

Assignment I. – Panel data

Milan Peter

Student number: 2868506

Email: m.peter2@student.vu.nl

INTRODUCTION

The three-factor CAPM model has basic formula

$$y_{it} = \alpha_i + f_t \beta_i + \epsilon_{it},$$

where α_i is a company-specific intercept, and β_i a company specific set of weights for the three Fama-French factors (Fama and French, 1993). For each stock this relation can be estimated anew, without looking at possible relations between stocks.

However, we might want to check for similarities between the companies, and move to the Panel Data model

$$y_{it} = \alpha_i + f_t \beta + \epsilon_{it}. \tag{1}$$

This equation (1) assumes that the effect of the factors is the same between stocks.

Possibly this assumption is rather strict, but it may be a valid starting point for the current exercise.

To investigate how strict the assumption of a common β is, this year the attention of the exercise is to release this restriction somewhat. That is: Is this β indeed common between all companies? Or would it work better to separate out a specific β per sector, as in

$$y_{it} = \alpha_i + f_t \beta_s + \epsilon_{it}? \tag{2}$$

DATA AND SETUP

I was provided with a large panel dataset on the SP500 stocks, with daily prices on the stocks which are components of the index, together with the Fama-French three factors, and the daily risk-free rate.

For those stocks, as far as available, I also received data on the industry, the sector, and the number of employees. I may disregard industry/employee data; this year only the link of ticker to sector is of interest.

The assignment description can be found in the *Panel Data* PDF, but I copy it in this notebook for convenience. The data is in the *capmff* folder.

The *prices* file contains the prices (adjusted close; source: Yahoo Finance), the *ff* file has the Fama-French 3-factor data with the risk free rate (see also FF Factors , with the daily risk-free rate), and the *sector* information in the last file.

Focus on either the extension to the group of sectors:

- Basic Materials, Communication Services, Consumer Cyclical, Consumer Defensive
- **Energy, Financial Services, Healthcare**
- Industrials, Real Estate, Technology, Utilities

My choice is not free: If I am in group G_n , (where n is the number of *my* group), calculate

$$d = (n \bmod 3) + 1$$

and take sector set d from the list above.

In my case: Group G_{13} has $n = 13$, so $d = (n \bmod 3) + 1 \Leftrightarrow (13 \bmod 3) + 1 \Leftrightarrow 1 + 1 = 2$, hence I (Group G_{13}) will study the sectors Energy, Financial Services, Healthcare as this is the second entry in the list above.

2.1. Data and scope

I began by loading the daily stock price data, sector classifications, and Fama-French factor data. I converted all date fields to proper datetime format and restricted the stock price dataset so that its dates exactly match the factor dates.

From the sector file, I dropped the unnecessary fields on industry, country, and employees, and then filtered the dataset to only include the assigned sectors - Energy, Financial Services, and Healthcare. Based on this filtered list, I kept only the corresponding tickers in the price data.

Next, I separated the risk-free rate from the factor dataset and retained only the three Fama-French factors (Mkt-RF, SMB, HML). Finally, I computed the daily stock excess returns by converting stock prices into daily percentage changes and subtracting the daily risk-free rate, using the equation $y_{it} = R_{it} - R_{ft}$. This ensured that the factor dates and trading dates were aligned and that missing values would not create unwanted artifacts in the sample.

2.2. Panel structure

I reshaped the excess returns into a long panel format with firm identifiers (i), dates (t), and returns (y_{it}). I removed missing return values to keep the dataset clean.

Next, I merged in the Fama-French factors by date (f_t) and dropped any rows with missing factor values. I then added the sector identifiers (s) by merging with the sector file and created dummy variables so that each firm was mapped to its sector.

After that, I set up the panel structure with a multi-index of ticker and date, as required for panel regressions. Finally, I generated sector-factor interaction terms by multiplying each Fama-French factor with the corresponding sector dummy variable ($f_t \times D_{is}$), so the model can capture sector-specific factor loadings.

