Application of factor models to risk premium estimation

A Thesis Submitted for the Degree of Doctor of Philosophy

By

Vittorio Penco

Department of Mathematics, Brunel University London

2025

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Cormac Lucas, for his guidance and support throughout this research.

I am also grateful to Prof. Robert Brooks, whose research was essential to the development of this thesis and who generously provided timely feedback on my paper review request.

My thanks also go to the review panel (Dr. Elena Boguslavskaya, Dr. Dalia Chakrabarty, Prof. Paresh Date, Dr. Jiawei Lim) for their annual evaluations and constructive feedback.

Finally, I wish to thank Dr. Diana Roman and Dr. Andrea Mazzoran, whose advice helped to better shape this study.

Abstract

We provide a rigorous mathematical approach to the beta-pricing model, starting from the standard two-step cross-sectional regression, through Nonlinear Seemingly Unrelated Regression (NSUR) and Generalized Method of Moments (GMM), and finally compare the results with several linear approximation methods. The use of the linear approximation applied to a single-factor nonlinear system of equations is new in the literature and is one of the major contributions of this work. Our results show that, in the presence of heavy-tailed distributions, the L1-norm methods proposed in this study are more appropriate (exhibiting lower bias and variance) for risk price estimation than traditional L2-norm approaches.

It is also the first time that the Capital Asset Pricing Model (CAPM) is applied systematically to compare the integration and segmentation between different markets and a given portfolio set. Our study, Penco and Lucas (2024), applies a two-factor integration model to the economies of Asia, Europe, Japan and North America, showing integration between the European and North American economies.

We also extend the integration model to commodity markets. To capture more accurately the cross-sectional pricing of commodity risk we use the Cochrane factor mimicking approach and compare the results with alternative dependence-based integration measures using copulas. We show how the copula correlation between the Stochastic Discount Factor (SDF) and returns differentiates the contribution of joint dependence from the contribution of the risk prices.

Finally, we introduce a penalised p-value Fama-MacBeth Generalized Least Squares (GLS) regularisation, which provides several advantages over other methods as it ensures that retained factors contribute not only to statistical fit but also to risk pricing. Unlike other approaches, this method regularises the pricing kernel directly. Factors that lack significance or explanatory power are penalised and removed, while priced and relevant sources of risk are preserved.

Contents

List of Abbreviations

Abbreviation	Full Form / Description
BFGS	Broyden–Fletcher–Goldfarb–Shanno algorithm
CAPM	Capital Asset Pricing Model
CBC	COIN-OR Branch and Cut
CDF	Cumulative Distribution Function
CPU	Central Processing Unit
CS	Cross-Section
CVXPY	Convex optimisation in Python
DCP	Disciplined Convex Programming
EIV	Error-In-Variables
ETF	Exchange Traded Fund
FF	Fama French Equities experiments
FM	Fama-MacBeth Regression
GLS	Generalized Least Squares
GMM	Generalized Method of Moments
GRB	Gurobi
IR	Integration Rejected
IV	Instrumental Variable
LASSO	Least Absolute Shrinkage and Selection Operator model
LP	Linear Programming
LU	Lower-Upper Decomposition
MC	Monte Carlo
MILP	Mixed Integer Linear Programming
MIQP	Mixed Integer Quadratic Programming
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
NSUR	Nonlinear Seemingly Unrelated Regression

Continued on next page

Abbreviation	Description
OLS	Ordinary Least Squares
OSQP	Operator Splitting Quadratic Program
QP	Quadratic Programming
PCA	Principal Component Analysis
PDF	Probability Density Function
PI	Partial Integration
PIT	Probability Integral Transform
PS	Partial Segmentation
QCQP	Quadratically Constrained Quadratic Programming
RMSE	Root Mean Squared Error
RYS	Rolling Yield Synthetic
SAE	Sum of Absolute Errors
SAS	Statistical Analysis System
SDF	Stochastic Discount Factor
SE	Standard Error
SLSQP	Sequential Least Squares Quadratic Programming
SR	Segmentation Rejected
SSR	Sum of Squared Residuals
SUR	Seemingly Unrelated Regression
TI	Total Integration
TS	Total Segmentation
VIX	Volatility Index

List of Tables

List of Figures

List of Experiments

For each of the experiments, we use two sets of data to test our models:

- an application to global economies' equity markets,
- an application to commodities markets.

