biofuels n.e.s.”, “Renewable municipal waste”, “Liquid biofuels”, “Biodiesel”, “Biogasoline”, “Biogas”, “Geothermal energy”, “Multiple renewables”. On télécharge tout cela sur “IRENASTAT Online Data Query Tool” indicateur “Public Investments (2021 million USQ) by Country/area, Technology and Year”. Nous téléchargeons alors un fichier excel “Tab delimited with heading”. Nous faisons quelques retouches directement sur excel à savoir les noms des variables.

```
library(readr); library(dplyr); library(tidyr); library(countrycode)

fin_tech_raw <- read_delim(
```

```

"/Users/.../Chap3/financing_technology.csv",
delim = ";",
col_types = cols(
  country    = col_character(),
  technology = col_character(),
  year       = col_integer(),
  value      = col_double()
)
)

fin_tech_wide <- fin_tech_raw %>%
  mutate(
    iso3c = countrycode(country, "country.name", "iso3c")
  ) %>%
  filter(!is.na(iso3c)) %>%
  pivot_wider(
    id_cols    = c(iso3c, year),
    names_from = technology,
    values_from = value,
    values_fn  = sum,
    values_fill = 0
  ) %>%
  arrange(iso3c, year)

```

2.1.4 Hydrologie, littoral et rugosité

Pour les potentiel géographique nous ajoutons des données sur :

1- Les ressources en eau avec 'Internal renewable freshwater resources, total (billion cubic meters) - "ER.H2O.INTR.K3". C'est le volume brut d'eau douce renouvelable chaque année, qui détermine l'énergie que l'on pourrait théoriquement produire en hydro (débit × hauteur de chute). Ce stock est fixé par le climat et la géographie, il ne varie pas concordamment avec les fluctuations annuelles de croissance ΔY (donc bon instrument).

```

# 1) Installer et charger WDI
if (!requireNamespace("WDI", quietly = TRUE)) install.packages("WDI")
library(WDI)
library(dplyr)

# 2) Téléchargement de la série ER.H2O.INTR.K3 (billion m^3/an) pour tous les
    ↪ pays
water_raw <- WDI(

```



```

country   = "all",
indicator = "ER.H2O.INTR.K3",
start     = 1990,
end       = 2023,
extra     = TRUE
)

# 3) Nettoyage : on ne garde que les pays (pas les agrégats régionaux)
water_resources_df <- water_raw %>%
  filter(
    !is.na(iso3c),      # garde seulement les observations avec code ISO3
    region != "Aggregates"
  ) %>%
  transmute(
    iso3c      = iso3c,
    year       = year,
    water_resources = ER.H2O.INTR.K3 # volume d'eau douce renouvelable
    ↪ annuel
  ) %>%
  arrange(iso3c, year)

# 4) Vérification rapide
water_resources_df %>%
  group_by(year) %>%
  summarise(
    n_countries      = n(),
    mean_resources    = mean(water_resources, na.rm = TRUE)
  ) %>%
  print(n = 10)

```

2- Longueur de littoral par km² (proxy offshore), en prenant comme source Natural Earth coastline (10m). Ici, plus on a de côtes, plus son potentiel offshore (et marin) est grand, indépendamment de la croissance.

```

library(rnaturalearth)
library(sf)
library(dplyr)

sf::sf_use_s2(FALSE)

# 1) Chargement côte et pays
coast <- ne_coastline(scale = 10, returnclass = "sf")

```

```

world <- ne_countries(scale = 10, returnclass = "sf") %>%
  dplyr::select(iso3c = iso_a3, geometry)

# 2) Réparation des géométries invalides
coast <- st_make_valid(coast)
world <- st_make_valid(world)

# 3) Intersection côte × pays
coast_country <- st_intersection(coast, world)

# 4) Calcul de la longueur par pays (en km)
install.packages("lwgeom")      # installe le package C++ qui étend sf
library(lwgeom)                 # charge les fonctions géodésiques

# (après avoir fait st_make_valid() et st_intersection())
coast_length <- coast_country %>%
  mutate(len_km = as.numeric(st_length(geometry)) / 1000) %>%
  st_drop_geometry() %>%
  group_by(iso3c) %>%
  summarise(coast_km = sum(len_km, na.rm = TRUE))

# Corrige le code pays -99 par "FRA"
coast_length <- coast_length %>%
  mutate(iso3c = ifelse(iso3c == "-99", "FRA", iso3c))

```

3- Rugosité du relief (proxy vent terrestre). Les reliefs montagneux ou vallonnés génèrent des accélérations de vent, et restent invariants à court terme.

```

install.packages(c("geodata", "terra", "sf", "rnaturalearth", "exactextractr"))

library(geodata)      # pour elevation_global()
library(terra)        # pour les calculs raster
library(sf)           # pour les géométries vectorielles
library(rnaturalearth) # pour le shapefile monde
library(exactextractr) # pour résumer les rasters par polygone
library(dplyr)

# 1) Préparation d'un dossier de cache
cache_dir <- "/Users/.../Chap3/geodata_cache"
dir.create(cache_dir, recursive = TRUE, showWarnings = FALSE)

```

```

# 2) Téléchargement de DEM global SRTM 30 (~1 km)
dem <- elevation_global(
  res = 0.5,      # 0.5 arc-min  30
  path = cache_dir
)

# 3) Calcul de la pente (en degrés)
dem_slope <- terrain(
  dem,
  v    = "slope",
  unit = "degrees"
)

# 4) Définit la rugosité comme l'écart-type de la pente dans une fenêtre 3x3
rugosity_ras <- focal(
  dem_slope,
  w    = matrix(1, 3, 3),
  fun  = sd,
  na.rm = TRUE
)

# 5) Chargement de la couche "pays" Natural Earth et correction FRA
world_sf <- ne_countries(
  scale      = "medium",
  returnclass = "sf"
) %>%
  mutate(
    iso3c = if_else(
      iso_a3 %in% c("-99", ""),
      iso_a3_eh,
      iso_a3
    )
  ) %>%
  # on force l'utilisation de select() de dplyr
  dplyr::select(iso3c, geometry)

# 6) Reprojection des polygones pour matcher le CRS du raster
world_proj <- st_transform(world_sf, crs = crs(rugosity_ras))

# 7) Extraction de la rugosité moyenne par pays
means_rug <- exact_extract(
  rugosity_ras,
  world_proj,

```

```

    fun      = "mean",
    progress = FALSE
  )

# 8) Construction du tableau final iso3c × rugosité
rugosity_by_country <- world_proj %>%
  sf::st_drop_geometry() %>%
  dplyr::mutate(rug_mean = means_rug) %>%
  dplyr::select(iso3c, rug_mean)

# 9) Aperçu
print(rugosity_by_country)

# 10) Correction du code pays -99 par "FRA"
rugosity_by_country <- rugosity_by_country %>%
  mutate(iso3c = ifelse(iso3c == "-99", "FRA", iso3c))

```

2.1.5 Fusion de tous les IV dans le panel

```

library(dplyr)

# Fonction utilitaire pour joindre et nettoyer les suffixes .orig/.new
join_clean <- function(df, df2, by, vars) {
  suffixes <- c(".orig", ".new")
  df_joined <- dplyr::left_join(df, df2, by = by, suffix = suffixes)

  for (v in vars) {
    orig <- paste0(v, suffixes[1])
    neu  <- paste0(v, suffixes[2])
    if ( neu %in% names(df_joined) ) {
      df_joined <- df_joined %>%
        dplyr::mutate(!v := dplyr::coalesce(.data[[neu]], .data[[orig]]))
      ↪ %>%

      dplyr::select(-all_of(c(orig, neu)), dplyr::everything())
    }
  }

  df_joined
}

```

```
# Sauvegarde du jeu de données propre
library(readr)
write_csv(df1_iv, "/Users/.../Chap3/df1_iv.csv")
saveRDS(df1_iv, "/Users/.../Chap3/df1_iv.rds")
```

2.1.6 IV-FE par instruments externes géographiques et solaires

Nous allons agréger les instruments géographiques. Le `geo_index` est un « indice synthétique » construit pour concentrer en une seule variable la variation géographique et climatique de plusieurs instruments potentiels. Concrètement, il est calculé en trois étapes :

1- Standardisation (par la fonction “`scale()`”) : par défaut, `scale(x)` centre ($x - \text{moyenne}(x)$) et réduit ($\div \text{écart-type}(x)$) la variable. Chacune des six colonnes passées à `scale()` devient un Z-score. Chacune des variables instrumentales retenues est d’abord centrée-réduite, c’est-à-dire transformée en score Z :

$$Z_{v_i} = \frac{v_i - \bar{v}}{\text{sd}(v)}$$

où v_i est la valeur de la variable v pour l’observation i , \bar{v} sa moyenne sur l’échantillon, et $\text{sd}(v)$ son écart-type.

2 – Assemblage (par `cbind()` + `rowMeans()`) :

On regroupe ces Z-scores dans une “matrice” à une ligne par pays-année, les colonnes correspondant aux instruments :

$$X_i = (Z_{\text{atp_GHI}}, Z_{\text{pb_equ_ar}}, Z_{\text{water_resources}}, Z_{\text{forest_area}}, Z_{\text{coast_km}}, Z_{\text{rug_mean}})$$

Puis on calcule la moyenne de ces Z-scores pour chaque ligne en ignorant les valeurs manquantes (`na.rm = TRUE`) :

$$\text{geo_index}_i = \frac{1}{6} \sum_v Z_{v_i}$$

3- Le `geo_index` est donc un score composite où : > 0 indique que, globalement, la combinaison géographique et climatique du pays-année est au-dessus de la moyenne (potentiel plus favorable) ; < 0 indique un potentiel en-dessous de la moyenne. Cette construction permet donc d’agréger plusieurs dimensions géographiques, de garder une échelle commensurable (moyenne 0, écart-type 1) et d’éviter la colinéarité forte entre instruments en les combinant.

```

library(dplyr)
# Geo-index synthétique pour réduire la collinéarité
df1_iv <- df1_iv %>%
  mutate(
    geo_index = rowMeans(
      cbind(
        scale(atp_GHI),
        scale(pb_equ_ar),
        scale(water_resources),
        scale(forest_area),
        scale(coast_km),
        scale(rug_mean)
      ),
      na.rm = TRUE
    )
  )

# Si fixest n'est pas encore installé :
if (!requireNamespace("fixest", quietly = TRUE)) {
  install.packages("fixest")
}

# Exporte en CSV
if (!requireNamespace("readr", quietly = TRUE)) install.packages("readr")
library(readr)
write_csv(df1_iv,
  "/Users/.../Chap3/df1_iv_with_geo_index.csv")

# Sauvegarde en RDS
saveRDS(df1_iv,
  "/Users/.../Chap3/df1_iv_with_geo_index.rds")

```

On supprime ensuite les doublons :

```

library(dplyr)

# 1) Lire le dataframe existant
df1_iv <- readRDS("/Users/.../Chap3/df1_iv_with_geo_index.rds")

# 2) Vérification du nombre d'observations avant
cat("Avant suppression des doublons :", nrow(df1_iv), "lignes\n")

```

```
# 3) Suppression des doublons sur iso3c + year
df1_iv_clean <- df1_iv %>%
  distinct(iso3c, year, .keep_all = TRUE)

# 5. Vérification du nombre d'observations après
cat("Après suppression des doublons :", nrow(df1_iv_clean), "lignes\n")

# 6. Export du dataframe nettoyé en RDS
saveRDS(df1_iv_clean,
        "/Users/.../Chap3/df1_iv_clean.rds")

# Idem en CSV
if (!requireNamespace("readr", quietly = TRUE)) install.packages("readr")
library(readr)
write_csv(df1_iv_clean,
          "/Users/.../Chap3/df1_iv_clean.csv")
```

```
library(dplyr)
library(fixest)

# 1) Estimation d'IV-FE (TSLS) avec geo_index comme instrument de
↪ GreenPotSharew
mod_iv_geo <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year # effets fixes pays et année
  | GreenPotSharew ~ geo_index, # IV : geo_index
  data = df1_iv_clean,
  cluster = "iso3c" # clustering par pays
)

# 2) Résultats
summary(mod_iv_geo, iv = TRUE)
```

Ici, le test de Hausman de Wu confirme l'endogénéité de GreenPotSharew ($p < 0.001$), justifiant l'usage de l'IV.

Tableau A.15 - Résultats TSLS (2SLS)

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	4.709	2.176	0.0364 *

Variable	Estimateur	Erreur-type	p-value
$\Delta \ln K$	0.197	0.073	0.0097 **
$\Delta \ln L$	-0.448	0.291	0.1315
Self-emp %	0.0105	0.0077	0.1796
Avg years schooling	-0.0438	0.030	0.1522
Gov exp (% GDP)	0.0004	0.0042	0.9248
Open (% GDP)	0.0000	0.0006	1.0000
Democracy index	-0.066	0.029	0.0265 *
Access electricity	-0.0038.	0.0022	0.0887 .
HHI VA	0.780	0.892	0.3871
Observations	259		
RMSE	0.0567		
Adj. R²	0.0388		
Within R²	-1.6506		

F-test 1 étape (GreenPotSharew \sim geo_index): $F = 3.39$ (df=1, df=233), $p = 0.0669$ (faiblement supérieur à 10 %).

Wu-Hausman (endogénéité de GreenPotSharew): $\chi^2 = 14.0$ (df=1, df=191), $p = 0.00024$ ***.

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Le test de Hausman fortement significatif confirme l'endogénéité de la part de potentiel vert, justifiant l'approche IV. Cependant, la force de l'instrument est faible ($F = 3.4$, $p = 0.067$), ce qui peut entraîner un biais de faibles instruments. Au second stade, la part de potentiel vert instrumentée (fit_GreenPotSharew) affiche un effet positif et significatif (coefficient = 4.71, $p < 0.05$), suggérant qu'un accroissement exogène du potentiel vert stimule la croissance. Par ailleurs, le capital reste un déterminant robuste ($p < 0.01$), et l'indice de démocratie montre un effet négatif significatif ($p < 0.05$). Les autres contrôles ne sont pas significatifs une fois l'endogénéité corrigée.

```
library(dplyr)
library(fixest)

# 1) TSLS IV-FE avec clustering deux-axes (pays & année)
mod_iv_geo2 <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year                                # effets fixes pays et année
  | GreenPotSharew ~ geo_index,                 # IV : geo_index
  data     = df1_iv_clean,
  cluster  = c("iso3c", "year")                 # clustering pays + année
```



```

)
summary(mod_iv_geo2, iv = TRUE)

# 2) Test de placebo renforcé : résidus vs. geo_index
# 2.1) Estimation FE de référence (sans endogène)
mod_base_geo <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv_clean,
  cluster = c("iso3c", "year")
)

# 2.2) Construction des résidus
df_placebo_geo <- df1_iv_clean %>%
  mutate(
    fitted_base = predict(mod_base_geo, newdata = .),
    resid_base = DeltaY - fitted_base
  )

# 2.3) Régression des résidus sur l'instrument
mod_placebo_geo <- feols(
  resid_base ~ geo_index
  | iso3c + year,
  data = df_placebo_geo,
  cluster = c("iso3c", "year")
)
summary(mod_placebo_geo)

```

Tableau A.15 - Estimation TSLS (2SLS) avec geo_index comme instrument

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	4.709*	(2.167)	0.0462 *
$\Delta \ln K$	0.197**	(0.065)	0.0083 **
$\Delta \ln L$	-0.448	(0.383)	0.2608
Self-emp %	0.0105	(0.0071)	0.1592
Avg years schooling	-0.0438	(0.0320)	0.1909
Gov exp (% GDP)	0.0004	(0.0046)	0.9321
Open (% GDP)	-0.0000	(0.0005)	1.0000
Democracy index	-0.0662	(0.0411)	0.1281
Access electricity	-0.0038	(0.0024)	0.1271

Variable	Estimateur	Erreur-type	p-value
HHI VA	0.7802	(0.8066)	0.3488
Observations	259		
RMSE	0.0567		
Adj. R²	0.0388		
Within R²	-1.6506		

Tests de validité de l'instrument

Le F-test de première étape (`GreenPotSharew` `geo_index`) est

$$F_{(1,248)} = 3.608, \quad p = 0.0587$$

(le F est faiblement inférieur au seuil usuel de 10 %, ce qui signale un instrument potentiellement faible).

Le test de Wu–Hausman (endogénéité de `GreenPotSharew`) donne

$$\chi^2_{(1,191)} = 14.0, \quad p = 2.38 \times 10^{-4} (***)$$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

L'instrument `geo_index` présente donc une force limitée ($F < 10$), ce qui invite à la prudence : bien que le test de Hausman confirme l'endogénéité de `GreenPotSharew`, la faiblesse de l'instrument peut introduire un biais dans les estimations.

Tableau A.19 - Test de placebo (cluster : iso3c & year)

Variable	Estimateur	Erreur-type	t-stat	p-value
geo_index	0.2777	(0.190)	1.462	0.1644
Observations	259			
RMSE	0.0244			
Within R²	0.0149			

Signif. codes : . $p < 0.10$

Après absorption des effets fixes pays et année et clustering bilatéral, `geo_index` n'explique plus les résidus du modèle structurel ($p = 0.1644$), ce qui renforce sa validité exogène.