EXERCISE

Each group gets a specific target of investigation. I am told to look at a specific set of sectors or years. For this purpose, the following steps are advised to be taken:

- Check the precise setup of the model (as (2)). Is it too general, should all data be kept together, or should the panel setup be simplified?
- Estimate with the sector specific β_s , and compare the results with the common β estimate in (1). What can be told from the results?
- Can it be tested whether the β_s are different indeed? How? Is it significantly better to go back to (1), with a common β ?

My task is to discuss in my report what the results mean to me, whether I used Pooled Regression, Fixed Effects, or Random Effects, with/without robust standard errors, through what procedure/package etc.

3.1. Estimation

I start the assignment by estimating the panel data models. First, I estimate the pooled OLS regression model as a benchmark. Then, I continue with extending the model to include the sector-factor interaction terms. Next, I estimate the fixed effects (FE) model using the within transformation to control for unobserved firm-specific effects. Finally, I estimate the random effects (RE) model via GLS, which assumes that firm-specific effects are uncorrelated with the regressors. In my implementation, I ran the pooled regression with `statsmodels`, regressing excess returns on the three Fama–French factors with a constant term. For the fixed effects and random effects models, I used the `linearmodels` package. This package provides a convenient way to estimate FE (applying the within transformation) and RE models using GLS.

3.1.1. Pooled OLS Regression

I began with the pooled OLS regression as a benchmark model. Pooled OLS combines all firm-time observations into a single dataset and estimates one regression line across them. This means that I assumed:

- one common intercept for all firms,
- one common set of slope coefficients for the factors,
- and no distinction between firms or time periods in the error structure.

The regression equation is:

$$y_{it} = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + u_{it}, \quad (3.1)$$

where y_{it} is the excess return of firm i on day t , and the three factors are the Fama-French market, size, and value factors.

This model is easy to estimate and provides a baseline for comparison. However, it ignores firm- or time-specific heterogeneity: differences in average returns across firms (firm fixed effects) or across time (time fixed effects) are captured only in the error term. As a result, if unobserved firm characteristics are correlated with the regressors, the estimates will be biased.

The results give me a simple benchmark of how well the factors explain stock returns, but I expect it to be outperformed by more refined panel models (fixed effects and random effects) that account for firm-level heterogeneity. See Table a.1 in appendix.

I find a negative and statistically significant intercept $\alpha = -0.0048$ and positive, highly significant loadings on the market, size, and value factors, indicating all three factors contribute to explaining daily excess returns in the pooled OLS benchmark. The model is jointly significant with $R^2 = 0.307$ and an enormous F -statistic ($p = 0.000$), so the factors have strong explanatory power in aggregate.

- Intercept: $\alpha = -0.0048$, highly significant ($t = -204.35$, $p = 0.000$), implying a negative average excess return unexplained by the factors in this pooled specification.
- Market factor: $\beta_{MKT} = 0.0099$, very precisely estimated ($t = 457.53$, $p = 0.000$), consistent with a strong positive sensitivity to the market excess return.
- Size factor: $\beta_{SMB} = 0.0009$, positive and significant ($t = 21.81$, $p = 0.000$), indicating a small but nontrivial tilt toward smaller firms on average.
- Value factor: $\beta_{HML} = 0.0040$, positive and significant ($t = 133.51$, $p = 0.000$), suggesting returns co-move positively with the value factor.
- Overall fit and joint significance: $R^2 = 0.307$ with $F \approx 8.283 \times 10^4$ ($p = 0.000$), confirming the factors are jointly informative in this pooled model, albeit with substantial residual variation remaining.

3.1.2. Pooled OLS with Sector-Specific Slopes (Interaction Model)

I extended the pooled OLS model by allowing the factors to vary across sectors. To do this, I created interaction terms between each of the Fama-French factors and the sector dummy variables, ($f_t \times D_{is}$). This setup means that instead of estimating one common set of slopes, the regression estimates separate slope coefficients for each sector.