We implement all experiments in Python or SAS. The code is available on GitHub at: https://github.com/vittoriopenco/thesis (access available on request).

Chapter 1

Introduction

1.1 Research Background and Context

Risk premium estimation is a central issue in finance. Its study describes the relation between asset prices, risk, and expected returns. All asset pricing models can be traced back to the basic consumption-based pricing equation, where the price equals the expected discounted payoff. A key implication is the beta-pricing representation:

$$R_{i,t} = \lambda_0 + \beta_i(\tilde{f}_t + \lambda_f) + \varepsilon_{i,t}, \quad \forall i, t$$

where $R_{i,t}$ is the excess return of asset i at time t, given by the return minus the risk-free rate; f_t is the excess return of the factor at time t; μ_f is the factor mean; $\tilde{f}_t$ is the demeaned factor at time t, $\tilde{f}_t = f_t - \mu_f$; λ_0 is the model intercept; β_i is the factor exposure of asset i, the sensitivity of its excess return to $\tilde{f}$; λ_f is the factor risk price; and $\varepsilon_{i,t}$ is an error term. In empirical applications, each i typically denotes a test portfolio rather than a single security.

In this model, the expected excess return on an asset is proportional to its covariance with the stochastic discount factor, or equivalently, to its exposure (beta) with respect to systematic factors. This insight underlies CAPM and multifactor models.

Formally, these models involve solving a system of nonlinear equations linking expected returns, portfolio exposures, and factors' risk prices. Estimation can proceed in several ways: the Fama-MacBeth procedure, which uses cross-sectional regressions of returns on betas; SUR-GLS methods, which exploit the system structure by stacking all equations together into a single system and account for error correlation; and the Generalized Method of Moments (GMM), which estimates parameters from moment conditions implied by the pricing equation. Each approach offers a way to test whether linear factor models adequately explain the cross-section of expected returns.

However, the nonlinear nature of the underlying model raises computational and identification challenges, especially when factors are weak or highly collinear. This provides motivation for the introduction of linear approximation methods, which approximate nonlinear product terms (the

risk premium, which in our notation is the product of the factor exposure β_i of portfolio i and the risk price λ_f of the factor f) by tractable linear or piecewise-linear forms. These approximations simplify estimation, improve numerical stability, and allow the use of efficient linear or convex programming techniques, while still capturing the essence of risk-return trade-off in beta-pricing models.

We explore three main methods (Taylor approximation, Convex Approximation and Product factor) with several variants used with linear approximation for product functions, including piecewise linear functions with Linear Programming (LP) and Mixed-Integer Linear Programming (MILP). Our results show that, in the presence of heavy-tailed distributions, the L1-norm methods proposed in this study exhibit lower bias and variance for risk price estimation than traditional L2-norm approaches. The linear approximation methods represent a novel contribution to the literature. Especially, the Taylor Convex and the Product Factor approximation methods have the merit of matching the accuracy of other approaches while offering superior CPU time performance.

The risk premium estimation of a two-factor model, presented in Chapter 2, evolved into an article on global economic integration that we published, Penco and Lucas (2024). We apply classical estimation methods (GMM and SUR) to study the integration of global economies (Europe, North America, Asia, and Japan). Our results show integration between European portfolios and the U.S. stock market, as well as between Asian portfolios and the U.S. stock market, over the full period analysed: twenty years from January 2003 to December 2022. Furthermore, by means of a rolling regression, we track how relationships between variables evolve over time, capturing regime shifts, assessing parameter stability, and detecting transitions between integration and segmentation.

