```
/*******************************************************************************************
Project Title: ESG Performance and Green Investor Attraction

Description:
This study investigates whether higher ESG (Environmental, Social, and Governance) performance
helps firms attract a greater number of green investors. Using firm-level panel data, this analysis
employs multiple empirical strategies to establish causal relationships and ensure robustness.

Empirical Approach:
- Data validation and diagnostic tests
- Panel data techniques (OLS, FE, RE)
- High-dimensional fixed effects models
- Instrumental variable estimation
- Dynamic panel models (Arellano-Bond)
- Heterogeneous treatment effects
- Extensive robustness checks and sensitivity analyses
- Data visualization and results interpretation

*******************************************************************************************/

// Set working directory
cd "/Users/zouzhaoling/Desktop/Final paper"

//======================== 1. DATA PREPARATION ========================//
// Import data
import excel "ESG.xlsx", first clear
save "Data1.dta", replace
use "Data1.dta", clear

// Document data structure
describe
codebook greennum ESG Size Lev ROA

// Data cleaning
// Remove pre-sample period
drop if year == 2009

// Set up panel structure
xtset stock year
xtdescribe

// Generate industry fixed effects
egen ind = group(Industry2)

// Create log transformations for skewed variables
gen ln_Size = ln(Size)
gen ln_Age = ln(Age)

//======================== 2. DATA VALIDATION ========================//
// Check for missing values and patterns
misstable summarize
misstable patterns greennum ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep
Dual

// Visualize distributions of key variables
foreach var of varlist greennum ESG Size Lev ROA {
    histogram `var', normal title("Distribution of `var'")
    graph export "hist_`var'.png", replace
}

// Create data quality report
mdesc greennum ESG Size Lev ROA BM TobinQ
asdoc sum greennum ESG Size Lev ROA BM TobinQ, save(summary_stats.doc) replace

// Check for outliers
graph box greennum ESG Size, saving(boxplot, replace)

//======================== 3. WINSORIZATION AND VARIABLE CONSTRUCTION ========================//
```

```
// Winsorize continuous variables at 1% and 99% levels
winsor2 greennum ESG ESG_score BloombergESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep, replace cuts(1 99)

// Also create 10/90 percentile winsorization for sensitivity analysis
winsor2 greennum ESG ESG_score BloombergESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep, generate(w1090_) cuts(10 90)

// Create standardized variables
foreach var of varlist ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep {
    egen z_`var' = std(`var')
}

// Generate interaction terms
gen ESG_Size = ESG * Size
gen ESG_Lev = ESG * Lev

// Create categorical variables for heterogeneity analysis
egen ESG_quartile = xtile(ESG), nq(4)
egen Size_quartile = xtile(Size), nq(4)

// Generate year and industry dummies
tab year, gen(year_)
tab ind, gen(ind_)

//======================== 4. DESCRIPTIVE STATISTICS ========================//
// Basic descriptive statistics
estpost tabstat greennum greendum ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual, ///
    statistics(n mean sd min p25 p50 p75 max) columns(statistics)
esttab using "descriptive_stats.rtf", replace ///
    cells("count mean(fmt(3)) sd(fmt(3)) min(fmt(3)) p25(fmt(3)) p50(fmt(3)) p75(fmt(3)) max(fmt(3))") ///
    title("Descriptive Statistics") nonumber label

// Time trends in key variables
preserve
collapse (mean) ESG greennum, by(year)
twoway (line ESG year) (line greennum year, yaxis(2)), ///
    title("Trends in ESG Performance and Green Investors") ///
    ytitle("Average ESG Score") ytitle("Avg. Green Investors", axis(2)) ///
    legend(order(1 "ESG Score" 2 "Green Investors"))
graph export "time_trends.png", replace
restore

//======================== 5. CORRELATION ANALYSIS ========================//
// Correlation matrix with significance stars
pwcorr_a greennum ESG Size Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep, star(.05) bonferroni

// Visual correlation matrix
corr greennum ESG Size Lev ROA TobinQ Cashflow FIXED