Formally, the model is:

$$y_{it} = \sum_{s=1}^S D_{is} (\beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t) + u_{it}, \quad (3.2)$$

where D_{is} indicates whether firm i belongs to sector s , and each sector has its own slope vector $(\beta_{1s}, \beta_{2s}, \beta_{3s})$. Because the slopes are calculated for each sector separately, using dummy variables, the model omits the sector-specific intercepts, α . Thus, I avoid the dummy variable trap, and the model is interpretable.

In practice, this specification allows me to capture heterogeneity in how different sectors respond to the Fama-French factors. For example, the market factor might matter more for Financial Services, while the value factor might be more relevant for Energy. The trade-off is that the model becomes more parameter-heavy, since each sector requires a separate set of factor coefficients.

The results are in Table a.2 in appendix. I find that the sector-specific loadings are all jointly significant, with uncentered $R^2 = 0.310$ and a very large F -statistic ($p = 0.000$), confirming that allowing slopes to vary by sector adds explanatory power in this pooled specification.

- Market beta: Energy (0.0108), Financial Services (0.0107), and Healthcare (0.0081) are all positive and precisely estimated, with Healthcare's market sensitivity noticeably lower than Energy and Financials (all $p = 0.000$).
- Size (SMB): The tilt is positive and significant in all three sectors – largest in Energy (0.0026) and smaller but still significant in Financials (0.0008) and Healthcare (0.0008) (all $p = 0.000$).
- Value (HML): Strong positive exposures in Energy (0.0101) and Financials (0.0061), but a small negative, yet significant, exposure in Healthcare (-0.0005), indicating a relative growth tilt there (all $p = 0.000$).
- Precision: All reported t -statistics are large in magnitude, and every coefficient is statistically significant at conventional levels, underscoring clear cross-sector heterogeneity in factor sensitivities within this pooled framework.

3.1.3. Fixed Effects Estimation (FE) using Within Transformation

I next estimated a fixed effects model, which controls for firm-specific heterogeneity. The idea is that each firm may have its own average return level due to characteristics that do not change over time (e.g., business model, risk exposure, management style). If these unobserved firm traits are correlated with the factors, pooled OLS would give biased estimates.

The fixed effects specification is:

$$y_{it} = \alpha_i + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + u_{it}, \quad (3.3)$$

where α_i is a firm-specific intercept.

Instead of including a dummy variable for each firm (the least squares dummy variable approach), I applied the within transformation: I subtracted each firm's mean from its time series. This demeaning removes the firm intercepts, leaving only within-firm variation:

$$\ddot{y}_{it} = y_{it} - \bar{y}_i, \quad \dot{X}_{it} = X_{it} - \bar{X}_i,$$

and the regression becomes:

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + \ddot{u}_{it}.$$

This way, the model estimates how deviations in a firm's returns from its own mean are explained by deviations in the factors from their means. The advantage is that any time-invariant firm characteristics are controlled for automatically, without estimating them explicitly.

Additionally, I computed clustered standard errors at the firm level. The reason is that residuals for a given firm are likely to be serially correlated and heteroskedastic over time - shocks to a firm's return today may carry over into future days. If I ignored this dependence and used conventional standard errors, the inference would be biased, typically underestimating the true variability of the estimates and overstating statistical significance.

The results are in Table a.3 in appendix. I find that with firm fixed effects and clustered SEs, the market, size, and value loadings remain positive and highly significant, with $R^2_{\text{within}} = 0.3077$ and a large robust F -statistic ($p = 0.000$) indicating strong explanatory power in within-firm variation.

- Intercept: the estimated constant is negative and precisely estimated (-0.0048 , $p = 0.000$), reflecting a negative average excess return after controlling for firm effects in this specification.
- Market beta: 0.0099 with a large t -statistic (51.68 , $p = 0.000$), confirming a strong positive sensitivity to the market factor within firms over time.
- Size (SMB): 0.0009 and statistically significant ($t = 4.38$, $p = 0.000$), indicating a modest but reliable positive within-firm exposure to the size factor.
- Value (HML): 0.0040 and significant ($t = 8.96$, $p = 0.000$), implying returns co-move positively with value after removing firm-specific means.
- Model fit and inference: $R^2_{\text{within}} = 0.3077$ and $R^2_{\text{overall}} = 0.3072$, with a robust F -statistic of 1336.7 ($p = 0.000$), showing strong joint significance under firm-clustered standard errors.
- Poolability: the F -test rejects equal intercepts across firms ($F = 8.699$, $p = 0.000$), so I conclude firm fixed effects are needed over pooled OLS.