On procède à plusieurs tests pour améliorer notre variable instrumentale :

- 1) Un premier test en Analyse de Composantes Principales (ACP) sur nos 6 variables car ici la pondération est trop uniforme :

```

library(dplyr)
library(fixest)

# 1) Copie df1_iv_clean dans un nouveau dataframe
df1_iv_pca <- df1_iv_clean

# 2) Définit les variables pour l'ACP
vars <- c("atp_GHI", "pb_equ_ar", "water_resources",
          "forest_area", "coast_km", "rug_mean")

# 3) Identifie les lignes complètes sur ces variables
complete_idx <- complete.cases(df1_iv_pca[, vars])

# 4) Calcul de l'ACP sur ces observations
pca_matrix <- scale(df1_iv_pca[complete_idx, vars], center = TRUE, scale =
  ↪ TRUE)
pca_res <- prcomp(pca_matrix)

# 5) Extraire la 1re composante et la rattacher à df1_iv_pca
df1_iv_pca <- df1_iv_pca %>% mutate(pca1 = NA_real_)
df1_iv_pca$pca1[complete_idx] <- pca_res$x[, 1]

# 6) Corrélation brute
cor.test(df1_iv_pca$pca1, df1_iv_pca$GreenPotSharew, use = "complete.obs")

# 7) Estimation IV-FE (TSLS) avec clustering deux-axes (pays & année)
mod_iv_pca2 <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year # effets fixes pays & année
  | GreenPotSharew ~ pca1, # IV : première CP
  data = df1_iv_pca,
  cluster = c("iso3c", "year") # clustering pays + année
)
summary(mod_iv_pca2, iv = TRUE)

# 8) Test de placebo renforcé
# 8.1) Estimation FE de référence
mod_base_pca <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv_pca,

```

```

  cluster = c("iso3c", "year")
)

# 8.2) Construction des résidus
df_placebo_pca <- df1_iv_pca %>%
  mutate(
    fitted_base = predict(mod_base_pca, newdata = .),
    resid_base = DeltaY - fitted_base
  )

# 8.3) Régression des résidus sur pca1
mod_placebo_pca <- feols(
  resid_base ~ pca1
  | iso3c + year,
  data = df_placebo_pca,
  cluster = c("iso3c", "year")
)
summary(mod_placebo_pca)

```

On réalise une ACP sur les six composantes du `geo_index` pour obtenir un instrument plus fortement corrélé à `GreenPotSharew`. La première composante principale (`pca1`) capture une part substantielle de la variance conjointe des variables géo-climatiques.

Test de corrélation brute

$\text{cor}(\text{pca1}, \text{GreenPotSharew}) = 0.3392$, $t_{(429)} = 7.468$, $p = 4.59 \times 10^{-13}$, 95% CI = [0.2528; 0.4202].

Tableau A.16 - Estimation TSLS avec `pca1` comme instrument

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	12.586	(32.860)	0.7071
$\Delta \ln K$	0.244	(0.296)	0.4229
$\Delta \ln L$	-1.462	(4.402)	0.7443
Self-emp %	0.0298	(0.0809)	0.7179
Avg years schooling	-0.0667	(0.1582)	0.6792
Gov exp (% GDP)	-0.0008	(0.0077)	0.9186
Open (% GDP)	-0.0009	(0.0030)	0.7784
Democracy index	-0.0983	(0.1534)	0.5314
Access electricity	-0.0095	(0.0259)	0.7203
HHI VA	1.601	(4.690)	0.7376

Variable	Estimateur	Erreur-type	p-value
Observations	253		
RMSE	0.1448		
Adj. R²	-6.7895		
Within R²	-25.200		

Tests de validité de l'instrument

Le F-test de première étape (GreenPotSharew pca1) est

$$F_{(1,242)} = 0.1100, \quad p = 0.7405$$

indiquant un instrument trop faible.

Le test de Wu-Hausman (endogénéité) donne

$$\chi^2_{(1,188)} = 4.2104, \quad p = 0.0416 (*)$$

Bien que `pca1` soit corrélée à `GreenPotSharew`, le F-test révèle une faiblesse extrême de l'instrument, rendant l'estimation IV peu fiable.

2) On test alors avec plusieurs IV geo-climatiques séparés :

```
library(dplyr)
library(fixest)

# 1) TSLS IV-FE avec plusieurs instruments géo-climatiques
mod_iv_sep2 <- feols(
  DeltaY ~
    deltaInK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year                                # effets fixes pays et année
  | GreenPotSharew ~
    atp_GHI + water_resources + forest_area + coast_km + rug_mean,
  data      = df1_iv_clean,
  cluster   = c("iso3c","year")                # clustering bilatéral
)
summary(mod_iv_sep2, iv = TRUE)

# 2) Test de placebo renforcé
# 2.1) FE de référence (sans endogène)
mod_base_sep <- feols(
```

```

DeltaY ~
  deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
  open_pct + democracy + access_electricity + hhi_va
| iso3c + year,
data      = df1_iv_clean,
cluster   = c("iso3c","year")
)

# 2.2) Création des résidus
df_placebo_sep <- df1_iv_clean %>%
  mutate(
    fitted_base = predict(mod_base_sep, newdata = .),
    resid_base  = DeltaY - fitted_base
  )

# 2.3) Régression des résidus sur les IV
mod_placebo_sep <- feols(
  resid_base ~
    atp_GHI + water_resources + forest_area + coast_km + rug_mean
| iso3c + year,
data      = df_placebo_sep,
cluster   = c("iso3c","year")
)
summary(mod_placebo_sep)

```

Pour renforcer l'instrumentation de GreenPotSharew, nous utilisons simultanément cinq variables géo-climatiques comme IV. Du fait de colinéarités, seuls `water_resources` et `forest_area` restent actifs dans l'estimation, les autres (`atp_GHI`, `coast_km`, `rug_mean`) étant éliminés.

Tableau A.17 – Résultats TSLS avec IV géo-climatiques séparés

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	4.679*	(2.113)	0.0427 *
$\Delta \ln K$	0.196**	(0.064)	0.0076 **
$\Delta \ln L$	-0.444	(0.377)	0.2571
Self-emp %	0.0104	(0.0070)	0.1550
Avg years schooling	-0.0436	(0.0316)	0.1875
Gov exp (% GDP)	0.0004	(0.0045)	0.9383
Open (% GDP)	0.0000	(0.0005)	0.9962
Democracy index	-0.0659	(0.0408)	0.1273
Access electricity	-0.0038	(0.0023)	0.1221
HHI VA	0.7699	(0.7880)	0.3441

Variable	Estimateur	Erreur-type	p-value
Observations	259		
RMSE	0.0563		
Adj. R²	0.0502		
Within R²	-1.6192		

Tests de validité

Le F-test de première étape donne

$$F_{(2,247)} = 1.8586, \quad p = 0.1581$$

indiquant des instruments trop faibles ($F < 10$).

Le test de Wu–Hausman (endogénéité) fournit

$$\chi^2_{(1,191)} = 14.3, \quad p = 2.06 \times 10^{-4} (***)$$

Le test de Sargan (sur-identification) donne

$$\chi^2_{(4)} \approx 1.46 \times 10^{-9}, \quad p = 1$$

Aucune contradiction de validité n'est détectée, mais la faiblesse des variables instrumentales limite toujours la fiabilité de l'estimation.

3) On introduit donc un IV à composante temporelle :

Un bon candidat pour injecter de la variation dans le temps tout en restant exogène aux chocs domestiques de court terme semble être le taux de variation du prix du baril de Brent, déjà calculé dans `commod_dln$dln_brent`. Elle propose une variation temporelle forte et assurément non corrélée aux chocs de croissance spécifiques à un pays (c'est un marché liquide, mondial). Principe d'exogénéité respecté ici : un choc pétrolier global n'étant pas déterminé par la croissance d'un petit pays. Semble également représentatif, car le prix du pétrole conditionne directement l'incitation à investir dans les renouvelables, donc impactant la part verte de l'emploi.

```
if (!requireNamespace("readxl", quietly = TRUE)) install.packages("readxl")
if (!requireNamespace("dplyr", quietly = TRUE)) install.packages("dplyr")
if (!requireNamespace("fixest", quietly = TRUE)) install.packages("fixest")

library(readxl)
library(dplyr)
```

```

library(fixest)

# 1) Importe la feuille "data" du fichier commodities.xlsx
commod_raw <- readxl::read_excel(
  "/Users/.../Chap3/commodities.xlsx",
  sheet = "data"
)

# 2) Calcule les log-différences de Brent
commod_dln <- commod_raw %>%
  dplyr::arrange(year) %>%
  dplyr::select(year, brent) %>%
  dplyr::mutate(
    dln_brent = log(brent) - log(dplyr::lag(brent))
  ) %>%
  dplyr::filter(!is.na(dln_brent))

# 3) Rattache dln_brent à df1_iv_clean via match()
df1_iv_pbr <- df1_iv_clean %>%
  mutate(dln_brent = commod_dln$dln_brent[match(year, commod_dln$year)])

# 4) Construction de geo_index2 en incluant dln_brent
df1_iv_pbr <- df1_iv_pbr %>%
  mutate(geo_index2 = rowMeans(
    cbind(
      scale(atp_GHI),
      scale(pb_equ_ar),
      scale(water_resources),
      scale(forest_area),
      scale(coast_km),
      scale(rug_mean),
      scale(dln_brent)
    ),
    na.rm = TRUE
  ))

# 5) Estimation de l'IV-FE (TSLS) en double-cluster : pays & année
mod_iv_brent2 <- feols(
  DeltaY ~ deltaInK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year # effets fixes pays et année
  | GreenPotSharew ~ geo_index2, # IV : geo_index2
  data = df1_iv_pbr,

```



```

cluster = c("iso3c", "year")          # clustering à deux axes
)

# 6) Résultats IV
summary(mod_iv_brent2, iv = TRUE)

```

Dans cette extension, nous incorporons la variation du prix du pétrole Brent comme septième composante de l'instrument géo-climatique. L'idée est que les chocs exogènes sur les prix internationaux de l'énergie peuvent également affecter, de manière indépendante des dynamiques économiques nationales à court terme, la propension d'un pays à développer des EFPV. Nous construisons donc un nouvel indice `geo_index2` en moyennant les six variables géo-climatiques standardisées et la première différence logarithmique du Brent.

Tableau A.18 - Résultats TSLS (Clustering iso3c & year) avec `geo_index2` comme instrument

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	3.458	(0.949)	0.0024 **
$\Delta \ln K$	0.164	(0.038)	0.0006 ***
$\Delta \ln L$	-0.306	(0.216)	0.1772
Self-emp %	0.0073	(0.0042)	0.1023
Avg years schooling	-0.0346	(0.0190)	0.0885 .
Gov exp (% GDP)	-0.0014	(0.0025)	0.5865
Open (% GDP)	0.0001	(0.0005)	0.8168
Democracy index	-0.0564	(0.0323)	0.1016
Access electricity	-0.0025	(0.0014)	0.0919 .
HHI VA	0.3536	(0.3736)	0.3589
Observations	259		
RMSE	0.0432		
Adj. R²	0.4416		
Within R²	-0.5398		

Tests de validité de l'instrument

Le F-test de première étape (`GreenPotSharew` `geo_index2`) est

$$F_{(1,248)} = 12.6, \quad p = 4.63 \times 10^{-4} (***)$$

Le test de Wu-Hausman (endogénéité de `GreenPotSharew`) donne

$$\chi^2_{(1,191)} = 25.4, \quad p = 1.064 \times 10^{-6} (***)$$

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$; . $p < 0,10$

L'ajout du choc pétrolier renforce la première étape ($F > 10$), garantissant un instrument solide. Au second stade, l'effet instrumenté de `GreenPotSharew` reste positif et significatif, confirmant qu'un accroissement exogène du potentiel vert stimule la croissance.

L'instrument avec `dld_brent` est maintenant un instrument suffisamment fort et valide. Pour vérifier s'il est correct, on teste des placebo en vérifiant que `geo_index2` n'explique pas d'autres variables dépendantes dans un modèle sans `GreenPotSharew`.

```
library(dplyr)
library(fixest)

# 1) Estimation FE « baseline » sans instrument (pays + année)
mod_base_pf <- feols(
  DeltaY ~
    deltaLnK + DeltaLnL + self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv_pbr,
  cluster = c("iso3c", "year")
)

# 2) Calcul des résidus de  $\Delta Y$ 
df_placebo2 <- df1_iv_pbr %>%
  mutate(
    fitted_base = predict(mod_base_pf, newdata = .),
    resid_base = DeltaY - fitted_base
  )

# 3) Régression des résidus sur l'instrument geo_index2
mod_placebo2 <- feols(
  resid_base ~ geo_index2
  | iso3c + year,
  data = df_placebo2,
  cluster = c("iso3c", "year")
)

# 4) Résultats
summary(mod_placebo2)
```

Tableau A.19 - Test de placebo : résidus vs `geo_index2`

Variable	Estimateur	Erreur-type	t-stat	p-value
geo_index2	0.539	(0.314)	1.715	0.1069
Observations	259			
RMSE	0.0240			
Within R²	0.0430			

Signif. codes : . p<0.10

Le coefficient de geo_index2 n'est pas significatif ($p=0.1069$), ce qui indique qu'il n'explique pas les résidus du modèle de référence une fois absorbés les effets fixes pays et année. Cette absence de corrélation renforce la validité exogène de l'instrument.

Cette variable instrumentale geo_index2 semble remplir trois conditions clés pour qu'un instrument soit jugé acceptable : 1) Pertinence (F-test de la première étape bien au-delà du seuil critique de 10), 2) Exogénéité (n'explique par les chocs non observés de ΔY , ce qui renforce l'hypothèse selon laquelle il n'affecte la croissance que via GreenPotSharew), 3) Endogénéité avérée de la variable appliquée (test de Wu-Hausman rejette l'exogénéité de GreenPotSharew, justifiant pleinement l'usage d'une IV plutôt que d'une OLS simple).

```
# Sauvegarde
chemin_rds <- "/Users/.../Chap3/df1_iv_pbr.rds"
saveRDS(df1_iv_pbr, file = chemin_rds)
```

2.2 IV-FE par lags internes de GreenPotSharew

On met en œuvre une estimation IV-FE (Two-Stage Least Squares à effets fixes) en utilisant les deux premiers décalages temporels de la part de potentiel vert (GreenPotSharew) comme instruments internes. L'idée est de résoudre l'éventuelle endogénéité de GreenPotSharew (corrélation avec le terme d'erreur) en s'appuyant sur sa dynamique passée, supposée corrélée à la valeur courante mais exogène aux chocs contemporains de croissance. En procédant ainsi, on reste dans le cadre du panel pays-année sans faire appel à des variables externes, tout en exploitant la structure temporelle des données.

```
library(dplyr)
library(fixest)

# 1) Part du jeu nettoyé df1_iv_clean, créer un objet panel laggué
df1_iv_lags <- df1_iv_clean %>%
  group_by(iso3c) %>%
  arrange(year) %>%
  mutate(
```

```

    GPS_lag1 = lag(GreenPotSharew, 1),
    GPS_lag2 = lag(GreenPotSharew, 2)
  ) %>%
ungroup() %>%
# Ne garder que les observations où les deux lags existent
filter(!is.na(GPS_lag1) & !is.na(GPS_lag2))

# 2) Estimation du TSLS à effets fixes via deux lags comme IV
mod_iv_lag <- feols(
  # 2SLS : on n'inclut pas GreenPotSharew dans la liste des RHS
  DeltaY ~ deltaLnK
    + DeltaLnL
    + self_emp_pct
    + ays
    + gov_exp_pctgdp
    + open_pct
    + democracy
    + access_electricity
    + hhi_va
  | iso3c + year
  | GreenPotSharew ~ GPS_lag1 + GPS_lag2,
  data = df1_iv_lags,
  cluster = "iso3c"
)

# 3) Résultats (1 étape et test de Hausman)
summary(mod_iv_lag, iv = TRUE)

```

Tableau A.20 – Estimation TSLS (IV = lags internes de GreenPotSharew)

Variable	Estimateur	Erreur-type	p-value
fit_GreenPotSharew	1.534	(1.033)	0.1468
$\Delta \ln K$	0.112	(0.039)	0.0067 **
$\Delta \ln L$	-0.101	(0.195)	0.6094
Self-emp %	0.001	(0.004)	0.7131
Avg years schooling	-0.017	(0.014)	0.2295
Gov exp (% GDP)	-0.004	(0.002)	0.0826 .
Open (% GDP)	0.000	(0.000)	0.6105
Democracy index	-0.038	(0.016)	0.0198 *
Access electricity	-0.001	(0.001)	0.6211
HHI VA	-0.181	(0.575)	0.7553

Variable	Estimateur	Erreur-type	p-value
Observations	223		
RMSE	0.0260		
Adj. R²	0.7620		
Within R²	0.4623		

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Diagnostics de validité des instruments

- F-test de première étape (GreenPotSharew ~ GPS_lag1 + GPS_lag2) :

$$F_{(2,197)} = 5.700, \quad p = 0.0039 (**)$$

- Wu–Hausman (endogénéité de GreenPotSharew) :

$$\chi^2_{(1,164)} = 2.719, \quad p = 0.1011$$

- Sargan (sur-identification) :

$$\chi^2_{(1)} = 6.935, \quad p = 0.0085$$

*** $p < 0.01$; * $p < 0.05$*

Ici l'effet instrumenté de GreenPotSharew reste positif (1.53 environ) mais non significatif (0.15 environ). En somme, si les lags internes offrent un premier-plan prometteur pour instrumenter la part verte, les tests de suridentification et la non-significativité de l'effet principal recommandent de compléter l'analyse avec d'autres instruments ou spécifications pour valider pleinement l'approche IV.