In Chapter 3, we realised that our investigation could have a potentially wider theoretical goal, the idea to better understand the integration and segmentation definitions and help to visualise them, by means of other related concepts as Copula and Stochastic Discount Factor (SDF). We developed the SDF copula analysis, an original method that helps to visualize whether strong dependence arises primarily from exposure or from pricing. We show that the sum of the factor realizations depends only on exposures and explains return variability regardless of whether factors are priced. By contrast, the SDF-return copula correlation depends directly on which factors are priced in the model: setting a factor's risk price λ to zero removes its impact on the SDF while it does still contribute to return variation. This distinction illustrates the essential difference between exposure and pricing in factor models: exposure drives realised return variation through the factor loadings, while only non-zero risk prices affect the expected return and the Stochastic

Discount Factor.

The SDF copula analysis is also flexible, as it can be extended to Student-t copulas or other time-dependent copulas. This is particularly relevant since financial data commonly exhibit extreme co-movements during crises.

Finally, we propose a copula-based maximum likelihood estimator as a novel approach for risk price estimation in the one-factor asset pricing model. While traditional methods such as Fama-MacBeth, GMM, and SUR rely on linear covariance structures, the copula framework allows explicit modelling of nonlinear and tail dependencies in residuals. In particular, the results show that the multivariate copula density likelihood recovers cross-sectional Generalized Least Squares (GLS) estimates, while the univariate copula likelihood recovers the Ordinary Least Squares (OLS) estimates under the assumptions of homoscedasticity and no cross-sectional correlation. The copula formulation thus offers a direct comparison between linear correlation-based approaches and dependence modelling based on transformed ranks. Although computationally more intensive, the copula method can be extended to more flexible specifications such as the Student-t copula, which provides robustness to heavy tails, and to dynamic copulas capturing time-varying dependence.

1.2 Organization of Thesis

After the literature review, in Section 2.2, we introduce the risk price estimation, which is the theoretical foundation of the thesis, and define the beta-pricing model, which we identify as a system of nonlinear equations where the unknowns are the parameters. In Section 2.3 we propose several linear approximations of the model, including an integer programming method, that to our knowledge is for the first time applied to risk price estimation. In Section 2.4, we define a general system of nonlinear equations, which provides the framework for the estimation of the parameters and the covariance matrix according to the main methods used in the literature: Seemingly Unrelated Regression (SUR), Generalized Least Squares (GLS), and Generalized Method of Moments (GMM).

We compare the results between the classic estimation methods and the linear approximation in the one-factor model experiment, Section 2.6.1.3, which is part of an article that has been submitted for publication.

Our results suggest that the L1-norm methods are more appropriate for risk price estimation under heavy-tails, as observed in the case of the EU market factor. In Section 2.6.1.4, we report

the main results from the global economies article, and provide additional comments about the errors-in-variables problem and weak factor issue related to risk price estimation.

In Section 2.6.2, we extend the integration model to commodity markets. To better capture the cross-sectional pricing of commodity risk, we use the factor mimicking approach, which is described in Section 2.6.2.5. As expected, the mimicked factors outperform standard factors in explaining commodity risk.

In Chapter 3, we define the elliptical copula density likelihood, Section 3.2.2, and a stochastic discount factor based copula, Section 3.2.3. Then, we build on this approach, by applying the mimicking factor construction to a broader set of commodity indices for risk price estimation in multifactor integration and segmentation models using the Stochastic Discount Factor and the Copula Density Maximum Likelihood Estimation earlier defined. The results are summarised in Section 3.4.2, and will be part of another article in preparation.

Finally, in the regularisation experiment 3.4.2.4, we test a Penalised p-value Fama-MacBeth (FM) GLS Lasso[1], which applies shrinkage directly to the stochastic discount factor. Insignificant factors are penalised and removed, while priced and relevant sources of risk are retained. However, when factors are weak and betas are highly collinear—as is often the case with commodity indices—the regularisation procedure can become very unstable.