3.1.4. Random Effects Estimation (RE)

I also estimated a random effects model, which takes a different approach to firm-specific heterogeneity. Instead of treating the firm intercepts as fixed parameters to be eliminated, I assumed that they are random variables drawn from a distribution with mean zero and that they are uncorrelated with the regressors.

The specification is:

$$y_{it} = a + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + (\alpha_i + u_{it}), \quad (3.4)$$

where a is the overall intercept, α_i is the firm-specific random effect, and u_{it} is the idiosyncratic error.

Because the error term now contains two components - a firm-specific effect and an idiosyncratic shock - ordinary least squares is no longer efficient. Instead, the model is estimated

using generalized least squares (GLS), which accounts for the correlation structure in the composite error.

The advantage of random effects is that it can be more efficient than fixed effects, since it uses both the within-firm and between-firm variation in the data. However, this efficiency gain relies on the strong assumption that the regressors are uncorrelated with the firm-specific effects. If this assumption fails, RE estimates will be biased.

I used clustered standard errors at the firm level for similar reasons as with the fixed effects model to account for the fact that residuals within the same firm may be serially correlated and heteroskedastic over time. Clustering ensures that inference remains valid in the panel setting by allowing arbitrary dependence of errors within firms, while still assuming independence across firms.

The results are in Table a.4 in appendix. I find that the random effects estimates with firm-clustered standard errors show a negative and significant intercept alongside positive, highly significant loadings on market, size, and value, indicating strong within- and overall explanatory power for excess returns. The model fit is solid with $R^2_{\text{within}} = 0.3077$, $R^2_{\text{overall}} = 0.3072$, and a large robust F -statistic ($p = 0.000$) confirming joint significance.

- Intercept: $\alpha = -0.0049$, precisely estimated and strongly negative ($t = -77.82$, $p = 0.000$), consistent with a negative average excess return under the RE specification.
- Market beta: $\beta_{\text{MKT}} = 0.0099$ with a very large t -statistic (51.68 , $p = 0.000$), confirming a strong positive sensitivity to market excess returns.
- Size (SMB): $\beta_{\text{SMB}} = 0.0009$, positive and statistically significant ($t = 4.39$, $p = 0.000$), indicating a modest but reliable size tilt.
- Value (HML): $\beta_{\text{HML}} = 0.0040$, positive and significant ($t = 8.96$, $p = 0.000$), implying returns co-move with the value factor.
- Fit and inference: $R^2_{\text{within}} = 0.3077$, $R^2_{\text{overall}} = 0.3072$, and robust $F = 1336.4$ ($p = 0.000$), showing strong joint significance under clustered SEs and coefficients that closely mirror the FE estimates in magnitude and significance.

3.2. Comparing results

In this section, I will compare the results of the 4 different methods for the panel data set. First, I inspect the model statistics, then the parameter estimates.

3.2.1. Model statistics

All four models deliver very similar explanatory power around 31%, with the interaction model edging the highest uncentered R-squared while fixed and random effects match on within and overall fit; however, pooled OLS is preferred by AIC/BIC among the OLS variants, and FE/RE provide the panel-decomposed diagnostics and clustered inference advantages. The robust F-statistics are huge and $p = 0.000$ in every case, so joint significance holds across specifications. See Table 3.1 for a summary of key statistics across the four models.

Fit and power

- R-squared: Pooled 0.3072, Interaction 0.3104, FE 0.3077, RE 0.3075, implying only a marginal uplift from sector interactions and virtually identical fit for FE/RE at the aggregate level.

- Within/Overall: FE and RE both post $R^2_{\text{within}} = 0.3077$ and $R^2_{\text{overall}} = 0.3072$, indicating that the factors explain within-firm time variation similarly under either estimator.