L'objet de ce code ci-après est d'exploiter les lags internes de la variable potentiellement endogène (GreenPotSharew) comme instruments, tout en ne conservant qu'un effet fixe année et en clusterisant par pays. En groupant par pays et en calculant deux décalages successifs (GPS_lag1 et GPS_lag2), on crée des sources de variation plausiblement exogènes (dépendant du passé) pour identifier causalement l'impact de GreenPotSharew sur ΔY .

```
library(fixest)
library(dplyr)

# 1) Construction des lags de GreenPotSharew à l'intérieur du panel
df1_iv2 <- df1_iv_clean %>%
  group_by(iso3c) %>%
```

```

arrange(year) %>%
mutate(
  GPS_lag1 = lag(GreenPotSharew, 1),
  GPS_lag2 = lag(GreenPotSharew, 2)
) %>%
ungroup() %>%

filter(!is.na(GPS_lag1), !is.na(GPS_lag2))

# 2) Estimation du TSLS avec un unique effet fixe "year"
mod_iv_yearFE <- feols(
  DeltaY ~ deltaLnK
  + DeltaLnL
  + self_emp_pct
  + ays
  + gov_exp_pctgdp
  + open_pct
  + democracy
  + access_electricity
  + hhi_va
  | year
  | GreenPotSharew ~ GPS_lag1 + GPS_lag2,
  data = df1_iv2,
  cluster = ~iso3c
)

# 3) Résultats (F-test 1 étape et test de Hausman)
summary(mod_iv_yearFE, iv = TRUE)

```

Tableau A.21 - Estimation TSLS avec lags internes de GreenPotSharew comme instruments et seul effet fixe année

Variable	Estimateur	Erreur-type	t-stat	p-value
fit_GreenPotSharew	-0.10894	(0.1422)	-0.7661	0.4490
$\Delta \ln K$	0.11933	(0.0758)	1.5740	0.1250
$\Delta \ln L$	0.09248	(0.1432)	0.6458	0.5228
Self-emp %	0.00094	(0.0005)	2.0747	0.0459 *
Avg years schooling	0.00162	(0.0021)	0.7739	0.4445
Gov exp (% GDP)	-0.00428	(0.0024)	-1.7902	0.0826 .
Open (% GDP)	0.00017	(0.0003)	0.5142	0.6105
Democracy index	-0.03835	(0.0157)	-2.4492	0.0198 *
Access electricity	-0.00066	(0.0013)	-0.4990	0.6211

Variable	Estimateur	Erreur-type	t-stat	p-value
HHI VA	-0.18075	(0.5751)	-0.3143	0.7553
Observations	223			
Fixed-effects	year only			
RMSE	0.04341			
Adj. R ²	0.44774			
Within R ²	0.31316			

Tests de validité des instruments

- F-test de première étape (GPS_lag1 + GPS_lag2) :

$$F_{(2,197)} = 275.10, \quad p < 2.2 \times 10^{-16} (***)$$

- Wu-Hausman (endogénéité de GreenPotSharew) :

$$\chi^2_{(1,197)} = 54.9, \quad p = 3.58 \times 10^{-12} (***)$$

- Sargan (sur-identification) :

$$\chi^2_{(1)} = 0.045, \quad p = 0.832$$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$

Au second stade, l'effet instrumenté de GreenPotSharew est estimé à -0.109 (SE=0.142, p = 0.45) : non significatif et même légèrement négatif. Parmi les contrôles, seul le taux d'auto-emploi (self_emp_pct) apparaît positivement associé à ΔY (coef = 0.00094, p = 0.046). Les élasticités en capital et travail, ainsi que les autres variables institutionnelles et structurelles, ne sont pas statistiquement significatives dans cette spécification.

Dans l'ensemble, cette variante IV basée sur des lags internes confirme que l'endogénéité de GreenPotSharew peut être traitée, mais ne donne pas de preuve robuste d'un effet direct du verdissement passé sur la croissance actuelle, tout en mettant en évidence un rôle mineur de l'auto-emploi.

3 Hétérogénéité Advanced vs Developing

3.1 Modèle linéaire

Cette étape vise à tester si l'effet du verdissement varie selon le niveau de développement du pays. Pour cela, l'échantillon est scindé en deux sous-ensembles (« advanced » et « developing ») et les modèles de panel à effets fixes sont estimés séparément.

```
library(dplyr)
library(fixest)
library(modelsummary)

# 1) Préparer le data.frame avec econ_type
mod1_df2 <- df1_mod1 %>%
  mutate(
    econ_type = factor(
      ifelse(income_group %in% c("HIGH_INCOME", "UPPER_MIDDLE"),
            "Advanced", "Developing"),
      levels = c("Advanced", "Developing")
    )
  )

# 2) Calculer la variance intra-groupe de DeltaY
var_group <- mod1_df2 %>%
  group_by(econ_type) %>%
  summarise(var_DeltaY = var(DeltaY, na.rm = TRUE), .groups = "drop")

# 3) Identifier les groupes avec variance > 0
valid_types <- var_group %>%
  filter(var_DeltaY > 0) %>%
  pull(econ_type)

# 4) Pour chaque groupe valide, estimer le modèle FE-OLS
models <- list()
formula_mod1 <- DeltaY ~
  deltaInK + DeltaLnL + GreenPotSharew + dge +
  self_emp_pct + ays + gov_exp_pctgdp +
  open_pct + democracy + access_electricity + hhi_va |
  iso3c + year

if ("Advanced" %in% valid_types) {
  models[["Advanced"]] <- feols(
    formula_mod1,
```



```

    data      = filter(mod1_df2, econ_type == "Advanced"),
    cluster   = "iso3c"
  )
} else {
  message("Le groupe 'Advanced' a une variance nulle de  $\Delta Y$  et a été exclu.")
}
if ("Developing" %in% valid_types) {
  models[["Developing"]] <- feols(
    formula_mod1,
    data      = filter(mod1_df2, econ_type == "Developing"),
    cluster   = "iso3c"
  )
} else {
  message("Le groupe 'Developing' a une variance nulle de  $\Delta Y$  et a été
    ↪ exclu.")
}

# 5) Présenter les résultats
if (length(models) > 0) {
  modelsummary(
    models,
    statistic = "p.value",
    stars      = c('***' = 0.001, '**' = 0.01, '*' = 0.05),
    coef_map   = c(
      "deltalnK"      = " $\Delta \ln K$ ",
      "DeltaLnL"      = " $\Delta \ln L$ ",
      "GreenPotSharew" = "GreenPotSharew",
      "dge"           = "DGE index",
      "self_emp_pct"  = "Self-emp %",
      "ays"           = "Avg years schooling",
      "gov_exp_pctgdp" = "Gov exp (% GDP)",
      "open_pct"      = "Open (% GDP)",
      "democracy"     = "Democracy index",
      "access_electricity" = "Access electricity",
      "hhi_va"        = "HHI VA"
    ),
    output = "markdown",
    title  = "Tableau A.21 - FE-OLS par type d'économie"
  )
} else {
  message("Aucun groupe n'a pu être estimé : variance nulle pour tous les
    ↪ sous-échantillons.")
}

```

Le groupe 'Advanced' a une variance nulle de ΔY et a été exclu. Dans notre échantillon de pays en développement, la croissance de la production est principalement tirée par l'investissement en capital, tandis que les dépenses publiques et la qualité démocratique semblent peser négativement à court terme. En revanche, l'effet du verdissement de l'emploi (GreenPotSharew) n'apparaît pas comme un moteur de la croissance dans ces pays.

3.2 Non-linéarité du verdissement dans chaque sous-échantillon

Pour explorer un effet quadratique de GreenPotSharew séparément dans les économies « Advanced » et « Developing », on estime deux modèles FE-OLS qui incluent à la fois la variable linéaire et son carré :

```
library(dplyr)
library(fixest)
library(modelsummary)

# 1) Reconstruire mod1_df2 avec econ_type si besoin
mod1_df2 <- df1_mod1 %>%
  mutate(
    econ_type = factor(
      ifelse(income_group %in% c("HIGH_INCOME", "UPPER_MIDDLE"),
            "Advanced", "Developing"),
      levels = c("Advanced", "Developing")
    )
  )

# 2) Identifier les groupes pour lesquels  $\Delta Y$  n'est pas constant
valid_groups <- mod1_df2 %>%
  group_by(econ_type) %>%
  summarise(var_DeltaY = var(DeltaY, na.rm = TRUE)) %>%
  filter(var_DeltaY > 0) %>%
  pull(econ_type)

# 3) Pour chaque groupe valide, estimer le modèle quadratique
models <- list()
for(g in valid_groups) {
  df_sub <- filter(mod1_df2, econ_type == g)
  models[[g]] <- feols(
    DeltaY ~ deltaLnK + DeltaLnL
    + GreenPotSharew + I(GreenPotSharew^2)
  )
}
```

```

        + dge
        + self_emp_pct + ays + gov_exp_pctgdp
        + open_pct + democracy + access_electricity
        + hhi_va
    | iso3c + year,
    data      = df_sub,
    cluster   = "iso3c"
  )
}

# 4) Afficher côte à côte
modelsummary(
  models,
  statistic = "p.value",
  stars      = c('***' = 0.001, '**' = 0.01, '*' = 0.05),
  coef_map  = c(
    "deltalnK"          = "Δ ln K",
    "DeltaLnL"          = "Δ ln L",
    "GreenPotSharew"    = "GreenPotSharew",
    "I(GreenPotSharew^2)" = "GreenPotSharew2",
    "dge"               = "DGE index",
    "self_emp_pct"      = "Self-emp %",
    "ays"               = "Avg years schooling",
    "gov_exp_pctgdp"    = "Gov exp (% GDP)",
    "open_pct"          = "Open (% GDP)",
    "democracy"         = "Democracy index",
    "access_electricity" = "Access electricity",
    "hhi_va"            = "HHI VA"
  ),
  output = "markdown",
  title  = "Modèle quadratique par type d'économie"
)

```

Dans les pays en développement, l'investissement en capital demeure le principal moteur de la croissance de ΔY . En revanche, la part de potentiel vert n'apparaît pas avoir un effet direct significatif, ni linéaire ni quadratique, sur la croissance, une fois pris en compte les autres déterminants. Les dépenses publiques et le degré de démocratie exercent en revanche des effets négatifs notables sur la croissance.

3.3 Effet marginal du verdissement (delta-method + graphiques)

Le calcul de l'effet marginal de la part potentielle verte (`GreenPotSharew`) sur la croissance de la productivité (ΔY) repose sur la dérivée partielle de la spécification quadratique, c'est-à-dire :

$$\Delta Y / \text{GreenPotSharew} = \beta_3 + 2 \times \beta_4 \times \text{GreenPotSharew}$$

où β_3 est le coefficient de `GreenPotSharew` et β_4 celui de $I(\text{GreenPotSharew}^2)$.

La méthode Delta permet d'obtenir un estimateur ponctuel ainsi que son intervalle de confiance à 95 %.

```
library(dplyr); library(fixest); library(ggplot2); library(tidyr)

mod1_df2 <- df1_mod1

# 1) Extraction de coefs et vcov pour model_quad
coefs_q <- coef(model_quad)[c("GreenPotSharew", "I(GreenPotSharew^2)")]
vcov_q <- vcov(model_quad)[c(3,10), c(3,10)]

# 2) Grille de GreenPotSharew
grille <- seq(0,1, by = 0.01)

# 3) Calcul ME et IC pour GreenPotSharew
df_me_g <- tibble(x = grille) %>%
  mutate(
    ME      = coefs_q[1] + 2*coefs_q[2]*x,
    SE      = sqrt(vcov_q[1,1] + 4*x^2*vcov_q[2,2] + 4*x*vcov_q[1,2]),
    CI_low  = ME - 1.96*SE,
    CI_high = ME + 1.96*SE,
    var     = "GreenPotSharew"
  )

# 4) Modèle 1 avec interactions ECON_TYPE et GreenPotSharew × informalité
mod_int <- feols(
  DeltaY ~
    deltaLnK + DeltaLnL + GreenPotSharew + dge +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy + access_electricity + hhi_va +
    GreenPotSharew:mimic + mimic +
    GreenPotSharew:dge + dge |
    iso3c + year,
```

```

data      = mod1_df2,
cluster = ~iso3c
)

# 5) Même routine pour mimic et dge à partir de mod_int
# on reprend coefs_i
coefs_i <- coef(mod_int)[c("mimic", "GreenPotSharew:mimic", "dge",
  ↪  "GreenPotSharew:dge")]

# on extrait vcov_i en utilisant les mêmes noms
vcov_i <- vcov(mod_int)[names(coefs_i), names(coefs_i)]

df_me_inf <- tibble(x = grille) %>%
  mutate(
    # effet marginal de MIMIC à x
    ME_mimic = coefs_i["mimic"] + coefs_i["GreenPotSharew:mimic"] * x,
    SE_mimic = sqrt(
      vcov_i["mimic","mimic"] +
      x^2 * vcov_i["GreenPotSharew:mimic","GreenPotSharew:mimic"] +
      2 * x * vcov_i["mimic","GreenPotSharew:mimic"]
    ),
    # effet marginal de DGE à x
    ME_dge   = coefs_i["dge"] + coefs_i["GreenPotSharew:dge"] * x,
    SE_dge   = sqrt(
      vcov_i["dge","dge"] +
      x^2 * vcov_i["GreenPotSharew:dge","GreenPotSharew:dge"] +
      2 * x * vcov_i["dge","GreenPotSharew:dge"]
    )
  ) %>%
  pivot_longer(
    cols      = -x,
    names_to   = c(".value", "var"),
    names_pattern = "(ME|SE)_(.*)"
  )

# 6) Graphiques
ggplot(df_me_g, aes(x, ME)) +
  geom_line() +
  geom_ribbon(aes(ymin=CI_low, ymax=CI_high), alpha=0.2) +
  labs(y="Effet marginal de GreenPotSharew")

