Berlin Firms Comparison

Load required libraries ——————————————————————-

library(tidyverse) library(gt) library(scales)

Load data ——————————————————————————

df <- readRDS("/workspaces/rct25/data/generated/Orbis_Berlin_Data/orbis_panel_berlin.rds")

Ensure df is a data.frame ——————————————————————

stopifnot(is.data.frame(df))

Prepare data ——————————————————————————

df <- df %>% filter(city_native == "Berlin") %>% mutate(total_assets = toas, equity_ratio = shfd / toas, log_total_assets = log1p(toas), group = if_else(postcode == 10437, "Postal Code 10437", "Other Berlin Firms")) %>% filter(!is.na(total_assets), !is.na(equity_ratio))

Count original sample size ——————————————————————-

n_total_original <- nrow(df)

Clean data: remove very small total assets ——————————————————————

df <- df %>% filter(total_assets >= 1000) n_removed_assets <- n_total_original - nrow(df)

Summary statistics ——————————————————————————

summary_stats <- df %>% group_by(group) %>% summarise(Mean_Total_Assets = mean(total_assets, na.rm = TRUE), SD_Total_Assets = sd(total_assets, na.rm = TRUE), Mean_Equity_Ratio = mean(equity_ratio, na.rm = TRUE), Median_Equity_Ratio = median(equity_ratio, na.rm = TRUE), SD_Equity_Ratio = sd(equity_ratio, na.rm = TRUE), n = n(), .groups = "drop")

Extract values explicitly by group ——————————————————————

get_stat <- function(var, group_name) summary_stats %>% filter(group == group_name) %>% pull({{ var }})

mean_total_assets_10437 <- get_stat(Mean_Total_Assets, "Postal Code 10437") mean_total_assets_other <- get_stat(Mean_Total_Assets, "Other Berlin Firms")

mean_equity_ratio_10437 <- get_stat(Mean_Equity_Ratio, “Postal Code 10437”) mean_equity_ratio_other <- get_stat(Mean_Equity_Ratio, “Other Berlin Firms”)

median_equity_ratio_10437 <- get_stat(Median_Equity_Ratio, “Postal Code 10437”) median_equity_ratio_other <- get_stat(Median_Equity_Ratio, “Other Berlin Firms”)

sd_equity_ratio_10437 <- get_stat(SD_Equity_Ratio, “Postal Code 10437”) sd_equity_ratio_other <- get_stat(SD_Equity_Ratio, “Other Berlin Firms”)

sd_total_assets_10437 <- get_stat(SD_Total_Assets, “Postal Code 10437”) sd_total_assets_other <- get_stat(SD_Total_Assets, “Other Berlin Firms”)

n_10437 <- get_stat(n, “Postal Code 10437”) n_other <- get_stat(n, “Other Berlin Firms”)

Statistical tests ————————————————————————-

t_total_assets <- t.test(total_assets ~ group, data = df, var.equal = FALSE) t_equity_ratio <- t.test(equity_ratio ~ group, data = df, var.equal = FALSE) t_sd_equity_ratio <- t.test(df$equity_ratio[df$group == “Postal Code 10437”], df$equity_ratio[df$group != “Postal Code 10437”], var.equal = FALSE) t_sd_total_assets <- t.test(df$total_assets[df$group == “Postal Code 10437”], df$total_assets[df$group != “Postal Code 10437”], var.equal = FALSE) t_median_equity_ratio <- wilcox.test(equity_ratio ~ group, data = df)