Information criteria

- Among OLS variants, pooled OLS has lower (better) AIC/BIC than the interaction model despite a slightly lower R-squared, reflecting the penalty for additional sector-factor parameters in the interaction specification.
- Log-likelihoods: FE is highest (1,469,008) followed by RE (1,468,557), pooled (1,468,361), and interaction (1,458,184), but AIC/BIC are only available for the OLS models, so I compare those directly there.

Panel diagnostics

- Negative between R-squared for FE (-0.0136) and RE (-0.0054) suggests the factors capture within-firm dynamics but do not explain cross-firm differences in average returns, consistent with strong entity heterogeneity in intercepts.
- Identical within R-squared for FE and RE, together with nearly identical slopes reported earlier, indicates RE gains come mainly from efficiency assumptions rather than improved in-sample fit.

Inference and tests

- All models show massive F-statistics with $p = 0.000$, confirming joint significance of included regressors under each estimator.

Table 3.1: Model comparison statistics

	Pooled OLS	Interaction OLS	Fixed Effects	Random Effects
R-squared	0.307	0.31	0.308	0.307
Adj. R-squared	0.307	0.31	–	–
R-squared (Within)	–	–	0.308	0.308
R-squared (Between)	–	–	-0.0136	-0.0054
R-squared (Overall)	–	–	0.307	0.307
Log-Likelihood	1.47e+06	1.46e+06	1.47e+06	1.47e+06
AIC	-2.94e+06	-2.92e+06	–	–
BIC	-2.94e+06	-2.92e+06	–	–
F-statistic	8.28e+04	2.8e+04	8.3e+04	8.3e+04
Prob(F-statistic)	0	0	0	0
Observations	560418	560418	560418	560418
Entities	–	–	150	150
Time Periods	–	–	3.9e+03	3.9e+03
Estimator/Notes	OLS	OLS with interactions	PanelOLS, clustered SEs	RandomEffects, clustered SEs

3.2.2. Parameter estimates

All core factor betas are virtually identical across pooled OLS, FE, and RE, and they are uniformly highly significant with $p \approx 0.000$; sector interactions reveal economically meaningful heterogeneity by sector while remaining strongly significant throughout. In short, I see stable signs and magnitudes for MKT, SMB, and HML across estimators, with interactions amplifying or flipping exposures in specific sectors (notably HML in

Healthcare). See Table 3.2 for a summary of key parameter estimates across the four models.

Core factors (pooled vs FE vs RE)

- Const: Pooled $-0.0048(p \approx 0.000)$, FE $-0.0048(p \approx 0.000)$, RE $-0.0049(p \approx 0.000)$, all indicating a negative average excess return under each specification.
- Market: $\beta_{\text{MKT}} \approx 0.0099$ with $p \approx 0.000$ in Pooled, FE, and RE, confirming a strong, stable market sensitivity across estimators.
- Size: $\beta_{\text{SMB}} \approx 0.0009$ with $p \approx 0.000$ across Pooled, FE, and RE, a modest but robust size tilt in all models.
- Value: $\beta_{\text{HML}} \approx 0.0040$ with $p \approx 0.000$ across Pooled, FE, and RE, indicating consistent positive value exposure.

Sector interactions (betas with p-values)

- Energy: $\beta_{\text{MKT,EN}} = 0.0108(p = 0.000)$, $\beta_{\text{SMB,EN}} = 0.0026(p = 0.000)$, $\beta_{\text{HML,EN}} = 0.0101(p = 0.000)$, showing high market and value sensitivity plus a sizeable size tilt.
- Financial Services: $\beta_{\text{MKT,FS}} = 0.0107(p = 0.000)$, $\beta_{\text{SMB,FS}} = 0.0008(p = 0.000)$, $\beta_{\text{HML,FS}} = 0.0061(p = 0.000)$, indicating strong market and positive value exposures with a smaller size tilt.
- Healthcare: $\beta_{\text{MKT,HC}} = 0.0081(p = 0.000)$, $\beta_{\text{SMB,HC}} = 0.0008(p = 0.000)$, $\beta_{\text{HML,HC}} = -0.0005(p = 0.000)$, showing high market sensitivity, modest size tilt, and a statistically significant negative value (growth) exposure.