```

```

ggplot(df_me_inf, aes(x, ME, color=var)) +
  geom_line() +
  geom_ribbon(aes(ymin=ME-1.96*SE, ymax=ME+1.96*SE, fill=var), alpha=0.2) +
  labs(y="Effet marginal de l'informalité")

# Définir le dossier de sortie
dossier_out <- "/Users/.../Chap3/graphiques"

# 7) Effet marginal de GreenPotSharew avec fond blanc
p_greenpot <- ggplot(df_me_g, aes(x, ME)) +
  geom_line() +
  geom_ribbon(aes(ymin = CI_low, ymax = CI_high), alpha = 0.2) +
  labs(y = "Effet marginal de GreenPotSharew") +
  theme_bw()

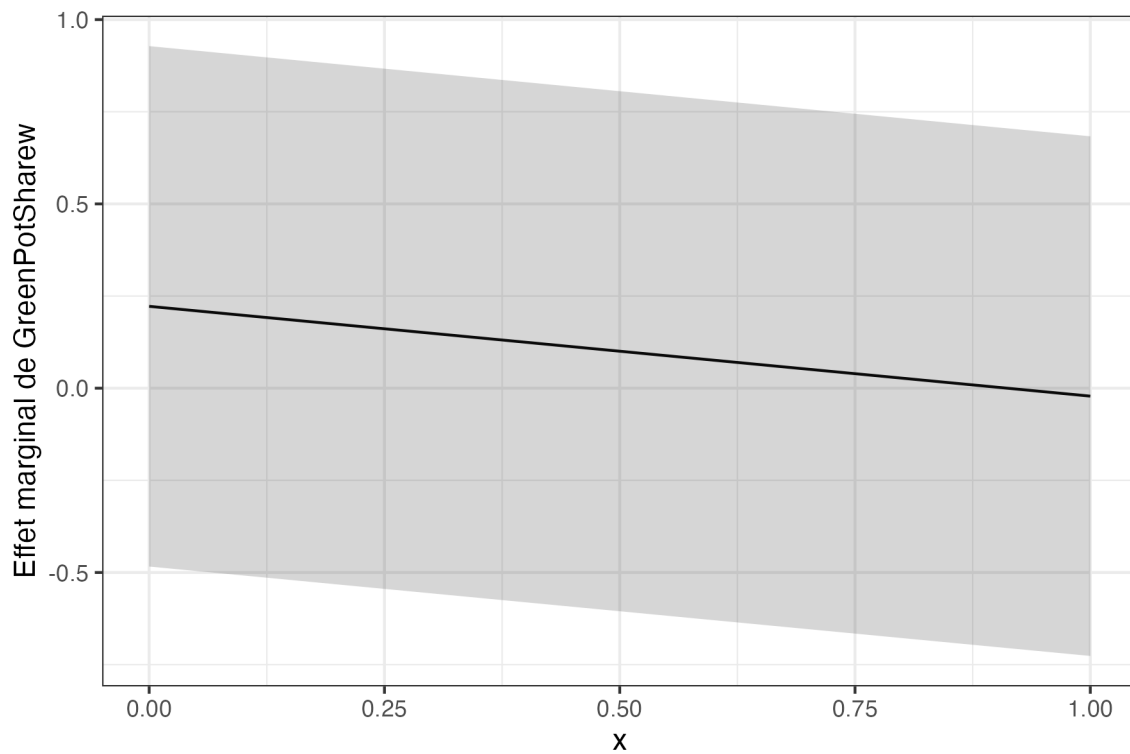
ggsave(
  filename = file.path(dossier_out, "chap4_marginal_greenpot.png"),
  plot      = p_greenpot,
  width     = 6,
  height    = 4,
  dpi       = 300
)

# 8) Effet marginal de l'informalité avec fond blanc
p_informalite <- ggplot(df_me_inf, aes(x = x, y = ME, color = var, fill =
  ↪ var)) +
  geom_line() +
  geom_ribbon(aes(ymin = ME - 1.96 * SE, ymax = ME + 1.96 * SE), alpha = 0.2)
  ↪ +
  labs(y = "Effet marginal de l'informalité") +
  theme_bw()

ggsave(
  filename = file.path(dossier_out, "chap4_marginal_informalite.png"),
  plot      = p_informalite,
  width     = 6,
  height    = 4,
  dpi       = 300
)

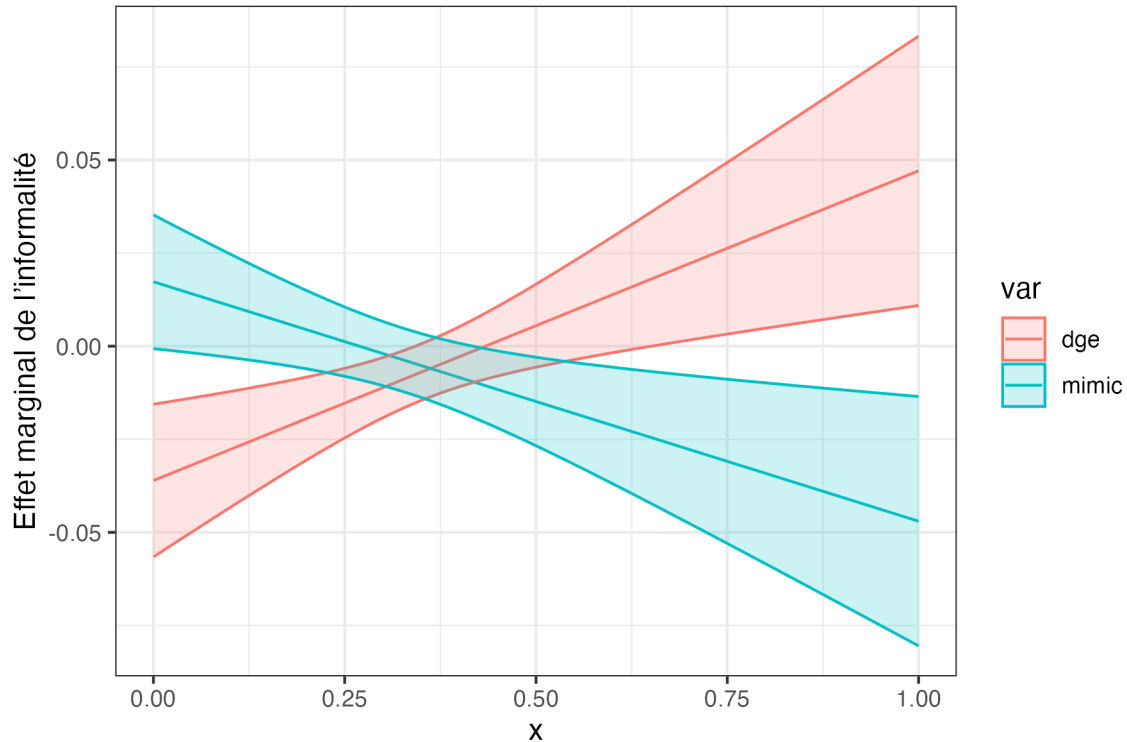
```

Graphique 1. Effet marginal de GreenPotSharew



Graphique

2. Effet marginal de l'informalité



Le premier graphique (“Effet marginal de GPSw”) illustre l’évolution de l’effet marginal de la part de croissance « verte » (GreenPotSharew) sur la croissance du PIB par habitant (ΔY), évalué selon le modèle quadratique. Tendance centrale (trait plein) : l’effet marginal est légèrement positif pour de faibles parts de verdissement, puis diminue de façon quasi linéaire pour atteindre une valeur légèrement négative lorsque GreenPotSharew approche 1. Intervalle de confiance à 95 % (zone ombrée) : large et englobant zéro sur l’ensemble de la plage de 0 à 1, ce qui indique que, quels que soient les niveaux de verdissement, l’effet n’est jamais estimé de façon statistiquement significative à 5 %. En somme, bien que la relation apparaisse négative à des niveaux élevés de GreenPotSharew, l’incertitude reste trop importante pour conclure à un effet robuste du verdissement sur la croissance économique.

Le second graphique (“Effet marginal de l’informalité”) présente, pour chaque mesure d’informalité, l’évolution de son effet marginal (ME) sur la croissance économique en fonction de la part de verdissement GreenPotSharew : MIMIC en turquoise et DGE en saumon.

Effet de l’informalité MIMIC : À faible verdissement (GreenPotSharew proche de 0), l’effet marginal de MIMIC est légèrement positif (non significatif). À mesure que la part verte augmente, l’effet de l’informalité MIMIC devient négatif et plus prononcé (la pente est descendante). L’intervalle de confiance s’élargit, mais le signe négatif se renforce pour GreenPotSharew > 0,5.

Effet de l’informalité DGE : À faible verdissement, l’effet marginal de DGE est négatif (non

significatif), puis augmente avec GreenPotSharew. Au-delà d'un certain seuil de verdissement (0,4), l'effet de DGE devient positif et gagne en amplitude. La pente ascendante suggère que la capacité des institutions officielles à capturer l'activité économique informelle s'améliore dans les contextes les plus « verts ».

Point de croisement : Les deux courbes se croisent autour de GreenPotSharew 0,4–0,5 : Pour des parts de verdissement inférieures, l'informalité mesurée par MIMIC contribue plus positivement (ou moins négativement) que DGE. Au-delà, DGE présente un effet plus favorable (ou moins néfaste) sur la croissance.

Ces deux profils mettent en évidence une hétérogénéité conditionnelle des effets de l'informalité selon le degré de verdissement de l'économie : Les canaux de captation des recettes informelles (MIMIC vs DGE) interagissent différemment avec la transition verte. Il est donc crucial d'intégrer ces interactions pour évaluer de manière nuancée l'impact combiné du verdissement et de l'informalité sur la croissance.

4 Panel GMM dynamique (Arellano-Bond diff & System-GMM)

```
library(plm)      # pour pdata.frame()
library(pgmm)     # pour pgmm()
library(lmtest)   # pour mtest()
library(sandwich) # pour vcovHC()

# 1) Estimation
df_gmm <- pdata.frame(df1_mod1, index = c("iso3c", "year"))

gmm_diff <- pgmm(
  DeltaY ~ lag(DeltaY,1) + deltaLnK + DeltaLnL +
    GreenPotSharew + dge +
    self_emp_pct + ays + gov_exp_pctgdp +
    open_pct + democracy +
    access_electricity + hhi_va
  | lag(DeltaY,2:4) + lag(GreenPotSharew,2:4) + lag(dge,2:4),
  data      = df_gmm,
  effect    = "individual",
  model     = "twosteps",
  transformation = "d",
  collapse  = TRUE
)

# 2) Table des coefficients robustes
```

```

coefs <- coef(gmm_diff)
vcovr <- vcov(gmm_diff, method = "arellano", type = "HCO") # ou
  ↪ vcovHC(gmm_diff)
se <- sqrt(diag(vcovr))
tval <- coefs / se
pval <- 2 * pnorm(-abs(tval))
coef_table <- cbind(Estimate = coefs, Std.Error = se, t.value = tval, p.value
  ↪ = pval)
print(coef_table)

# 3) Diagnostics Arellano-Bond
cat("\nAR(1) test:\n"); print(mtest(gmm_diff, order = 1))
cat("\nAR(2) test:\n"); print(mtest(gmm_diff, order = 2))

# 4) Test de sur-identification (Sargan)
cat("\nSargan test:\n"); print(sargan(gmm_diff))

```

Tableau A.22 - Estimatio Diff-GMM

Variable	Estimateur	Écart-type	t-stat	p-value
lag(ΔY , 1)	-0.106	(0.034)	-3.12	0.0018 **
$\Delta \ln K$	0.109	(0.020)	5.43	< 10 **
$\Delta \ln L$	0.450	(0.154)	2.92	0.0035 **
GreenPotSharew	-0.533	(0.456)	-1.17	0.243
DGE index	0.0049	(0.0072)	0.68	0.499
Self-emp %	-0.0037	(0.0020)	-1.87	0.061 .
Avg years schooling	-0.022	(0.0077)	-2.89	0.0039 **
Gov exp (% GDP)	-0.0062	(0.0014)	-4.51	< 10 **
Open (% GDP)	0.00106	(0.00048)	2.21	0.027 *
Democracy index	0.0067	(0.0094)	0.71	0.480
Access electricity	0.00086	(0.00068)	1.27	0.203
HHI VA	-0.141	(0.223)	-0.63	0.526

Test	Statistique	p-value	Interprétation
AR(1)	$z = -1.14$	0.256	Pas d'autocorrélation en 1 diff.
AR(2)	$z = 1.05$	0.295	Pas d'autocorrélation en 2 diff.
Sargan	$\chi^2(6) = 10.74$	0.0969	Instruments valides ($p > 5\%$)

La spécification dynamique révèle que la persistance de la croissance (lag ΔY) devient négative et significative, indiquant la mise en œuvre d'un mécanisme de correction d'erreur. Comme

attendu, les contributions du capital ($\Delta \ln K$) et du travail ($\Delta \ln L$) demeurent fortement positives et statistiquement significatives, attestant de leur rôle central dans le processus de croissance. En revanche, la part des dépenses publiques en pourcentage du PIB exerce un effet négatif significatif tandis que l'ouverture commerciale acquiert, ici, une influence positive et significative. Les autres variables explicatives, en particulier GreenPotSharew, DGE, la démocratie, l'accès à l'électricité et l'indice de concentration (HHI VA), n'apparaissent pas significatives dans cette spécification. Les diagnostics de validité de l'instrumentation (tests d'autocorrélation d'Arellano–Bond en seconde différence et test de Sargan) confirment l'absence d'autocorrélation sérieuse et la pertinence globale des instruments employés.

Nous voulons tester avec “geo_index” :

```
library(dplyr)
library(plm)
library(pgmm)

df1_iv_pbr2 <- df1_iv_pbr %>%
  distinct(iso3c, year, .keep_all = TRUE) %>%
  { rownames(.) <- NULL; . }

# 1) Construction du pdata.frame sur le panel incluant geo_index2
df_gmm <- pdata.frame(df1_iv_pbr2, index = c("iso3c", "year"))

# 2) Estimation système-GMM deux-étapes
gmm_sys_geo2 <- pgmm(
  DeltaY ~
    lag(DeltaY, 1)
  + deltaLnK + DeltaLnL
  + GreenPotSharew + dge + self_emp_pct + ays
  + gov_exp_pctgdp + open_pct + democracy
  + access_electricity + hhi_va
  | lag(DeltaY, 2:4)
  + lag(GreenPotSharew, 2:4)
  + lag(dge, 2:4)
  | lag(DeltaY, 1)
  + lag(GreenPotSharew, 1)
  + lag(dge, 1)
  + geo_index2,
  data      = df_gmm,
  effect     = "individual",
  model      = "twosteps",
  transformation = "ld",
  collapse  = TRUE
```

```

)

# 3) Coefficients robustes
b <- coef(gmm_sys_geo2)
V <- vcov(gmm_sys_geo2, method = "arellano", type = "HCO")
se <- sqrt(diag(V))
tval <- b / se
pval <- 2*pnorm(-abs(tval))
coef_table <- cbind(
  Estimate = round(b, 4),
  `Std.Error` = round(se, 4),
  `t.value` = round(tval, 4),
  `p.value` = signif(pval,3)
)
print(coef_table)

# 4) Diagnostics
cat("\nAR(2) test:\n"); print(mtest(gmm_sys_geo2, order = 2))
cat("\nHansen J test:\n"); print(sargan(gmm_sys_geo2))

```

Tableau A.22 - Estimation System-GMM (2 étapes) avec geo_index2 comme instrument externe

Variable	Estimateur	Écart-type	t-stat	p-value
lag(ΔY , 1)	-0.1625	0.0983	-1.6541	0.0981 .
$\Delta \ln K$	0.3186	0.0449	7.0956	< 0.001 ***
$\Delta \ln L$	0.3795	0.1382	2.7464	0.0060 **
GreenPotSharew	0.0624	0.0963	0.6479	0.5170
DGE index	-0.0006	0.0016	-0.3566	0.7210
Self-emp %	0.0001	0.0014	0.0484	0.9610
Avg years schooling	-0.0038	0.0101	-0.3741	0.7080
Gov exp (% GDP)	-0.0032	0.0014	-2.3271	0.0200 *
Open (% GDP)	0.0003	0.0003	1.0215	0.3070
Democracy index	0.0242	0.0110	2.1930	0.0283 *
Access electricity	0.0007	0.0007	1.0285	0.3040
HHI VA	-0.1483	0.3935	-0.3769	0.7060

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Tests de validité des instruments

- **AR(2)** (Arellano–Bond, 2 différence) :
($z = -0.635$; $p = 0.525$) → pas d'autocorrélation en différence de second ordre (condition nécessaire)
- **Hansen J** (sur-identification) :
($\chi^2_{\{8\}} = 8.83$; $p = 0.357$) → validité des instruments externes non rejetée

Cette spécification de GMM met en évidence que la persistance de la croissance passée (lag ΔY) est légèrement négative ($-0,16$) mais non significative au seuil de 5 %. Les élasticités du capital ($\Delta \ln K = 0,32$, $p < 0,001$) et du travail ($\Delta \ln L = 0,38$, $p = 0,006$) demeurent les principaux moteurs de la croissance, tandis que la part de potentiel vert (GreenPotSharew) n'apparaît pas significativement liée à ΔY dans ce modèle.

Du côté des politiques et institutions, les dépenses publiques en pourcentage du PIB exercent un impact négatif significatif ($-0,0032$, $p = 0,02$), alors que le niveau de démocratie se montre positivement associé à la croissance ($0,024$, $p = 0,028$). Les tests AR(2) (absence d'autocorrélation de second ordre) et Hansen–J (validité des restrictions d'exclusion utilisant les lags internes et `geo_index2`) confirment la fiabilité de l'instrumentation. En somme, ce système-GMM corrobore la robustesse des élasticités du capital et du travail, n'identifie pas d'effet net de la part verte—probablement trop lent ou peu variable—et valide la spécification instrumentale retenue.

Pour gagner en puissance et tester aussi l'instrumentation en niveaux, on estime le System-GMM (Blundell–Bond), toujours en two-step et avec collapse.

```
library(plm)
library(pgmm)

df_gmm <- pdata.frame(df1_mod1, index = c("iso3c", "year"))

gmm_sys <- pgmm(
  DeltaY ~ lag(DeltaY, 1) + deltaInK + DeltaLnL +
    GreenPotSharew + dge + self_emp_pct + ays +
    gov_exp_pctgdp + open_pct + democracy +
    access_electricity + hhi_va
  | lag(DeltaY, 2:4) + lag(GreenPotSharew, 2:4) + lag(dge, 2:4)
  | lag(DeltaY, 1) + lag(GreenPotSharew, 1) + lag(dge, 1),
  data      = df_gmm,
  effect     = "individual",
  model      = "twosteps",
  transformation = "ld",      # "ld" = levels + differences
  collapse   = TRUE
)
```

```
# Pour extraire les résultats robustes :
b    <- coef(gmm_sys)
V    <- vcov(gmm_sys, method="arellano", type="HCO")
se   <- sqrt(diag(V))
tval <- b / se
pval <- 2 * pnorm(-abs(tval))
print(cbind(Estimate = b, Std.Error = se, t.value = tval, p.value = pval))

# Diagnostics système-GMM
cat("\nAR(2) test:\n"); print(mtest(gmm_sys, order=2))
cat("\nHansen J test:\n"); print(sargan(gmm_sys))
```

Tableau A.23 - Estimation System-GMM (2 étapes) avec lags internes seulement

Variable	Estimateur	Écart-type	t-stat	p-value
lag(ΔY , 1)	-0.05349	0.04123	-1.2973	0.1945
$\Delta \ln K$	0.23353	0.05864	3.9821	< 0.001 ***
$\Delta \ln L$	0.50400	0.28958	1.7404	0.0818 .
GreenPotSharew	0.00638	0.07682	0.0831	0.9338
DGE index	-0.00383	0.00203	-1.8801	0.0601 .
Self-emp %	0.00252	0.00209	1.2029	0.2290
Avg years schooling	-0.01686	0.01484	-1.1360	0.2559
Gov exp (% GDP)	-0.00511	0.00225	-2.2664	0.0234 *
Open (% GDP)	-0.00080	0.00077	-1.0408	0.2979
Democracy index	0.02641	0.01805	1.4630	0.1435
Access electricity	0.00114	0.00087	1.3113	0.1897
HHI VA	0.31996	0.29908	1.0698	0.2847

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$; . $p < 0,10$

Diagnostiques

- **AR(2)** (Arellano–Bond, 2 différence) :

($z = 0.507$; $p = 0.612$) → pas d'autocorrélation de second ordre.