matrix C = r(C)
heatplot C, values(format(%3.2f)) color(hcl diverging, intensity(.6)) ///
    title("Correlation Matrix") xlabel(, angle(45)) ylabel(, angle(0))
graph export "correlation_heatmap.png", replace

//======================== 6. MULTICOLLINEARITY DIAGNOSTICS ========================//
// Variance inflation factor analysis
reg greennum ESG Size Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual
vif
estat vif

// Condition index test
coldiag2 ESG Size Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual

//======================== 7. BASELINE REGRESSION MODELS ========================//
// OLS Regression
```

```
ind, vce(robust)
est store ols
estadd local yearFE "Yes"
estadd local indFE "Yes"
estadd local firmFE "No"

// Test for normality of residuals
predict resid, residuals
kdensity resid, normal
graph export "residual_distribution.png", replace
sktest resid

// Test for heteroskedasticity
estat hettest
estat imtest, white

// Test for autocorrelation
xtserial greennum ESG Size Lev ROA TobinQ

// Fixed Effects Model
xtreg greennum ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual i.year, fe vce(robust)
est store fe
estadd local yearFE "Yes"
estadd local indFE "No"
estadd local firmFE "Yes"

// Random Effects Model
xtreg greennum ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual i.year i.ind, re vce(robust)
est store re
estadd local yearFE "Yes"
estadd local indFE "Yes"
estadd local firmFE "No"

// Breusch-Pagan LM test for random effects
xttest0

// Hausman Test for model selection
hausman fe re, sigmamore
local hausman_chi2 = r(chi2)
local hausman_p = r(p)

// High-dimensional Fixed Effects
reghdfe greennum ESG Size Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual, absorb(i.year i.ind) vce(robust)
est store hdfe1
estadd local yearFE "Yes"
estadd local indFE "Yes"
estadd local firmFE "No"

reghdfe greennum ESG Size Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual, absorb(i.year i.stock) vce(robust)
est store hdfe2
estadd local yearFE "Yes"
estadd local indFE "No"
estadd local firmFE "Yes"

//======================== 8. CAUSAL IDENTIFICATION STRATEGIES ========================//
// Instrumental variables approach
// First-stage regression
reg ESG industry_avg_ESG country_regulation Size Lev ROA TobinQ i.year i.ind
test industry_avg_ESG country_regulation
predict ESG_hat

// 2SLS
ivregress 2sls greennum (ESG = industry_avg_ESG country_regulation) Size Lev ROA TobinQ i.year i.ind, first
```

```
est store iv
estadd local yearFE "Yes"
estadd local indFE "Yes"
weakivtest

// Difference-in-differences setup (if applicable)
gen post_regulation = (year >= 2015) // Assuming regulatory change in 2015
gen high_ESG = (ESG > r(p50))
gen did = post_regulation * high_ESG

xtreg greennum did post_regulation high_ESG Size Lev ROA TobinQ i.year, fe vce(robust)
est store did
estadd local yearFE "Yes"
estadd local firmFE "Yes"

// Event study specification
forvalues t = -3/3 {
    if `t' != -1 {
        gen event_`t' = (year == (2015 + `t'))
        gen event_`t'_highESG = event_`t' * high_ESG
    }
}

xtreg greennum event_* Size Lev ROA TobinQ i.year, fe vce(robust)
est store event
coefplot, keep(event_*_highESG) vertical yline(0) xline(3.5) ///
    title("Event Study: Impact of ESG on Green Investors")
graph export "event_study.png", replace

// Dynamic panel model (Arellano-Bond)
xtabond2 greennum L.greennum ESG Size Lev ROA TobinQ, ///
    gmm(L.greennum ESG, lag(2 .)) iv(Size Lev ROA TobinQ i.year) twostep robust
est store dyn
estadd local yearFE "Yes"

//======================== 9. HETEROGENEOUS EFFECTS ========================//
// Interactions with firm characteristics
reg greennum c.ESG##c.Size c.ESG##c.Lev ROA TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual
i.year i.ind, vce(robust)
est store inter

// Marginal effects analysis
margins, dydx(ESG) at(Size=(10(10)50))
marginsplot, title("Marginal Effect of ESG by Firm Size") ///
    ytitle("Effect on Green Investor Number") ///
    xtitle("Firm Size") name(marg1, replace)
graph export "marginal_effects_size.png", replace

// Subsample analysis by industry
eststo clear
levelsof ind, local(industries)
foreach i of local industries {
    reg greennum ESG Size Lev ROA if ind == `i', robust