I read these results as stable, highly significant factor loadings across pooled OLS, FE, and RE, with sector interactions uncovering meaningful cross-sector differences - especially stronger value exposure in Energy and Financials and a negative value tilt in Healthcare - all with $p \approx 0.000$.

Table 3.2: Model comparison parameter estimates

	Pooled OLS	Interaction OLS	Fixed Effects	Random Effects
const	-0.0048 (0.0000)	—	-0.0048 (0.0000)	-0.0049 (0.0001)
MKT	0.0099 (0.0000)	—	0.0099 (0.0002)	0.0099 (0.0002)
SMB	0.0009 (0.0000)	—	0.0009 (0.0002)	0.0009 (0.0002)
HML	0.0040 (0.0000)	—	0.0040 (0.0004)	0.0040 (0.0004)
MKT_ENERGY	—	0.0108 (0.0001)	—	—
SMB_ENERGY	—	0.0026 (0.0001)	—	—
HML_ENERGY	—	0.0101 (0.0001)	—	—
MKT_FINSERVICES	—	0.0107 (0.0000)	—	—
SMB_FINSERVICES	—	0.0008 (0.0001)	—	—
HML_FINSERVICES	—	0.0061 (0.0000)	—	—
MKT_HEALTHCARE	—	0.0081 (0.0000)	—	—
SMB_HEALTHCARE	—	0.0008 (0.0001)	—	—
HML_HEALTHCARE	—	-0.0005 (0.0000)	—	—

Note: Parameter estimates with Standard errors in parentheses.

3.3. Comparing models using statistical tests

To compare the different models, I performed a series of statistical tests. First, I used F-tests to compare nested models: pooled OLS vs. pooled OLS with sector-specific slopes (interaction), fixed effects vs. pooled OLS, and fixed effects vs. pooled OLS with sector-specific slopes (interaction). These tests check whether the additional parameters in the more complex model significantly improve the fit. Next, I used the Hausman test to compare the fixed effects and random effects models. This test checks whether the difference in coefficients between the two models is systematic. If the test indicates correlation between regressors and firm effects, fixed effects is the consistent choice; otherwise, random effects can be used for efficiency.

3.3.1. *F-test: Pooled OLS vs Pooled OLS with Sector-Specific Slopes (Interaction)*

This test checks whether slope coefficients differ significantly between sectors.

- **Null hypothesis (H_0):** all sectors share the same slopes (β).
- **Alternative (H_1):** at least one sector has different slopes (β_s).

If I reject H_0 , it means the interaction model (sector-specific betas) explains more variation than the simple pooled model.

I do not reject H_0 : the sector-specific slopes do not improve fit over pooled slopes ($F \approx -134.11$, $df1 = 149$, $df2 = 560265$, $p \approx 1$), so the simple pooled specification suffices for this test. The interaction model's residual sum of squares is larger than the pooled model's, producing a negative F-statistic, which is consistent with no incremental explanatory power from adding sector-factor interactions here. Given this outcome, I retain the pooled slopes as my benchmark and do not adopt the parameter-heavier interaction specification for this dataset and scope.

3.3.2. *F-test: Fixed Effects vs. Pooled OLS*

Next, I want to test if firm-specific intercepts (α_i) are needed.

- **Null hypothesis (H_0):** all intercepts are equal ($\alpha_1 = \alpha_2 = \dots = \alpha_M$) \rightarrow pooled OLS is valid.
- **Alternative (H_1):** at least one α_i differs \rightarrow fixed effects needed.

If I reject H_0 , it indicates that controlling for firm-specific effects significantly improves the model fit, justifying the use of fixed effects.

I reject H_0 : firm-specific intercepts matter ($F = 8.6990$, $df1 = 149$, $df2 = 560265$, $p = 8.39e - 182$), so fixed effects are preferred over pooled OLS for this panel. This indicates intercepts α_i differ across firms and accounting for them significantly improves fit by removing time-invariant heterogeneity from the error term. The result is consistent with the poolability test reported in the FE output, reinforcing the choice of FE for inference on within-firm factor effects in this dataset.