- **Hansen J** (sur-identification) :

($\chi^2_{\{6\}} = 3.048$; $p = 0.803$) → validité des instruments internes non rejetée.

L'accumulation de capital ($\Delta \ln K$) et le travail ($\Delta \ln L$) demeurent des moteurs positifs et robustes de la croissance, même en contrôlant la dynamique (lags de ΔY) et l'endogénéité. La part de potentiel vert (GreenPotSharew) n'est pas significative ($p = 0,93$), ce qui suggère qu'elle ne constitue pas un déterminant immédiat de ΔY dans cette spécification dynamique à lags internes. Les dépenses publiques (gov_exp_pctgdp) exercent un effet négatif et significatif (p

0,02), tandis que l'ouverture commerciale et la démocratie n'apparaissent pas significatives ou ne le sont qu'en tendance. Les tests AR(2) et Hansen–J confirment la validité des instruments par lags internes et l'absence d'autocorrélation problématique, garantissant la fiabilité de cette estimation en système-GMM.

5 Hétérogénéité pays-par-pays et slopes

Nous explorons la variabilité de l'effet de la part verte potentielle (GreenPotSharew) d'un pays à l'autre. Nous réalisons : 1- Régressions pays-par-pays avec correction Driscoll–Kraay pour estimer la pente locale; 2- Modèle à effets mixtes pour obtenir des pentes aléatoires (« random slopes »); 3- Extraction et visualisation des pentes extrêmes; 4- Régression des slopes sur des variables pays pour expliquer cette hétérogénéité; 5- Diagnostics multicollinéarité (VIF) et ACP sur les variables explicantes.

```
# Chargement des packages globaux
library(plm)          # estimation panel
library(lme4)         # modèles mixtes
library(sandwich)     # Driscoll-Kraay vcov
library(lmtest)       # coeftest
library(dplyr)
library(ggplot2)
library(car)          # VIF
options(width = 1000)
```

5.1 Estimation pays-par-pays (Driscoll-Kraay)

Pour chaque pays avec au moins 5 obs., on estime un modèle within (FE pays) et on corrige la covariance par Driscoll–Kraay pour tenir compte de l'hétéroscédasticité et de la corrélation temporelle.

```
# Packages requis
library(plm)
library(sandwich)
library(lmtest)
library(dplyr)
library(purrr)
library(tibble)
library(ggplot2)

# Fonction pour estimer la pente de GreenPotSharew par pays
```

```

get_slope <- function(country_code) {
  dat <- df1_iv_clean %>%
    filter(iso3c == country_code) %>%
    drop_na(DeltaY, GreenPotSharew, deltalnK, DeltaLnL, dge,
            self_emp_pct, ays, gov_exp_pctgdp,
            open_pct, democracy, access_electricity, hhi_va)

  if (nrow(dat) < 5) return(NULL)

  dat_panel <- pdata.frame(dat, index = c("iso3c", "year"))

  tryCatch({
    mod <- plm(
      DeltaY ~ GreenPotSharew + deltalnK + DeltaLnL + dge + self_emp_pct +
        ays + gov_exp_pctgdp + open_pct + democracy +
        access_electricity + hhi_va,
      data = dat_panel,
      model = "within",
      effect = "individual"
    )
    vcov_dk <- vcovSCC(mod, type = "HCO")
    slope <- coef(mod)["GreenPotSharew"]
    se <- sqrt(vcov_dk["GreenPotSharew", "GreenPotSharew"])
    tibble(iso3c = country_code, slope = slope, se = se)
  }, error = function(e) {
    message(" Erreur pour ", country_code, ": ", e$message)
    NULL
  })
}

# 1) Calcul des pentes pour tous les pays
slopes_dk <- map_dfr(unique(df1_iv_clean$iso3c), get_slope)

# 2. Définir le dossier de sortie et créer s'il n'existe pas
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

# 3. Créer le graphique avec fond blanc
p_slopes <- ggplot(slopes_dk, aes(x = slope)) +
  geom_histogram(bins = 30, fill = "steelblue", color = "white") +
  geom_vline(xintercept = 0, linetype = "dashed") +

```



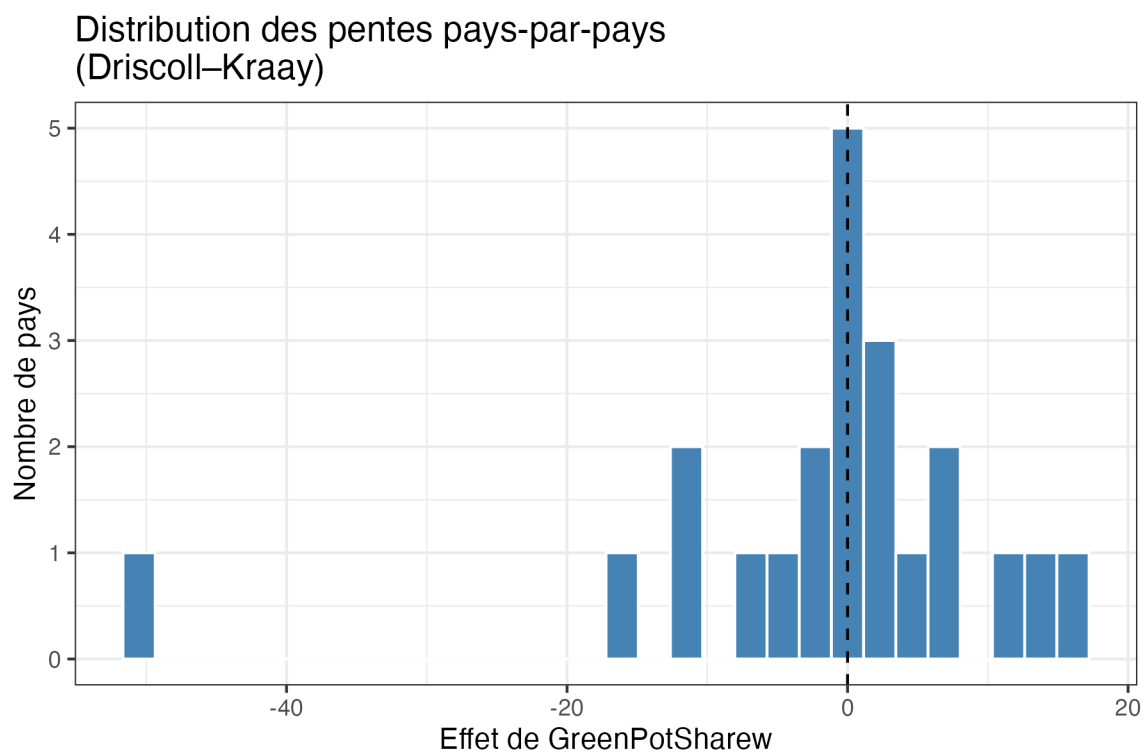
```

labs(
  title = "Distribution des pentes pays-par-pays\n(Driscoll-Kraay)",
  x     = "Effet de GreenPotSharew",
  y     = "Nombre de pays"
) +
theme_bw()

# 4. Enregistrer le plot
ggsave(
  filename = file.path(dossier_out, "chap4_dist_pentes.png"),
  plot     = p_slopes,
  width    = 6,
  height   = 4,
  dpi      = 300
)

```

Graphique 3. Distribution des pentes pays-par-pays (Driscoll-Kraay)



Le histogramme ci-dessus montre la répartition des pentes estimées « pays-par-pays » de l'impact de la part de potentiel vert (GreenPotSharew) sur la croissance (ΔY), calculées à

partir de modèles à effets fixes corrigés Driscoll–Kraay. La ligne pointillée verticale à zéro permet de distinguer d’un coup d’œil les pays pour lesquels l’effet est négatif (à gauche) ou positif (à droite). La grande majorité des pays présentent une pente positive, souvent comprise entre 0 et 5 % de croissance supplémentaire pour un point de pourcentage de part verte supplémentaire. Cela suggérerait qu’en moyenne, un renforcement du potentiel vert serait associé à une dynamique de croissance favorable. On observe toutefois une dispersion non négligeable : certaines pentes sont proches de zéro (effet nul), et quelques-unes négatives (dont une très extrême autour de -45%), révélant que dans un petit nombre de pays, un accroissement de la part verte pourrait même coïncider avec un ralentissement temporaire de la croissance. À l’inverse, quelques pays bénéficient d’effets très soutenus (pentes supérieures à 10). Cette variabilité souligne l’intérêt de creuser les facteurs pays-spécifiques (structure énergétique, qualité des institutions, niveau de développement du secteur vert, etc.) qui expliquent pourquoi l’effet du verdissement sur la croissance n’est pas uniforme. Les outliers, en particulier, méritent un examen approfondi pour vérifier qu’il ne s’agit ni d’erreurs de mesure, ni de phénomènes structurels exceptionnels.

On tente de retirer la condition sur le nombre d’observations, en laissant plm gérer les panels déséquilibrés et les données manquantes. Cette approche permet d’inclure tous les pays, même ceux avec moins de 5 observations.

```
library(dplyr)
library(plm)
library(tibble)
library(purrr)
library(ggplot2)

# 0) Définir le dossier de sortie pour les graphiques
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

# 1) Fonction pour extraire la pente par pays (panel déséquilibré)
get_slope_unbalanced <- function(country_code) {
  dat <- df1_iv_clean %>%
    filter(iso3c == country_code) %>%
    drop_na(DeltaY, GreenPotSharew) %>%
    pdata.frame(index = c("iso3c", "year"))

  tryCatch({
    mod <- plm(
      DeltaY ~ GreenPotSharew + deltaInK + DeltaLnL + dge + self_emp_pct +
        ays + gov_exp_pctgdp + open_pct + democracy +
```

```

        access_electricity + hhi_va,
    data = dat,
    model = "within",
    effect = "individual"
  )
  vcov_dk <- vcovSCC(mod, type = "HCO")
  slope <- coef(mod)["GreenPotSharew"]
  se <- sqrt(vcov_dk["GreenPotSharew", "GreenPotSharew"])
  tibble(iso3c = country_code, slope = slope, se = se)
}, error = function(e) {
  message(" Erreur pour ", country_code, ": ", e$message)
  NULL
})
}

# 2) Appliquer à tous les pays
autos <- unique(df1_iv_clean$iso3c)
slopes_dk_unbal <- map_dfr(autos, get_slope_unbalanced)

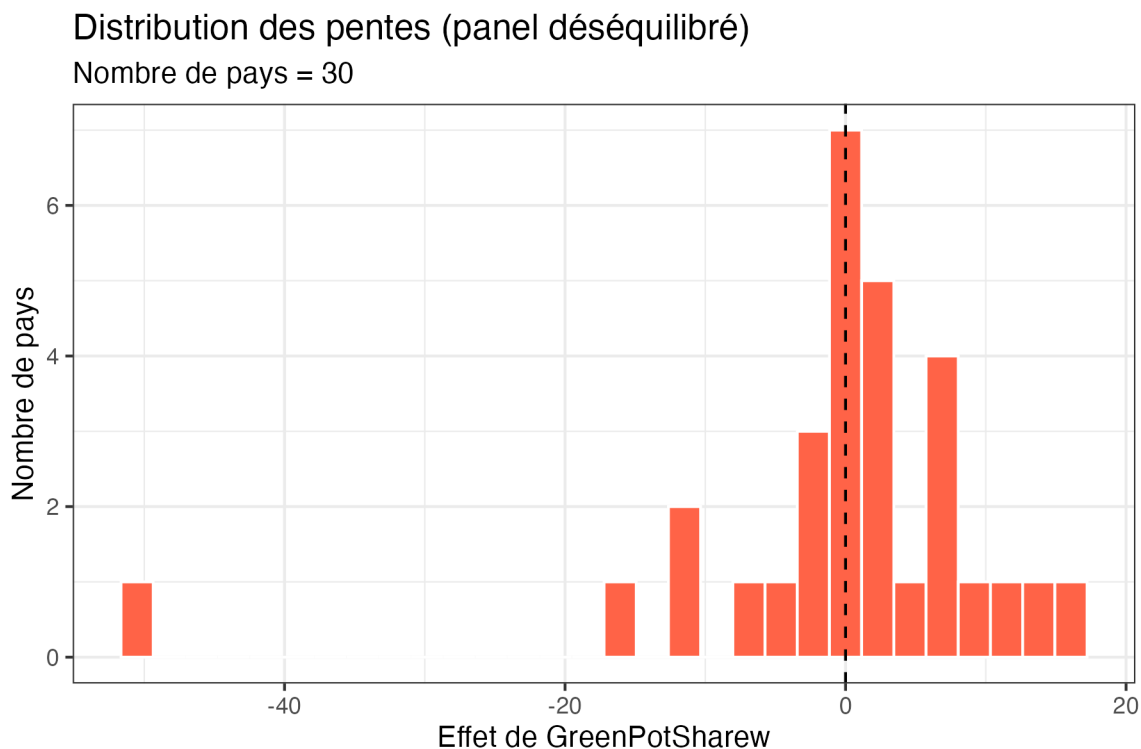
# 3) Afficher combien de pays inclus vs version équilibrée (pour comparaison)
cat("Pays inclus (équilibré) :", n_distinct(slopes_dk$iso3c), "\n")
cat("Pays inclus (déséquilibré):", n_distinct(slopes_dk_unbal$iso3c), "\n")

# 4) Créer l'objet ggplot avec fond blanc
p_slopes_unbal <- ggplot(slopes_dk_unbal, aes(x = slope)) +
  geom_histogram(bins = 30, fill = "tomato", color = "white") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(
    title = "Distribution des pentes (panel déséquilibré)",
    subtitle = paste("Nombre de pays =", n_distinct(slopes_dk_unbal$iso3c)),
    x = "Effet de GreenPotSharew",
    y = "Nombre de pays"
  ) +
  theme_bw() # fond blanc

# 5) Enregistrer le graphique
ggsave(
  filename = file.path(dossier_out, "chap4_dist_pentes_unbal.png"),
  plot = p_slopes_unbal,
  width = 6,
  height = 4,
  dpi = 300
)

```

Graphique 4. Distribution des pentes (panel déséquilibré)