Create results table ————————————————————————-

results_table <- tibble(`Postal Code 10437` = c(mean_total_assets_10437, sd_total_assets_10437, mean_equity_ratio_10437, median_equity_ratio_10437, sd_equity_ratio_10437, n_10437), `Other Berlin Firms` = c(mean_total_assets_other, sd_total_assets_other, mean_equity_ratio_other, median_equity_ratio_other, sd_equity_ratio_other, n_other), Difference = c(mean_total_assets_10437 - mean_total_assets_other, sd_total_assets_10437 - sd_total_assets_other, mean_equity_ratio_10437 - mean_equity_ratio_other, median_equity_ratio_10437 - median_equity_ratio_other, sd_equity_ratio_10437 - sd_equity_ratio_other, NA), `P-Value` = c(t_total_assets$p.value, t_sd_total_assets$p.value, t_equity_ratio$p.value, t_median_equity_ratio$p.value, t_sd_equity_ratio$p.value, NA), $Significance = c(case_when(t_total_assets$p.value < 0.01 ~ “**”, t_total_assets$p.value < 0.05 ~ “*”, t_total_assets$p.value < 0.1 ~ “”), TRUE ~ “”), case_when(t_sd_total_assets$p.value < 0.01 ~ “***”, t_sd_total_assets$p.value < 0.05 ~ “**”, t_sd_total_assets$p.value < 0.1 ~ “*”, TRUE ~ “”), case_when(t_equity_ratio$p.value < 0.01 ~ “**”, t_equity_ratio$p.value < 0.05 ~ “*”, t_equity_ratio$p.value < 0.1 ~ “”, TRUE ~ “”), case_when(t_median_equity_ratio$p.value < 0.01 ~ “***”, t_median_equity_ratio$p.value < 0.05 ~ “**”, t_median_equity_ratio$p.value < 0.1 ~ “*”, TRUE ~ “”), case_when(t_sd_equity_ratio$p.value < 0.01 ~ “**”, t_sd_equity_ratio$p.value < 0.05 ~ “*”, t_sd_equity_ratio$p.value < 0.1 ~ “”, TRUE ~ “”),””))

Format results ————————————————————————-

results_table_formatted <- results_table %>% mutate(rowname = c(“Total Assets (Mean)”, “Total Assets (SD)”, “Equity Ratio (Mean)”, “Equity Ratio (Median)”, “Equity Ratio (SD)”, “Number of Firms (N)”), `Postal Code 10437` = case_when(str_detect(rowname, “Equity”) ~ formatC(100 * as.numeric(`Postal Code 10437`), format = “f”, digits = 2, big.mark = “.”, decimal.mark = “,”) %>% paste0(“%”), str_detect(rowname, “Total|Firms”) ~ format(round(as.numeric(`Postal Code 10437`), 0), big.mark = “.”, decimal.mark = “,”, scientific = FALSE), TRUE ~ as.character(`Postal Code 10437`)), `Other Berlin Firms` = case_when(str_detect(rowname, “Equity”) ~ formatC(100 * as.numeric(`Other Berlin Firms`), format = “f”, digits = 2, big.mark = “.”, decimal.mark = “,”) %>% paste0(“%”), str_detect(rowname, “Total|Firms”) ~ format(round(as.numeric(`Other Berlin Firms`), 0), big.mark = “.”, decimal.mark = “,”, scientific = FALSE), TRUE ~ as.character(`Other Berlin Firms`)), Difference = case_when(str_detect(rowname, “Equity”) ~ formatC(100 * as.numeric(Difference), format = “f”, digits = 2, big.mark = “.”, decimal.mark = “,”) %>% paste0(“%”), str_detect(rowname, “Total|Firms”) ~

format(round(as.numeric(Difference), 0), big.mark = ".", decimal.mark = ",", scientific = FALSE), TRUE ~ as.character(Difference)), `P-Value` = case_when(is.na(`P-Value`) ~ "", `P-Value` < 0.001 ~"<0.001", TRUE ~ formatC(`P-Value`, format = "f", digits = 3, decimal.mark = ",")))

Create gt table

gt_table <- results_table_formatted %>% gt(rowname_col = "rowname") %>% tab_header(title = "Comparison of Firms in Postal Code 10437 vs. Other Berlin Firms", subtitle = "Total Assets and Equity Ratios (Most Recent Year)") %>% cols_label(`Postal Code 10437` = "Postal Code 10437", `Other Berlin Firms` = "Other Berlin Firms", Difference = "Difference", `P-Value` = "P-Value", Significance = "Signif.") %>% cols_align(align = "center") %>% tab_source_note(source_note = paste0("Note: Total assets in EUR. Equity ratio = Equity / Total Assets.", "P-values from Welch's t-test (means, SDs) and Wilcoxon rank-sum test (medians).", "Significance levels: * p<0.1, ** p<0.05, *** p<0.01.", "N = number of firms per group. All p-values <0.001 reported as <0.001.", "Sample restricted to Berlin firms with total assets
geq 1,000 EUR (", n_removed_assets, " excluded)."))