3.3.3. Hausman Test: Fixed Effects vs. Random Effects

To decide between FE and RE, I rely on the **Hausman test**, which checks whether the difference in coefficients between the two models is systematic. If the test indicates correlation between regressors and firm effects, fixed effects is the consistent choice; otherwise, random effects can be used for efficiency.

The Hausman test checks whether the RE assumption (regressors uncorrelated with firm effects) holds.

- **Null hypothesis (H_0):** RE is consistent and efficient ($\hat{\beta}_{RE} \approx \hat{\beta}_{FE}$).
- **Alternative (H_1):** RE is inconsistent (correlation between X and α_i), so FE is preferred.

If I reject H_0 , it suggests that the RE assumption is violated, and FE should be used for consistent estimates.

I do not reject H_0 in the Hausman test, so I treat Random Effects as consistent and efficient relative to Fixed Effects for this panel ($H = 0.0013$, $df = 3$, $p = 1$). Given the negligible FE–RE coefficient differences and identical within fit, I prefer RE for efficiency while retaining clustered standard errors for valid inference. Since FE and RE deliver nearly identical betas and R^2_{within} , I select RE for reporting and inference to exploit both within- and between-variation efficiently under the supported assumption.

3.4. Conclusion

A three-factor panel model explains within-firm daily excess returns well, fixed effects are required over pooled OLS, and random effects are supported by the Hausman test and preferred for efficiency; sector-specific slopes are individually significant but do not improve fit jointly over pooled slopes in this sample. Taken together, common factor betas are stable across estimators, firm intercept heterogeneity is material, and the RE specification with clustered errors is the recommended baseline for reporting and inference here.

Is one common β across firms reasonable, or should β vary by sector as in $y_{it} = \alpha_i + f_t\beta_s + \epsilon_{it}$? The pooled baseline delivers $R^2 \approx 0.307$, and allowing sector-specific slopes raises uncentered R^2 slightly to about 0.310, but the nested F -test against pooled slopes is not significant overall (negative F , $p \approx 1$), so I do not adopt sector-specific betas as a superior general specification, despite significant sector coefficients individually.

Do firm-specific intercepts matter? Yes, the F -test strongly rejects poolability in favor of fixed effects ($F \approx 8.699$, $p \approx 8.39e - 182$), indicating that firms differ in average returns and those differences need to be controlled to avoid bias in slope estimates.

FE vs RE: which to use? The Hausman test does not reject RE ($H \approx 0.0013$, $p = 1$), and FE and RE yield practically identical betas and R^2_{within} , so I prefer Random Effects.

REFERENCES

- [1] Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33.1, pp. 3–56.
- [2] Brooks, C. (2019). *Introductory Econometrics for Finance*. 4th edn. Cambridge: Cambridge University Press.
- [3] C.S. Bos (2025). *Econometrics for Quantitative Finance - Lecture Slides*. Vrije Universiteit Amsterdam, School of Business and Economics.

APPENDIX

Table a.1: Pooled OLS model results

Dep. Variable:	EXCESSRETURN	R-squared:	0.307
Model:	OLS	Adj. R-squared:	0.307
Method:	Least Squares	F-statistic:	8.283e+04
Date:	Wed, 24 Sep 2025	Prob (F-statistic):	0.00
Time:	15:37:44	Log-Likelihood:	1.4684e+06
No. Observations:	560418	AIC:	-2.937e+06
Df Residuals:	560414	BIC:	-2.937e+06
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0048	2.36e-05	-204.353	0.000	-0.005	-0.005
MKT	0.0099	2.16e-05	457.527	0.000	0.010	0.010
SMB	0.0009	3.95e-05	21.812	0.000	0.001	0.001
HML	0.0040	3e-05	133.509	0.000	0.004	0.004

Omnibus:	161005.541	Durbin-Watson:	1.668
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21632897.309
Skew:	0.144	Prob(JB):	0.00
Kurtosis:	33.436	Cond. No.	1.97