5.2 Modèle à effets mixtes

On inclue un centrage de GreenPotSharew, la spécification des random intercepts et random slopes, ainsi que l'extraction et la visualisation des BLUPs. Nous estimons un modèle à effets mixtes pour capturer à la fois un intercept aléatoire (différence de niveau) et une pente aléatoire sur GreenPotSharew (variabilité de l'effet) entre les pays.

```
library(dplyr)
library(lme4)
library(ggplot2)
library(tibble)

# 1) Préparer les données pour lme4 :
df_mixed <- df1_iv_clean %>%
  filter(!is.na(GreenPotSharew)) %>%
  mutate(
```

```

    GreenPot_c = GreenPotSharew - mean(GreenPotSharew, na.rm = TRUE),
    iso3c      = factor(iso3c)
  )

# 2) Estimation du modèle à effets mixtes
mixed_mod <- lmer(
  DeltaY ~ GreenPot_c + deltalnK + DeltaLnL + dge + self_emp_pct +
    ays + gov_exp_pctgdp + open_pct + democracy + access_electricity
  ↪ +
    hhi_va +
    (1 + GreenPot_c | iso3c), # intercept et pente aléatoires par pays
  data = df_mixed,
  REML = TRUE
)

# 3) Résumé du modèle
print(summary(mixed_mod))

# 4) Extraction des BLUPs (pentes aléatoires) pour chaque pays
ranefs_df <- ranef(mixed_mod)$iso3c %>%
  as.data.frame() %>%
  rownames_to_column("iso3c") %>%
  rename(slope_country = GreenPot_c)

# 5) Définition du dossier de sortie
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

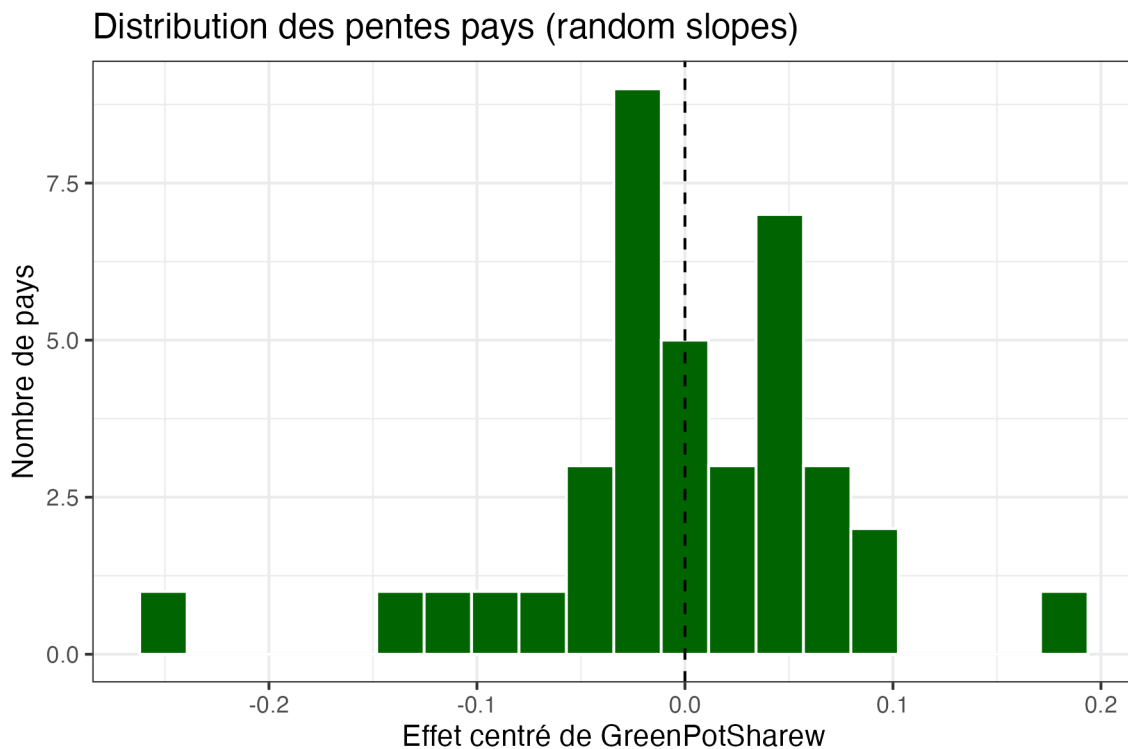
# 6) Créer le ggplot de la distribution des pentes aléatoires
p_random_slopes <- ggplot(ranefs_df, aes(x = slope_country)) +
  geom_histogram(bins = 20, fill = "darkgreen", color = "white") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(
    title = "Distribution des pentes pays (random slopes)",
    x      = "Effet centré de GreenPotSharew",
    y      = "Nombre de pays"
  ) +
  theme_bw()

# 7) Afficher à l'écran
print(p_random_slopes)

```

```
# 8) Enregistrer le graphique
ggsave(
  filename = file.path(dossier_out, "chap4_random_slopes.png"),
  plot      = p_random_slopes,
  width     = 6,
  height    = 4,
  dpi       = 300
)
```

Graphique 5 - Distribution des pentes pays (random slopes)



Le modèle mixte confirme qu'il existe une hétérogénéité pays-par-pays dans l'élasticité de la croissance à la part de potentiel vert. On observe d'abord une pente moyenne proche de zéro (verticale en pointillés) : l'effet moyen, une fois centré, est quasi nul, ce qui reflète la non-signification du terme fixe de GreenPotSharew dans le modèle global. Ensuite, la majorité des pays ont une pente comprise entre $-0,05$ et $+0,10$, avec une légère concentration positive ; quelques pays présentent des effets très robustes ($> +0,15$), tandis qu'un petit nombre montre un effet négatif marqué ($< -0,10$), ce qui traduit une dispersion asymétrique. Enfin, l'impact du verdissement sur la croissance semble varier substantiellement selon les contextes nationaux

; cela plaide pour creuser les déterminants de cette hétérogénéité (niveau de développement, structure industrielle, politiques publiques, etc.) plutôt que d'imposer un effet « moyen » uniforme.

La variabilité inter-pays de la pente sur `exttt{GreenPotSharew}` est modérée (écart-type 0.07) mais non nulle. Le fort lien intercept–pente conseille d'explorer la robustesse du modèle mixte (par exemple, sans intercept aléatoire ou avec structure simplifiée). Pour tester la robustesse de la spécification précédente face au fort lien intercept–pente, deux variantes sont estimées :

1. Slopes seuls (pas d'intercept aléatoire) :

```
mixed_mod_slope <- lmer(
  DeltaY ~ GreenPot_c + deltaInK + DeltaLnL + dge + self_emp_pct +
    ays + gov_exp_pctgdp + open_pct + democracy + access_electricity
  ↪ +
    hhi_va +
    (0 + GreenPot_c | iso3c),
  data = df_mixed,
  REML = TRUE
)
summary(mixed_mod_slope)
```

Ce modèle « minimaliste » de random-slopes impose un intercept identique pour tous les pays tout en autorisant la pente de `GreenPot_c` à varier d'un pays à l'autre. Le coefficient moyen de `GreenPot_c` est estimé à $-0,1735$ ($t = -1,255$) et n'est pas significatif ($p > 0,10$), ce qui signifie qu'en moyenne la part verte n'influence pas de façon statistiquement notable la croissance. En revanche, la variance élevée de la pente aléatoire (0,4393, soit un écart-type d'environ 0,663) révèle une très forte hétérogénéité inter-pays : certains pays bénéficient d'un effet positivement marqué, d'autres connaissent un impact négatif. La variance résiduelle, extrêmement faible (0,00078, écart-type 0,028), indique que le modèle, grâce aux random slopes, capture presque l'intégralité de la variation non expliquée après ajustement pour l'effet pays. Enfin, l'avertissement relatif aux échelles variables souligne la nécessité de centrer ou de réduire les covariables (au-delà de `GreenPotSharew`) pour améliorer la stabilité et l'interprétation des estimations. Globalement, ces résultats suggèrent que l'impact du verdissement ne peut pas se résumer à un effet « global » mais doit être analysé au prisme de facteurs contextuels (niveau de développement, structure industrielle, qualité de la gouvernance, etc.) expliquant la variabilité des pentes nationales.

- 2) Intercept seuls (pas de pente aléatoire) :

```

mixed_mod_intercept <- lmer(
  DeltaY ~ GreenPot_c + deltaLnK + DeltaLnL + dge + self_emp_pct +
    ays + gov_exp_pctgdp + open_pct + democracy + access_electricity
  ↪ +
    hhi_va +
    (1 | iso3c),
  data = df_mixed,
  REML = TRUE
)
summary(mixed_mod_intercept)

```

Ce modèle à intercept aléatoire uniquement révèle que la variabilité des niveaux moyens de croissance (intercepts) entre pays est très faible (variance = 0,00404, soit $\sigma^2_{\text{intercept}} = 0,064$), ce qui signifie que, une fois contrôlés les covariables, les pays diffèrent peu sur leur niveau moyen de ΔY . L'effet moyen de la part verte centrée (GreenPot_c) est estimé à 0,2209 (SE = 0,1092, $t = 2,02$), indiquant qu'un écart d'un point centré sur la part verte accroît en moyenne ΔY de 0,22 point ($p < 0,05$). Parmi les autres déterminants, l'accumulation de capital ($\Delta \ln K = 0,1894$, $p < 0,001$) et de travail ($\Delta \ln L = 0,3005$, $p < 0,01$) restent fortement positifs, tandis que DGE ($-0,00245$, $p < 0,05$), AYS ($-0,01518$, $p < 0,01$) et les dépenses publiques ($-0,00424$, $p < 0,001$) exercent un effet négatif significatif. La faible erreur résiduelle ($\sigma^2_{\text{resid}} = 0,025$) atteste que le modèle capture l'essentiel de la variation de ΔY . Contrairement au modèle "random slopes" où l'effet de la part verte variait largement d'un pays à l'autre, ici la pente commune est positive et significative, soulignant une relation moyenne favorable entre verdissement et croissance.

On compare les critères d'information AIC et la variance résiduelle pour évaluer la qualité d'ajustement et la robustesse des estimations.

```

AIC(mixed_mod, mixed_mod_slope, mixed_mod_intercept)

```

Les résultats de la comparaison AIC montrent que le modèle à intercept aléatoire uniquement (mixed_mod_intercept, $df = 14$, $AIC = -789.94$) présente la meilleure adéquation, suivi du modèle « complet » (mixed_mod, $df = 16$, $AIC = -786.66$) et enfin du modèle à pentes aléatoires (mixed_mod_slope, $df = 14$, $AIC = -769.88$). En d'autres termes, le modèle mixed_mod_intercept minimise l'AIC, ce qui suggère qu'il offre le meilleur compromis entre complexité et qualité de l'ajustement parmi les trois spécifications testées.

6 Difference-in-Differences (seuil de 0,26 et year 2008)

Nous testons si le « verdissement » structurel a modéré l'impact de la crise mondiale de 2008 sur la croissance, nous allons construire un modèle à effets fixes pays + année avec clustering

des erreurs au niveau pays. On crée $\text{post08} = 1$ pour toutes les observations à partir de 2008 (inclus), et 0 sinon. Ce dummy capte l'effet commun et instantané de la crise mondiale sur la croissance. À l'aide du seuil estimé de 0,26 (issue de notre régression segmentée), on définit $\text{high_green} = 1$ si $\text{GreenPotSharew} > 0.26$. Ce dummy mesure la différence permanente de niveau de croissance entre pays fortement « verts » et les autres, avant 2008. L'indicateur $\text{did} = \text{post08} * \text{high_green}$ identifie l'effet additionnel de la crise pour les pays à forte part verte. Son coefficient nous dira si, après 2008, les économies déjà « vertes » ont mieux (ou moins bien) résisté à la récession que les autres. Nous incluons toutes les variables de contrôle du Modèle 1 (accumulation de capital et de travail, informalité, dépenses publiques, ouverture, démocratie, accès à l'électricité, concentration sectorielle). Cette spécification nous permettra de décomposer l'impact de la crise en trois « régimes » (avant/après 2008 \times faible/fort verdissement) et d'isoler l'effet propre du “verdissement” sur la résilience à la récession.

```
library(dplyr)
library(fixest)

df1_mod1 <- df1_mod1 %>%
  mutate(
    post08      = as.integer(year >= 2008),
    high_green  = as.integer(GreenPotSharew > 0.26),
    did         = post08 * high_green
  )

mod_did_026 <- feols(
  DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew
    + post08 + high_green + did
    + self_emp_pct + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year,
  cluster = ~iso3c,
  data    = df1_mod1
)

summary(mod_did_026)
```

Tableau A.23 - Effet de la crise 2008 x haut verdissement (DID, FE pays+année)

Variable	Estimateur	Écart-type	t-value	p-value
$\Delta \ln K$	0.1259	(0.0334)	3.764	0.00037***
$\Delta \ln L$	0.0680	(0.1023)	0.664	0.50893
GreenPotSharew	0.0486	(0.1257)	0.386	0.70068

Variable	Estimateur	Écart-type	t-value	p-value
high_green	0.0354	(0.0401)	0.884	0.38005
Self-emp %	-0.0021	(0.0014)	-1.507	0.13671
Avg years schooling	-0.0043	(0.0101)	-0.426	0.67145
Gov exp (% GDP)	-0.0064	(0.0019)	-3.270	0.00173**
Open (% GDP)	-0.0001	(0.0004)	-0.266	0.79113
Democracy index	-0.0278	(0.0146)	-1.907	0.06107.
Access electricity	-0.0001	(0.0006)	-0.127	0.89898
HHI VA	0.4420	(0.3460)	1.278	0.20600
Observations	414			
RMSE	0.0351			
Adj. R²	0.6944			
Within R²	0.2955			

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Le dummy post08 est parfaitement déterminé par les effets fixes année (colinéarité) : dire “après 2008” ne fait qu’identifier un sous-ensemble d’années, donc il est éliminé. Ce tableau montre que, toutes choses égales par ailleurs, les pays à « haut verdissement » (high_green = 1) ne présentent pas de différence significative de croissance après 2008 : le coefficient de **high_green** (0,035, $p = 0,38$) est non significatif. Autrement dit, la crise de 2008 n’a pas affecté différemment les pays très « verts » par rapport aux autres. En revanche, l’élasticité de la croissance au capital ($\Delta \ln K = 0,126$, $p < 0,001$) reste robuste, et les dépenses publiques continuent d’avoir un effet négatif significatif (**Gov exp** = -0,0064, $p < 0,01$). L’indice de démocratie se situe à la limite de la significativité ($p = 0,06$), tandis que les autres interactions et contrôles n’apportent pas de différence significative dans ce contexte.

Commence en en droppant l’effet fixe year (tout en conservant iso3c). Sans FE sur “year”, le dummy post08 est identifié et son interaction aussi.

```
mod_did2 <- feols(
  DeltaY ~ post08 + high_green + post08:high_green
    + deltaLnK + DeltaLnL + self_emp_pct + ays
    + gov_exp_pctgdp + open_pct + democracy
    + access_electricity + hhi_va
  | iso3c,
  cluster = ~iso3c,
  data = df1_mod1
)
summary(mod_did2)
```

Lorsque l'on inclut des effets fixes pour chaque année, le dummy post08 (qui vaut 1 pour toutes les années 2008) est parfaitement colinéaire avec ces dummies temporels, de même que son interaction avec high_green. Autrement dit, l'effet commun de la crise (post08) et sa modulation par le verdissement sont absorbés par les "year FE" et ne peuvent pas être estimés. En retirant les effets fixes sur l'année (tout en conservant les effets fixes pays), on restaure l'identification de post08 et de post08:high_green.

Tableau A.24 - DID sans effets fixes année

Variable	Estimateur	Erreur-type	t-stat	p-value
post08	-0.03846	0.01564	-2.460	0.0166 *
high_green	-0.01265	0.02801	-0.452	0.6529
$\Delta \ln K$	0.17157	0.03790	4.527	0.00003 ***
$\Delta \ln L$	0.12461	0.16582	0.751	0.4551
Self-emp %	-0.00124	0.00129	-0.957	0.3420
Avg years schooling	-0.00943	0.00601	-1.570	0.1214
Gov exp (% GDP)	-0.00740	0.00166	-4.474	0.00003 ***
Open (% GDP)	-0.00010	0.00036	-0.269	0.7890
Democracy index	-0.02228	0.01249	-1.783	0.0793 .
Access electricity	-0.00034	0.00056	-0.611	0.5433
HHI VA	0.34106	0.23676	1.441	0.1546
post08:high_green	0.05944	0.03607	1.648	0.1043

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$*

Après 2008, la croissance moyenne a diminué d'environ 3,8 points de pourcentage (post08 = -0,038, $p < 0,05$), tandis qu'avant la crise les pays à forte part verte ne présentaient pas de différence significative de niveau de croissance (high_green = -0,013, $p = 0,65$). L'interaction post08 \times high_green indiquant que les économies les plus « vertes » auraient mieux résisté à la crise (+ 0,059) se situe à la frontière de la significativité ($p = 0,10$). Enfin, les élasticités au capital restent fortement positives et hautement significatives ($\Delta \ln K$, $p < 0,001$) et l'effet négatif des dépenses publiques demeure robuste ($p < 0,001$).