    est store ind_`i'
}
coefplot ind_*, keep(ESG) xline(0) sort(b) ///
    title("Effect of ESG by Industry")
graph export "industry_effects.png", replace

// Quantile regression to examine effects across the distribution
sqreg greennum ESG Size Lev ROA TobinQ, q(0.25 0.5 0.75) reps(100)
est store qreg

//======================== 10. ROBUSTNESS CHECKS ========================//
// Alternative dependent variables
reg greennum_pct ESG Size Lev ROA TobinQ i.year i.ind, vce(robust)
est store alt_dep1

```

```
foreach esg_var in ESG ESG_score BloombergESG {
    reghdfe greennum `esg_var' Size Lev ROA TobinQ, absorb(i.year i.ind) vce(robust)
    est store `esg_var'_model
}
coefplot ESG_model || ESG_score_model || BloombergESG_model, keep(*ESG*) xline(0) bycoefs ///
    title("Comparison of ESG Measures")
graph export "esg_measures_comparison.png", replace

// Alternative winsorization
reg greennum w1090_ESG Size Lev ROA TobinQ i.year i.ind, vce(robust)
est store alt_winsor

// Placebo tests
reg unrelated_outcome ESG Size Lev ROA TobinQ i.year i.ind, vce(robust)
est store placebo

// Logistic regression model
logit greendum ESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual i.year i.ind
est store logit
margins, at(ESG=(0(10)100))
marginsplot, title("Predicted Probability of Green Investor Presence") ///
    ytitle("Pr(Green Investor Present)") xtitle("ESG Score")
graph export "logit_predicted_prob.png", replace

// Addressing endogeneity: Control function approach
predict v, residuals
reg greennum ESG v Size Lev ROA TobinQ i.year i.ind, vce(robust)
est store cf

//======================== 11. REPORT RESULTS ========================//
// Create comprehensive results table
esttab ols fe re hdfe2 iv dyn using "comprehensive_results.rtf", replace ///
    b(%9.3f) se(%9.3f) star(* 0.10 ** 0.05 *** 0.01) ///
    s(N r2 r2_a yearFE indFE firmFE, label("Observations" "R-squared" "Adjusted R-squared" "Year FE" "Industry FE" "Firm FE")) ///
    mtitles("OLS" "FE" "RE" "HDFE" "IV" "Dynamic") ///
    addnote("Hausman test: Chi2 = `hausman_chi2', p-value = `hausman_p'") ///
    title("Impact of ESG Performance on Green Investor Attraction")

// Coefficient plot for main models
coefplot ols fe re hdfe2 iv, keep(ESG) xline(0) ///
    title("Coefficient of ESG Across Models") ///
    ciopts(recast(rcap)) citop ///
    xlabel(, angle(0)) ylabel(, angle(horizontal))
graph export "model_comparison.png", replace

// Heterogeneous effects summary
esttab inter qreg using "heterogeneous_effects.rtf", replace ///
    b(%9.3f) se(%9.3f) star(* 0.10 ** 0.05 *** 0.01) ///
    title("Heterogeneous Effects of ESG on Green Investor Attraction")

// Robustness checks summary
esttab alt_dep1 ESG_score_model BloombergESG_model alt_winsor placebo logit using "robustness_checks.rtf", replace ///
    b(%9.3f) se(%9.3f) star(* 0.10 ** 0.05 *** 0.01) ///
    title("Robustness Checks for ESG Impact on Green Investor Attraction")

//======================== 12. ADDITIONAL ANALYSES ========================//
// Mediation analysis
reg mediator ESG Size Lev ROA i.year i.ind, vce(robust)
predict mediator_hat
reg greennum ESG mediator_hat Size Lev ROA i.year i.ind, vce(robust)
est store mediation

// Matching analysis (PSM)
psmatch2 high_ESG Size Lev ROA i.ind, outcome(greennum) neighbor(3) common caliper(0.01)
pstest Size Lev ROA, treated(high_ESG) both graph
```

```
graph export "psm_balance.png", replace

psgraph, treated(high_ESG) pscore(_pscore)
graph export "psm_histogram.png", replace

// Testing for parallel trends assumption in DiD
gen trend = year - 2015
gen trend_highESG = trend * high_ESG

xtreg greennum trend trend_highESG if year < 2015, fe
est store parallel

//======================== 13. EXPORT CLEAN DATA FOR FURTHER ANALYSIS ========================//
// Create final analysis dataset
keep stock year greennum greendum ESG ESG_score BloombergESG Size Lev ROA BM TobinQ Cashflow FIXED Growth TOP5 Age Board Indep Dual ind ESG_Size ESG_Lev ESG_quartile Size_quartile

// Add model predictions
predict yhat_fe, xb
label var yhat_fe "Predicted values from FE model"

// Export final dataset
save "ESG_analysis_final.dta", replace
export delimited using "ESG_analysis_final.csv", replace

```