The Appendices contain the derivation of the objective function and the covariance matrix for the Ordinary Least Squares (OLS), Generalized Least Squares (GLS), and the General Method of Moments (GMM) in the context of a system of nonlinear equations. Although not entirely new, this contribution is not marginal, as the topic is complex and rarely presented with rigorous mathematical formalism. In particular, we highlight the differences among the methods and provide evidence on why heteroscedasticity and error autocorrelation cannot be ignored.

1.3 Contribution and Publications

We summarise below the main contributions of this thesis.

We introduce novel convex and linear approximation methods for risk premium estimation, demonstrating that L1-norm methods provide more reliable risk price estimates in the presence of heavy-tailed distributions. We also, as a corollary, establish robust equivalence across classical factor models, showing that centred and raw factor regressions provide consistent risk premium estimates under OLS, SUR, GMM, and Fama-MacBeth GLS.

[1] LASSO: Least Absolute Shrinkage and Selection Operator model

We apply a systematic CAPM integration framework to global equity markets, providing clear evidence of integration and segmentation patterns across Europe, North America, Asia, and Japan. In parallel, we extend the integration analysis to commodity markets, applying factor-mimicking techniques where we identify systematic integration patterns between Oil and Gas.

We introduce an original copula density maximum likelihood approach to risk price estimation, which generalizes classical factor models by capturing nonlinear and tail dependencies and recovers OLS and GLS as special cases. As an additional outcome, we develop an SDF–return copula analysis that separates factor exposure from priced risk, helping to identify spurious factors and clarify the distinction between return variation and expected return compensation.

Finally, we propose an heuristic penalised p-value Fama-MacBeth GLS Lasso model that directly regularises the pricing kernel, retaining only statistically and economically relevant risk factors.

Parts of this dissertation have been, or are being, submitted to peer-reviewed journals:

- Section 2.6.1.4: Financial Integration of the European, North America, Asian and Japanese stock markets from 2003 to present times, Penco and Lucas (2024). Available online: https://www.e-jei.org/journal/view.php?doi=10.11130/jei.2024012. Published June 2024.

- Sections 2.3, 2.6.1.3: Linear Approximation Methods for Beta Pricing, Penco and Lucas (2025). Submitted September 2025 to *Computational Economics*.

- Sections 2.6.2, 3.4.2: Copula Density MLE, Copula SDF and Penalised GLS in Commodity Asset Pricing, Penco, Roman and Lucas (2025). Submitted September 2025 to *Quantitative Finance*.

Chapter 2

Linear Approximation for Risk Premium Estimation

In this chapter, we will look at the use of linear approximation methods for risk premium estimation and at their application for the study of global markets and commodities indices integration.

2.1 Literature Review

This literature review discusses key contributions to beta-pricing estimation and risk prices measurement, focusing on the Fama-MacBeth regression, Cochrane's factor pricing approach, Seemingly Unrelated Regression, Generalized Method of Moments, and recent advancements in machine learning (ML).

Fama and MacBeth (1973) introduced a two-step regression approach to estimate risk prices in cross-sectional asset pricing models. Their method first estimates factor loadings using time-series regressions and then regresses asset returns on these estimated betas to obtain the risk prices. Despite its widespread application, the Fama-MacBeth procedure is prone to the errors-in-variables (EIV) problems, as beta estimates from the first step contain measurement errors that bias the second-step results.

Cochrane (2005) links beta-pricing models to the stochastic discount factor representation, addressing errors-in-variables in beta estimation. He investigated factor mimicking portfolios to mitigate noise in factor loadings, thereby improving the robustness of estimated risk prices. He also shows that when factors are weakly identified, the pricing equation may fail to hold in sample applications, leading to significant discrepancies in risk prices estimation.

The Seemingly Unrelated Regression introduced by Zellner (1962) jointly estimates asset return equations while allowing for correlated error terms. SUR, which accounts for cross-sectional dependencies among assets, is especially useful where asset returns share common factor structures.

Jagannathan and Wang (1996) studied conditional CAPM models, allowing for time-varying betas to improve parameter estimation. They also applied SUR to estimate risk prices in models with

multiple factors, where weak factors can lead to poor identification of risk prices due to small variations in betas across assets.