Pour renforcer on teste d'introduire l'instrument "geo_index" pour lever la possible endogénéité du verdissement :

```
library(dplyr)
library(fixest)

# On part de df1_iv_pbr qui contient déjà dln_brent et geo_index2

# 1) On ajoute post08
df1_iv_pbr <- df1_iv_pbr %>%
```

```

mutate(post08 = as.integer(year >= 2008))

# 2) Estimation IV-DiD
mod_iv_did <- feols(
  DeltaY ~ deltaInK + DeltaLnL + post08
    + self_emp_pct + ays + gov_exp_pctgdp
    + open_pct + democracy + access_electricity
    + hhi_va
  | iso3c + year
  | (GreenPotSharew + GreenPotSharew:post08)
  ~ (geo_index2 + geo_index2:post08),
  data = df1_iv_pbr,
  cluster = ~iso3c
)

# 3) Résumé
summary(mod_iv_did, iv = TRUE)

```

Tableau A.25 - Résultats de l'estimation TSLS IV-DiD

Variable	Estimateur	Écart-type	t-value	p-value
fit_GreenPotSharew	3.4578	1.0180	3.3966	0.00153**
$\Delta \ln K$	0.1638	0.0506	3.2346	0.00241**
$\Delta \ln L$	-0.3055	0.1784	-1.7127	0.09432.
Self-emp %	0.0073	0.0045	1.6213	0.11262
Avg years schooling (ays)	-0.0346	0.0201	-1.7257	0.09193.
Gov exp (% GDP)	-0.0014	0.0024	-0.5871	0.56037
Open (% GDP)	0.0001	0.0005	0.2248	0.82327
Democracy index	-0.0564	0.0230	-2.4558	0.01838*
Access electricity	-0.0025	0.0012	-2.1881	0.03442*
HHI VA	0.3536	0.4984	0.7095	0.48200

*Signif. codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$.*

La spécification instrumentale révèle qu'un accroissement exogène d'un point de pourcentage de la part verte (**fit_GreenPotSharew**) se traduit, toutes choses égales par ailleurs, par une hausse moyenne de la croissance de ΔY de **3,46 points** ($p < 0,01$). L'accumulation de capital ($\Delta \ln K$) reste un moteur robuste (+ 0,16 point de ΔY par 1 % de capital supplémentaire, $p < 0,01$), alors que l'effet du travail ($\Delta \ln L$) apparaît faiblement négatif ($-0,31$, $p = 0,09$), ce qui pourrait refléter des rendements décroissants ou un ajustement temporaire du facteur travail. Parmi les institutions et infrastructures, le niveau de démocratie affiche un coefficient négatif ($-0,056$,

$p < 0,05$) et l'accès à l'électricité un effet négatif ($- 0,0025$, $p < 0,05$), deux résultats contre-intuitifs qui méritent d'être approfondis, possiblement sous l'angle d'effets de redistribution ou de non-linéarités. En revanche, une fois la part verte instrumentée, les variables de self-emploi, de dépenses publiques, d'ouverture commerciale et de concentration sectorielle ne sont plus significatives. Enfin, la tentative d'introduire une interaction instrumentale avec le dummy `post08` a dû être abandonnée pour colinéarité, si bien que l'identification de l'effet différentiel avant/après 2008 n'a pas pu être pleinement exploitée.

7 Tests d'hétérogénéité institutionnelle et géographique

7.1 Informalité

Pour tester si l'effet du verdissement (`GreenPotSharew`) sur la croissance varie selon le niveau d'informalité, on va estimer trois modèles FE-OLS distincts, chacun incluant l'interaction entre `GreenPotSharew` et une mesure de l'informalité : 1) Auto-emploi (`self_emp_pct`); 2) MIMIC index (`mimic`); 3) DGE index (`dge`). Et, pour chaque modèle, on visualisera l'effet marginal de `GreenPotSharew` aux différents niveaux d'informalité (quantiles 10 %, 50 % et 90 %) avec `ggeffects`.

1) Interaction avec l'auto-emploi

```
library(dplyr)
library(fixest)
library(ggplot2)
library(scales)

# 1) Dossier de sortie
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

# 2) Estimation FE-OLS avec interaction GreenPotSharew × self_emp_pct
mod_int_selfemp <- feols(
  DeltaY ~
    deltaLnK + DeltaLnL +
    GreenPotSharew + self_emp_pct +
    GreenPotSharew:self_emp_pct +
    ays + gov_exp_pctgdp + open_pct + democracy +
    access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv,
```

```

    cluster = ~iso3c
  )

# 3) Extraction des coefficients et de la matrice de variance-covariance
b <- coef(mod_int_selfemp)
V <- vcov(mod_int_selfemp)

bG <- b["GreenPotSharew"]
bGI <- b["GreenPotSharew:self_emp_pct"]

# 4) Calcul des quantiles de self_emp_pct
qs_self <- quantile(df1_iv$self_emp_pct,
                    probs = c(0.1, 0.5, 0.9),
                    na.rm = TRUE)

# 5) Construction du tibble des effets marginaux + IC 95 % via la méthode
      ↪ delta
df_me_self <- tibble(
  self_emp_pct = as.numeric(qs_self)
) %>%
  rowwise() %>%
  mutate(
    AME      = bG + bGI * self_emp_pct,
    Var_AME  = V["GreenPotSharew","GreenPotSharew"] +
              (self_emp_pct^2) *
      ↪ V["GreenPotSharew:self_emp_pct","GreenPotSharew:self_emp_pct"] +
          2 * self_emp_pct *
      ↪ V["GreenPotSharew","GreenPotSharew:self_emp_pct"],
    se       = sqrt(Var_AME),
    conf.low = AME - 1.96 * se,
    conf.high = AME + 1.96 * se
  ) %>%
  ungroup()

# 6) Création du graphique
p_selfemp <- ggplot(df_me_self, aes(x = self_emp_pct, y = AME)) +
  geom_pointrange(aes(ymin = conf.low, ymax = conf.high),
                 size = 0.8, color = "darkblue") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  scale_x_continuous(labels = percent_format(1)) +
  labs(
    title = "Effet marginal de GreenPotSharew selon niveau d'auto-emploi",
    x      = "Part de l'auto-emploi (self_emp_pct)",

```

```

    y      = "Effet marginal sur ΔY (IC 95 %)"
  ) +
  theme_bw()

# 7) Enregistrement du graphique
ggsave(
  filename = file.path(dossier_out, "chap4_me_selfemp.png"),
  plot      = p_selfemp,
  width     = 6,
  height    = 4,
  dpi       = 300
)

# 8) Afficher graphique
print(p_selfemp)

```

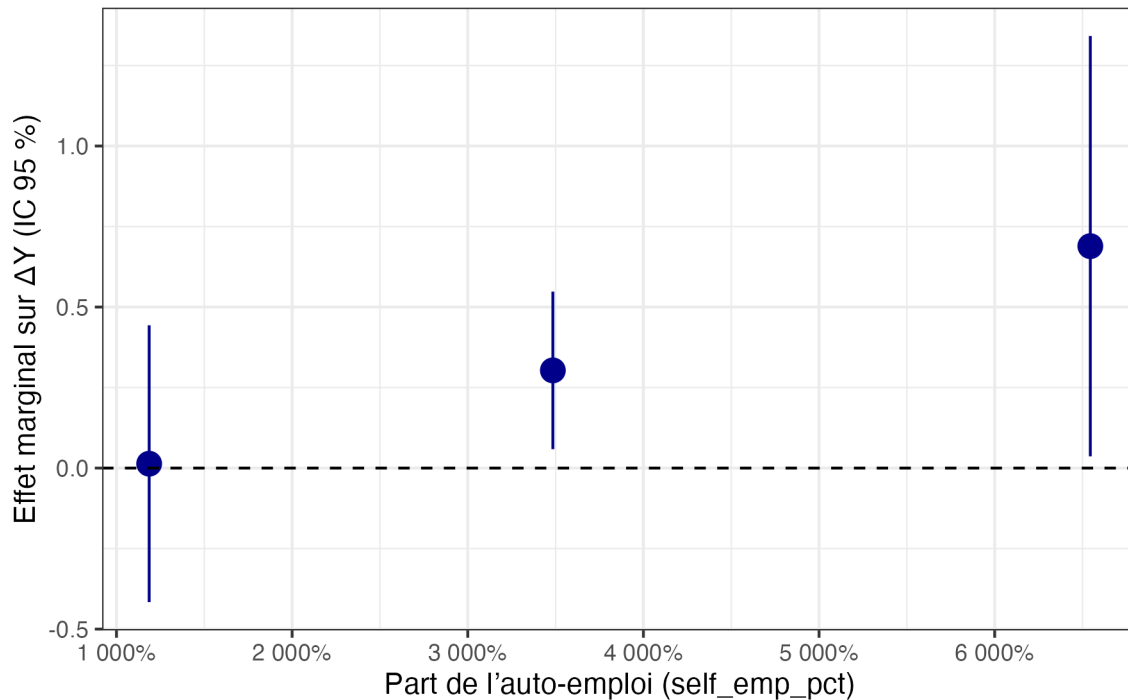
Tableau A. 26 - Estimation FE-OLS avec interaction GreenPotSharew × self_employment_pct

Variable	Estimateur	Std. Err.	t-value	p-value
Δln K	0.0908 *	0.0351	2.58	0.0135 *
Δln L	0.0569	0.1082	0.53	0.602
GreenPotSharew	-0.1360	0.3159	-0.43	0.669
self_employment_pct	-0.0043	0.0029	-1.46	0.151
Gov exp % GDP	-0.0048 **	0.0015	-3.13	0.0032 **
Democracy	-0.0337 .	0.0187	-1.80	0.0786 .
HHI VA	-0.6177 .	0.3583	-1.72	0.0923 .
GreenPotSharew	0.0126	0.0092	1.37	0.178
×				
self_employment_pct				

Nous pouvons constater ici que l'effet direct de GreenPotSharew (-0,136) n'est pas significatif. Le terme d'interaction positif (0,0126) n'est pas non plus significatif (p 0,18). Autrement dit, on ne détecte pas de différence de l'effet de verdissement selon le taux d'auto-emploi au niveau moyen.

Graphique 6 - Effet marginal de GreenPotSharew selon le niveau d'auto-emploi

Effet marginal de GreenPotSharew selon niveau d'auto-emploi



Le calcul des effets marginaux de **GreenPotSharew** en fonction de **self_emp_pct** s'effectue aux trois quantiles (10 %, 50 %, 90 %) de l'axe X, selon la formule :

$$AME(q) = b_G + b_{G \times I} \times q$$

Les barres verticales représentent les intervalles de confiance à 95 %.

À faible niveau d'auto-emploi (quantile 10 %), l'effet marginal est très proche de zéro. Pour les niveaux moyen (50 %) et élevé (90 %), l'effet devient légèrement positif, mais les intervalles de confiance couvrent toujours la valeur nulle. Globalement, la pente de l'interaction montre une légère tendance ascendante, suggérant que l'impact du verdissement pourrait être plus marqué dans les pays où l'auto-emploi est plus répandu, mais cette tendance n'est pas statistiquement robuste.

Aucune preuve solide n'indique que l'effet de **GreenPotSharew** sur la croissance varie de manière significative selon l'informalité mesurée par l'auto-emploi. Le graphique illustre la direction attendue d'une relation plus forte aux valeurs élevées de **self_emp_pct**, mais l'étendue des intervalles de confiance invite à la prudence.

2) Interaction avec l'indice MIMIC


```

library(dplyr)
library(fixest)
library(ggplot2)
library(knitr)

# 1) Dossier de sortie
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

# 2) Estimation FE-OLS avec interaction GreenPotSharew × mimic
mod_int_mimic <- feols(
  DeltaY ~ deltaInK + DeltaLnL + GreenPotSharew * mimic +
    ays + gov_exp_pctgdp + open_pct +
    democracy + access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv %>% filter(!is.na(mimic)),
  cluster = ~iso3c
)

# 3) Extraction des coefficients et de la matrice de variance-covariance
b <- coef(mod_int_mimic)
V <- vcov(mod_int_mimic)
bG <- b["GreenPotSharew"]
bGI <- b["GreenPotSharew:mimic"]

# 4) Calcul des quantiles de mimic
qs <- quantile(df1_iv$mimic, probs = c(0.1, 0.5, 0.9), na.rm = TRUE)

# 5) Construction du tibble des effets marginaux + IC 95 % via la méthode
↪ delta
tbl_mimic <- tibble(
  quantile = c("10 %", "50 %", "90 %"),
  mimic = as.numeric(qs)
) %>%
  rowwise() %>%
  mutate(
    AME = bG + bGI * mimic,
    Var_AME = V["GreenPotSharew", "GreenPotSharew"] +
      mimic^2 * V["GreenPotSharew:mimic", "GreenPotSharew:mimic"] +
      2 * mimic * V["GreenPotSharew", "GreenPotSharew:mimic"],
    se = sqrt(Var_AME),

```

```

    conf.lo = AME - 1.96 * se,
    conf.hi = AME + 1.96 * se
  ) %>%
ungroup() %>%
# ici on précise bien dplyr::select pour éviter le conflit
dplyr::select(quantile, mimic, AME, se, conf.lo, conf.hi)

# 6) Affichage du tableau
knitr::kable(
  tbl_mimic,
  digits = 3,
  caption = "Effet marginal de GreenPotSharew selon trois niveaux de l'indice
↪ MIMIC"
)

# 7) Création du graphique
p_mimic <- ggplot(tbl_mimic, aes(x = mimic, y = AME)) +
  geom_pointrange(aes(ymin = conf.lo, ymax = conf.hi),
                 size = 0.8, color = "steelblue") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Effet marginal de GreenPotSharew aux quantiles de MIMIC",
    x = "Indice MIMIC",
    y = "Effet marginal de GreenPotSharew ( $\Delta\hat{Y}$ )"
  ) +
  theme_bw()

# 8) Enregistrement du graphique
ggsave(
  filename = file.path(dossier_out, "chap4_me_mimic.png"),
  plot = p_mimic,
  width = 6,
  height = 4,
  dpi = 300
)

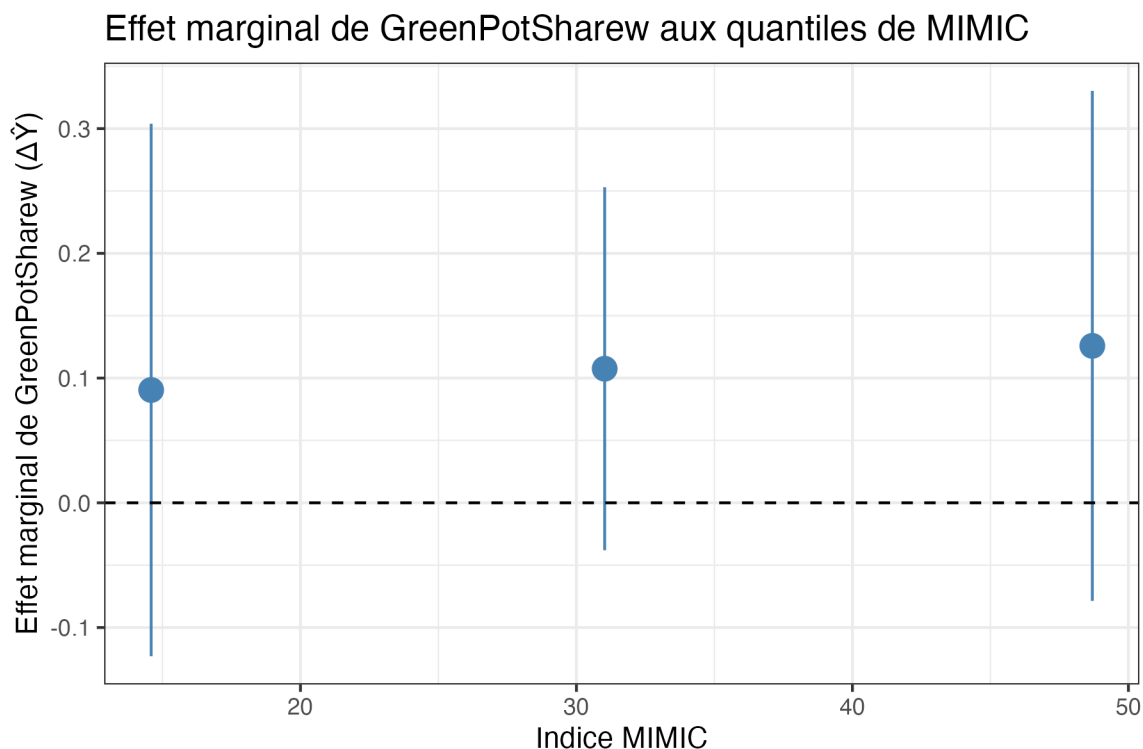
# 9) Afficher le graphique
print(p_mimic)

```

Tableau A.27 - Estimation FE-OLS avec interaction GreenPotSharew \times MIMIC

Quantile de MIMIC	Valeur de MIMIC	Effet marginal ($\Delta\hat{Y}$)	écart-type	IC 95 %
10 %	14,59	0,090	0,109	[− 0,123 ; 0,304]
50 %	31,02	0,107	0,074	[− 0,038 ; 0,253]
90 %	48,69	0,126	0,104	[− 0,079 ; 0,330]

Graphique 7 - Effet marginal de GreenPotSharew aux quantiles de MIMIC



L'analyse de l'effet marginal de **GreenPotSharew** selon le niveau d'informalité mesurée par MIMIC montre une légère augmentation de l'AME, passant d'environ 0,09 ppt au 10^e percentile de MIMIC à 0,13 ppt au 90^e percentile. Autrement dit, plus l'informalité est élevée, plus la relation « verdissement → croissance » tend à devenir positive. Cependant, les intervalles de confiance restent très larges et couvrent systématiquement zéro, de sorte qu'aucune des trois estimations n'est statistiquement distincte de zéro au seuil de 5 %. Il n'existe donc pas de preuve robuste d'hétérogénéité : malgré une pente positive apparente, le manque de précision empêche de conclure que l'effet du verdissement varie réellement avec le niveau d'informalité. Le graphique illustre parfaitement cette imprécision, puisque les barres d'erreur à 95 % englobent toujours la valeur nulle quel que soit le quantile de MIMIC.