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table a.2: Sector-specific OLS model results

Dep. Variable:	EXCESSRETURN	R-squared (uncentered):	0.310
Model:	OLS	Adj. R-squared (uncentered):	0.310
Method:	Least Squares	F-statistic:	2.803e+04
Date:	Wed, 24 Sep 2025	Prob (F-statistic):	0.00
Time:	15:37:47	Log-Likelihood:	1.4582e+06
No. Observations:	560418	AIC:	-2.916e+06
Df Residuals:	560409	BIC:	-2.916e+06
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
MKT_ENERGY	0.0108	5.79e-05	187.080	0.000	0.011	0.011
SMB_ENERGY	0.0026	0.000	24.571	0.000	0.002	0.003
HML_ENERGY	0.0101	8.01e-05	126.118	0.000	0.010	0.010
MKT_FINSERVICES	0.0107	3.26e-05	326.535	0.000	0.011	0.011
SMB_FINSERVICES	0.0008	5.97e-05	12.908	0.000	0.001	0.001
HML_FINSERVICES	0.0061	4.54e-05	134.895	0.000	0.006	0.006
MKT_HEALTHCARE	0.0081	3.46e-05	235.191	0.000	0.008	0.008
SMB_HEALTHCARE	0.0008	6.33e-05	12.287	0.000	0.001	0.001
HML_HEALTHCARE	-0.0005	4.81e-05	-9.474	0.000	-0.001	-0.000

Omnibus:	161576.103	Durbin-Watson:	1.539
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21802448.278
Skew:	0.157	Prob(JB):	0.00
Kurtosis:	33.555	Cond. No.	3.47

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table a.3: Fixed Effects model results

Dep. Variable:	EXCESSRETURN	R-squared:	0.3077
Estimator:	PanelOLS	R-squared (Between):	-0.0136
No. Observations:	560418	R-squared (Within):	0.3077
Date:	Wed, Sep 24 2025	R-squared (Overall):	0.3072
Time:	15:37:51	Log-likelihood	1.469e+06
Cov. Estimator:	Clustered		
		F-statistic:	8.3e+04
Entities:	150	P-value	0.0000
Avg Obs:	3736.1	Distribution:	F(3,560265)
Min Obs:	315.00		
Max Obs:	3895.0	F-statistic (robust):	1336.7
		P-value	0.0000
Time periods:	3895	Distribution:	F(3,560265)
Avg Obs:	143.88		
Min Obs:	130.00		
Max Obs:	150.00		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
	const	-0.0048	8.352e-06	-576.49	0.0000	-0.0048
	MKT	0.0099	0.0002	51.680	0.0000	0.0095
	SMB	0.0009	0.0002	4.3779	0.0000	0.0005
	HML	0.0040	0.0004	8.9640	0.0000	0.0031

F-test for Poolability: 8.6990

P-value: 0.0000

Distribution: F(149,560265)

Included effects: Entity

Table a.4: Random Effects model results

Dep. Variable:	EXCESSRETURN	R-squared:	0.3075
Estimator:	RandomEffects	R-squared (Between):	-0.0054
No. Observations:	560418	R-squared (Within):	0.3077
Date:	Wed, Sep 24 2025	R-squared (Overall):	0.3072
Time:	15:37:55	Log-likelihood	1.469e+06
Cov. Estimator:	Clustered		
		F-statistic:	8.297e+04
Entities:	150	P-value	0.0000
Avg Obs:	3736.1	Distribution:	F(3,560414)
Min Obs:	315.00		
Max Obs:	3895.0	F-statistic (robust):	1336.4
		P-value	0.0000
Time periods:	3895	Distribution:	F(3,560414)
Avg Obs:	143.88		
Min Obs:	130.00		
Max Obs:	150.00		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
	const	-0.0049	6.268e-05	-77.817	0.0000	-0.0050
	MKT	0.0099	0.0002	51.676	0.0000	0.0095
	SMB	0.0009	0.0002	4.3902	0.0000	0.0005
	HML	0.0040	0.0004	8.9628	0.0000	0.0031