Hansen (1982) introduced the GMM estimator as a generalization of instrumental variables techniques, requiring moment conditions to be imposed for estimation. Cochrane (2005) applied GMM to estimate factor pricing models, emphasizing its ability to correct for errors-in-variables in beta estimation. However, when factors are weakly correlated with returns, standard GMM approaches can suffer from large finite-sample biases, as highlighted by Stock and Wright (2000).

Sarisoy et al. (2024) found out that weak identification of factors can lead to rejection of valid asset pricing models, proposing bias-corrected GMM and the use of additional instrumental variables to bypass these problems.

In beta-pricing, the EIV problem arises in empirical asset research, where the independent variable beta (exposure) is measured independent of the risk price. This measurement error leads to bias toward zero, also known as downward bias: the standard errors of the estimated risk price increase, making statistical inference less reliable, while misestimated betas may cause non-beta factors (e.g., firm size, book-to-market ratio) to appear significant.

Black et al. (1972) were first in pointing out that beta estimates from historical data contain noise, leading to biased estimates of the risk price. Vasicek (1973) proposed Bayesian shrinkage estimation to reduce beta estimation noise. Litzenberger and Ramaswamy (1979) analysed how errors in beta affect asset pricing tests. Shanken (1992) further explored these limitations, showing that standard errors in the second-step regression are underestimated, leading to potential misinterpretations of risk prices significance.

However the most cited works in the literature are: the Instrumental Variable (IV) approach, Shanken (1992); Portfolio Grouping, where stocks are grouped into portfolios to reduce estimation noise, using a cross-sectional regression which however is prone to bias due to beta estimation errors, Fama and MacBeth (1973); Principal Component Analysis (PCA) that works extracting strong factors from noisy data, Connor and Korajczyk (1988).

Fama and French (1993) found a weak relationship between beta and average returns, partly due to beta mismeasurement caused by weak factors. Lewellen and Nagel (2006) showed that short-term beta estimates exacerbate the EIV problem reducing the explanatory power of traditional risk factors. Gagliardini et al. (2016) used a time varying factor model approach to improve beta estimation and address weak factor issues.

More recent works, such as Anatolyev and Mikusheva (2022), have studied the implications of

weak factors in the Fama-MacBeth setting, demonstrating that low-variance factors can inflate standard errors and distort cross-sectional tests.

The beta-pricing model assumes that asset returns depend on their exposure to systematic risk factors. However, if the measured betas or factors are weak, the risk prices estimation becomes unreliable. Weak betas generally occur when the cross-sectional variation in betas is small or contains substantial noise. Weak factors arise when factors have low explanatory power or weak correlation with asset returns. The consequences are the same identified for the EIV, being weak betas in part due to error measurement.

Machine learning (ML), neural networks and tree-based models, such as random forests and gradient boosting, introduced alternative approaches to beta-pricing estimation by leveraging high-dimensional datasets, improving latent factor selection and enhancing the robustness of risk prices estimation. Gu et al. (2020) demonstrated that supervised learning algorithms outperform traditional econometric methods in predicting expected returns, suggesting that nonlinearities in factor structures may be better captured using ML techniques.

However, while ML methods allow for flexible interactions among factors, the risk prices estimates derived from black-box models lack the economic intuition provided by traditional factor pricing theories.

2.2 Risk Premium Estimation

The Capital Asset Pricing Model (CAPM), Sharpe (1964) and Black et al. (1972), provides a theoretical foundation for asset pricing, linking expected returns to systematic risk.

CAPM is based on the following assumptions:

- Investors are rational and risk-averse, maximizing their expected utility.
- Markets are frictionless, with no transaction costs or taxes.
- All investors have homogeneous expectations about asset returns.
- There is a risk-free asset that investors can borrow and lend at the risk-free rate R_f.
- Investors hold efficient portfolios, meaning they only care about systematic risk.