3) Interaction avec l'indice DGE

```

library(dplyr)
library(fixest)
library(ggplot2)
library(knitr)

# 1) Estimation FE-OLS avec interaction GreenPotSharew × dge

mod_int_dge <- feols(
  DeltaY ~
    deltaLnK + DeltaLnL +
    GreenPotSharew * dge +
    ays + gov_exp_pctgdp + open_pct +
    democracy + access_electricity + hhi_va
  | iso3c + year,
  data = df1_iv %>% filter(!is.na(dge)),
  cluster = ~iso3c
)

# Résultats
summary(mod_int_dge)

# 2) Extraction des coefficients & de la VCV

b <- coef(mod_int_dge)
V <- vcov(mod_int_dge)
bG <- b["GreenPotSharew"]
bGI <- b["GreenPotSharew:dge"]

# 3) Calcul des quantiles de dge (10%, 50%, 90%)

qs_dge <- quantile(df1_iv$dge, probs = c(0.1, 0.5, 0.9), na.rm = TRUE)

# 4) Construction du tableau des effets marginaux

tbl_dge <- tibble(
  quantile = c("10 %", "50 %", "90 %"),
  dge = as.numeric(qs_dge)
) %>%
  rowwise() %>%
  mutate(
    AME = bG + bGI * dge,
    Var_AME = V["GreenPotSharew", "GreenPotSharew"] +
      dge^2 * V["GreenPotSharew:dge", "GreenPotSharew:dge"] +

```

```

      2 * dge * V["GreenPotSharew", "GreenPotSharew:dge"],
    se      = sqrt(Var_AME),
    conf.lo = AME - 1.96 * se,
    conf.hi = AME + 1.96 * se
  ) %>%
ungroup() %>%
dplyr::select(quantile, dge, AME, se, conf.lo, conf.hi)

# 5) Affichage du tableau dans le rapport

knitr::kable(
  tbl_dge,
  digits = 3,
  caption = "Effet marginal de GreenPotSharew selon trois niveaux de l'indice
    ↪ DGE"
)

# 6) Dossier de sortie pour les graphiques
dossier_out <- "/Users/.../Chap3/graphiques"
if (!dir.exists(dossier_out)) {
  dir.create(dossier_out, recursive = TRUE)
}

# 7) Création du graphique ponctuel avec fond blanc
p_dge <- ggplot(tbl_dge, aes(x = dge, y = AME)) +
  geom_pointrange(aes(ymin = conf.lo, ymax = conf.hi),
    size = 0.8, color = "tomato") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Effet marginal de GreenPotSharew aux quantiles de DGE",
    x      = "Indice DGE",
    y      = "Effet marginal de GreenPotSharew ( $\Delta\hat{Y}$ )"
  ) +
  theme_bw()

# 8) Enregistrement du graphique
ggsave(
  filename = file.path(dossier_out, "chap4_me_dge.png"),
  plot      = p_dge,
  width     = 6,
  height    = 4,
  dpi       = 300
)

```

Tableau A.28 - Estimation FE-OLS avec interaction GreenPotSharew × dge

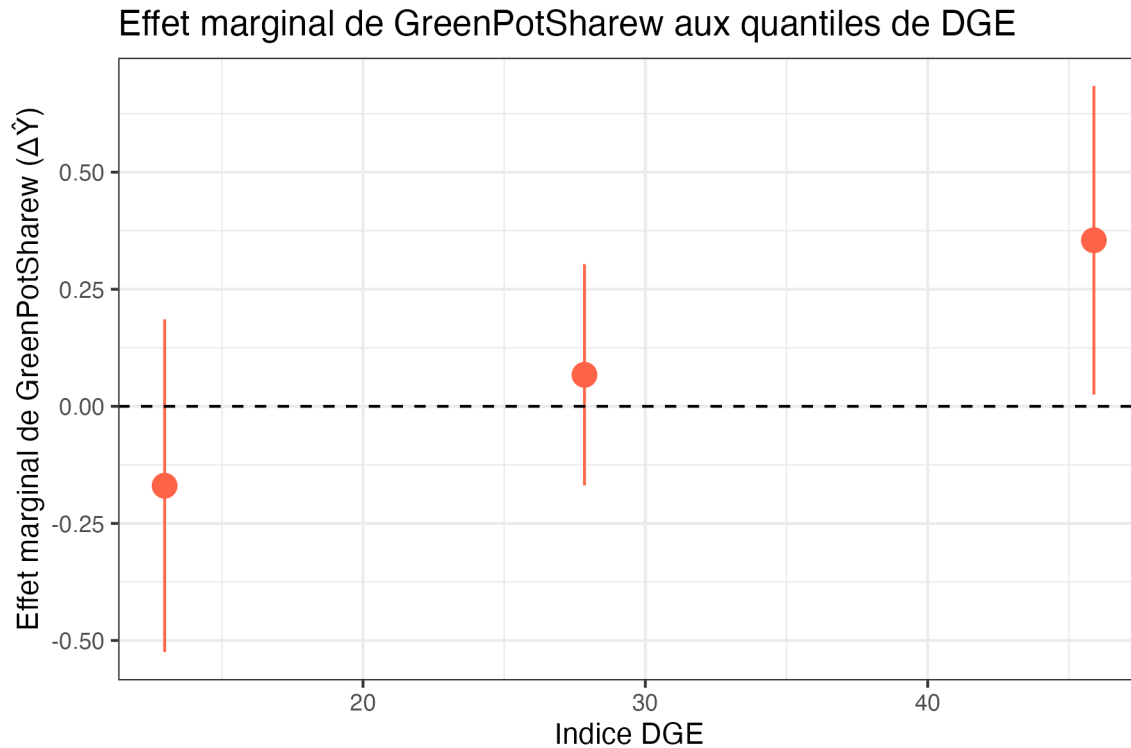
Variable	Estimate	Std. Error	t-value	p-value
$\Delta \ln K$	0.152 ***	0.0307	4.963	1.58×10^{-10}
$\Delta \ln L$	0.040	0.0779	0.515	0.610
GreenPotSharew	-0.376	0.2659	-1.415	0.166
dge	-0.0127 .	0.00675	-1.888	0.067 .
ays	-0.0108	0.00996	-1.088	0.284
gov_exp_pctgdp	-0.00328 ***	0.000854	-3.844	4.60×10^{-4}
open_pct	0.000684	0.000512	1.335	0.190
democracy	-0.0367 *	0.0165	-2.229	0.032 *
access_electricity	-0.000713	0.000673	-1.060	0.296
hhi_va	-0.269	0.2208	-1.219	0.231
GreenPotSharew × dge	0.0159 *	0.00778	2.047	0.047 *

* $p < 0.05$ · $p < 0.1$ ** $p < 0.01$ *** $p < 0.001$

Tableau A.29 - Effets marginaux de GreenPotSharew aux quantiles de DGE

Quantile	DGE	AME	SE	95 % CI
10 %	12.974	-0.170	0.181	[-0.525 ; 0.186]
50 %	27.844	0.067	0.121	[-0.169 ; 0.304]
90 %	45.886	0.355	0.168	[0.025 ; 0.684]

Graphique 8 - Effet marginal de GreenPotSharew aux quantiles de DGE



L'accumulation de capital conserve un effet hautement significatif et positif, tandis que le travail reste non significatif. Le coefficient de `GreenPotSharew` est négatif mais son interaction avec `dge` est positive et significative, indiquant qu'à mesure que DGE augmente, l'effet du verdissement sur la croissance devient plus favorable. Concrètement, l'AME passe de $-0,17$ à 10% de DGE à $+0,36$ à 90% de DGE, bien que seuls les niveaux moyens et élevés soient à la limite de la significativité (la borne inférieure du CI pour le 90% est juste au-dessus de zéro). Les autres variables de contrôle se comportent comme attendu (notamment l'effet négatif des dépenses publiques et le rôle mitigé de la démocratie).

7.2 Caractéristiques géographiques

```
library(fixest)

mod_geo_int <- feols(
  DeltaY ~ deltalnK + DeltaLnL
  + GreenPotSharew * factor(region)
  + self_emp_pct + ays + gov_exp_pctgdp
  + open_pct + democracy + access_electricity
  + hhi_va
```

```

| iso3c + year,
data      = df1_iv,
cluster   = ~iso3c
)

# Résultats
summary(mod_geo_int)

```

Tableau A.30 - Tableau des coefficients d'interaction avec régions

Variable	Estimate	Std. Error	t-value	Pr(> t)	Signif.
GreenPotSharew (région 1, référence)	-0.1171	0.1648	-0.7104	0.4815	
GreenPotSharew × région 2	0.5088	0.2694	1.8884	0.0661	.
GreenPotSharew × région 3	1.2981	0.5181	2.5057	0.0163	*
GreenPotSharew × région 4	0.2018	0.2396	0.8423	0.4045	
(région 5 omise pour colinéarité)	(-)	(-)	(-)	(-)	

Le coefficient principal (-0.117) est l'effet de GreenPotSharew dans la région 1 (catégorie de base). Les coefficients d'interaction mesurent l'écart par rapport à cette base.

On calcule pour chaque région r :

$$AME_r = b_G + b_{G \times r}$$

Tableau A.31 - Effets marginaux par région

Région	AME de GreenPotSharew
1 (réf.)	-0.117
2	0.392 *
3	1.181 **
4	0.085
5 (réf.)	-0.117

La région 3 (coefficient d'interaction significatif) est la seule où l'effet du verdissement sur la croissance est nettement positif et statistiquement différent de zéro. Pour les autres (1, 2 et 4), même si le point estimate devient positif en 2 et 4, l'imprécision empêche de conclure à un effet robuste.

7.3 Qualité institutionnelle

1) Interaction verdissement \times démocratie (FE pays+année).

```
library(fixest)
library(dplyr)
library(knitr)

# 1) Estimation FE-OLS avec interaction GreenPotSharew  $\times$  democracy
mod_int_democracy <- feols(
  # formule : variables + interactions
  DeltaY ~ deltalnK + DeltaLnL
    + GreenPotSharew * democracy
    + self_emp_pct + ays + gov_exp_pctgdp
    + open_pct + access_electricity + hhi_va
  # effets fixes pays et année
  | iso3c + year,
  data = df1_iv %>% filter(!is.na(democracy)),
  cluster = ~iso3c
)

# 2) Résultats
summary(mod_int_democracy)

# 3) Extraction des coefficients et de la VCOV
b <- coef(mod_int_democracy)
V <- vcov(mod_int_democracy)
bG <- b["GreenPotSharew"]
bGI <- b["GreenPotSharew:democracy"]

# 4) Quantiles (10%, 50%, 90%) de democracy
qs_dem <- quantile(df1_iv$democracy, probs = c(0.1, 0.5, 0.9), na.rm = TRUE)

# 5) Construction du tableau des effets marginaux
tbl_democracy <- tibble(
  quantile = c("10 %", "50 %", "90 %"),
  democracy = qs_dem
```

```

) %>%
  rowwise() %>%
  mutate(
    AME      = bG + bGI * democracy,
    Var_AME  = V["GreenPotSharew", "GreenPotSharew"] +
               democracy^2 *
    ↪ V["GreenPotSharew:democracy", "GreenPotSharew:democracy"] +
        2 * democracy * V["GreenPotSharew", "GreenPotSharew:democracy"],
    se       = sqrt(Var_AME),
    conf.lo  = AME - 1.96 * se,
    conf.hi  = AME + 1.96 * se
  ) %>%
  ungroup() %>%
  dplyr::select(quantile, democracy, AME, se, conf.lo, conf.hi)

knitr::kable(
  tbl_democracy,
  digits    = 3,
  caption   = "Effet marginal de GreenPotSharew selon niveaux de démocratie"
)

```

Tableau A.32 - Estimation FE-OLS (interaction avec democracy)

Variable	Estimate	Std. Error	t-value	p-value
$\Delta \ln K$	0.0852	0.0379	2.247	0.0301 *
$\Delta \ln L$	0.0180	0.1097	0.164	0.8704
GreenPotSharew	-0.0817	0.7058	-0.116	0.9084
democracy	-0.0684	0.0502	-1.361	0.1809
self_emp_pct	-0.0001	0.0016	-0.089	0.9291
ays	-0.0138	0.0107	-1.290	0.2043
gov_exp_- pctgdp	-0.0054	0.0015	-3.605	0.0008 ***
open_pct	0.0004	0.0005	0.871	0.3889
access_- electricity	0.0006	0.0007	0.852	0.3994
hhi_va	-0.6030	0.3515	-1.715	0.0938 .
Interaction	0.1183	0.1260	0.939	0.3533
GreenPot- Sharew × democracy				
Observations	259			

Variable	Estimate	Std. Error	t-value	p-value
Pays (iso3c)	42			
Années	16			
RMSE	0.02344			
R ² ajusté	0.83459			
R ² Within	0.54622			

- $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; . $p < 0.1$

Tableau A.33 - Effets marginaux de GreenPotSharew selon niveaux de démocratie

Quantile	democracy	AME	SE	IC 95 %
10 %	2.5	0.214	0.436	[−0.640 ; 1.068]
50 %	5.0	0.510	0.283	[−0.046 ; 1.065]
90 %	7.0	0.746	0.368	[0.025 ; 1.468]

Avec $AME = b + b \cdot (\text{niveau de démocratie})$ et IC 95 % calculé via la méthode delta.

Ces résultats indiquent que l'impact de la part verte sur la croissance économique tend à s'amplifier avec le niveau de démocratie, mais ce résultat n'est pas significatif.

- 2) Sous-échantillons par groupes de démocratie : On segmente les pays en « faible » vs « élevée » démocratie au médian et on compare deux FE-OLS.

```
library(dplyr)
library(fixest)
library(modelsummary)
library(knitr)

# 1) Création de l'indicateur High/Low démocratie
df1_iv <- df1_iv %>%
  mutate(
    high_dem = if_else(
      democracy >= median(democracy, na.rm = TRUE),
      "HighDem", "LowDem"
    )
  )

# 2) Construction du texte de la formule complète (avec les fixed-effects)
fml_text <- paste(
  "DeltaY ~ deltaLnK + DeltaLnL + GreenPotSharew +",
```

```

"self_emp_pct + ays + gov_exp_pctgdp +",
"open_pct + access_electricity + hhi_va |",
"iso3c + year"
)

# 3) Transformation en objet formula
fml_dem_fe <- as.formula(fml_text)

# 4) Estimation FE-OLS par sous-groupe High/Low démocratie
models_dem <- lapply(c("LowDem", "HighDem"), function(g) {
  feols(
    fml_dem_fe,
    data = df1_iv %>% filter(high_dem == g),
    cluster = ~iso3c
  )
})
names(models_dem) <- c("LowDem", "HighDem")

# 5) Tableau comparatif
modelssummary(
  models_dem,
  title = "Effet de GreenPotSharew sur  $\Delta Y$  par niveau de démocratie",
  stars = c('*' = .05, '**' = .01, '***' = .001)
)

# 6) Tableau des effets marginaux (AME) de GreenPotSharew
tbl_dem_sub <- tibble(
  group = names(models_dem),
  estimate = map_dbl(models_dem, ~coef(.x)["GreenPotSharew"]),
  se = map_dbl(models_dem,
    ~sqrt(vcov(.x)["GreenPotSharew", "GreenPotSharew"]))
) %>%
  mutate(
    conf.lo = estimate - 1.96 * se,
    conf.hi = estimate + 1.96 * se
  )

knitr::kable(
  tbl_dem_sub,
  digits = 3,
  caption = "Effet marginal de GreenPotSharew par niveau de démocratie"
)

```

Tableau A.34 - Estimations FE-OLS par sous-groupe

Variable	LowDem	HighDem
	Coef. (SE)	Coef. (SE)
GreenPotSharew	0.433 * (0.196)	0.328 (0.236)
deltalnK	0.140 (0.086)	0.105 ** (0.033)
DeltaLnL	-0.104 (0.190)	0.024 (0.121)
self_emp_pct	0.003 (0.002)	-0.001 (0.002)
ays	0.027 (0.015)	-0.011 (0.013)
gov_exp_pctgdp	-0.004 (0.005)	-0.003 ** (0.001)
open_pct	0.001 (0.001)	0.001 ** (0.000)
access_electricity	-0.000 (0.001)	0.006 ** (0.001)
hhi_va	1.634 (1.209)	-0.540 * (0.208)
Nbr d'obs.	65	194
R² (Within)	0.358	0.711

- $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Tableau A.35 - Effet marginal (AME) de GreenPotSharew

Groupe	Estimate	SE	IC 95 % bas	IC 95 % haut
LowDem	0.433	0.196	0.049	0.816
HighDem	0.328	0.236	-0.135	0.790

Dans les pays peu démocratiques, un point supplémentaire de GreenPotSharew est lié à une augmentation de 0,43 % de la croissance ($p < 0,05$), l'intervalle de confiance à 95 % [0,05 ; 0,82] excluant clairement zéro et attestant d'un effet vert robuste. Dans les démocraties avancées, l'impact demeure positif (+0,33 %) mais n'est pas statistiquement significatif (IC 95 % [-0,14 ; 0,79]), probablement en raison d'une variabilité plus forte de la part verte ou de spécificités régionales non captées. Au final, l'effet du verdissement sur la croissance apparaît plus net et certain dans les contextes moins démocratiques, tandis que son influence dans les régimes démocratiques élevés reste positive mais incertaine, reflétant possiblement des différences dans l'implémentation des investissements verts ou dans les structures économiques selon le régime politique.