Consider a standard factor model:

$$R_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.1}$$

where:

- $R_{f,t}$ = scalar random variable representing the risk-free rate at time t, we use the US zero coupon bond rate time series[1].

- $R^*_{i,t}$ = return of the local portfolio i at time t.

- $R_{i,t}$ = excess return of the local portfolio i at time t, i.e. $R_{i,t} = R^*_{i,t} - R_{f,t}$.

- $R^*_{m,t}$ is the market simple return at time t.

- f_t is the CAPM factor at time t, the market excess return, i.e. $f_t \equiv R^*_{m,t} - R_{f,t}$.

- α_i is the alpha of the asset, the excess return that is not explained by systematic factors at a given cross-section of the data. For a portfolio including all market securities, this value tends to converge toward zero.

- β_i is the sensitivity of asset i's excess return to f.

- $\varepsilon_{i,t}$ is an error term with $\mathbb{E}[\varepsilon_i] = 0,\ \forall i$.

The assumptions are: factor stationarity: the factors are assumed to be stationary with the following unconditional moments ($\mathbb{E}[f] = \mu_f, \text{Var}(f) = \text{Var}(\tilde{f}) = \mathbb{E}[(f - \mu_f)^2]$); independence of errors and factors: the asset error terms $(\varepsilon_{i,t})$ are uncorrelated with factors: $\text{Cov}(f, \varepsilon_i) = 0$; uncorrelated error terms: the error terms $(\varepsilon_{i,t})$ are assumed to be serially uncorrelated and do not exhibit cross-sectional correlation between different assets.

Consider an investor who chooses a portfolio containing:

- A risk-free asset with expected return $\mu_{Rf} = \mathbb{E}[R_f]$.

- A market portfolio with expected return $(\mathbb{E}[R^*_m])$.

According to CAPM, the expected return on asset i is given by[2]:

$$\mathbb{E}[R^*_i] \ = \ \mu_{Rf} \ + \ \beta_i\big(\mathbb{E}[R^*_m] - \mu_{Rf}\big), \qquad \mu_{Rf} = \mathbb{E}[R_f], \tag{2.2}$$

where $\mathbb{E}[R^*_m] - \mu_{Rf}$ is the market risk premium; and $\beta_i \equiv \frac{\text{Cov}(R_i, f)}{\text{Var}(f)}$.

Following Cochrane (2005), we use "risk price" for the factor price λ_f and "risk premium" for the product $\beta_i \lambda_f$.

The CAPM equation 2.2 states that an asset's expected return is determined by the expected risk-free rate plus compensation for systematic risk and expresses the asset risk premium as a

[1] Following standard econometric notation, the subscript f denotes "free" (risk-free) and is not an index.
[2] For a full derivation of the CAPM equation, we refer to Fama and MacBeth (1973)

function of beta and the market risk premium (the factor's risk price). Equation 2.2 is a special case of the beta-pricing model (Cochrane, 2005), which in general takes the form:

$$\mathbb{E}[R_i^*] = \mu_{Rf} + \beta_i \lambda_f, \quad \forall i \tag{2.3}$$

with $\beta_i = \frac{\text{Cov}(R_i, f)}{\text{Var}(f)}$ and $\lambda_f = \mathbb{E}[f]$ the risk price of factor f.

In the CAPM framework, the relevant factor is the market excess return

$$f_t = R^*_{m,t} - R_{f,t}.$$

Any time series can be decomposed into its mean and an innovation component, so that

$$f_t = \mu_f + \tilde{f}_t, \quad \forall t \qquad \mathbb{E}[\tilde{f}] = 0, \tag{2.4}$$

where $\mu_f = \mathbb{E}[f]$ represents the expected market risk premium and $\tilde{f}_t$ captures the mean-zero fluctuations around it. Equivalently, we can write

$$\tilde{f}_t = f_t - \mathbb{E}[f], \quad \forall t$$

which emphasizes that $\tilde{f}_t$ is simply the demeaned factor. Thus, the innovation $\tilde{f}_t$ can be viewed either as the shock relative to the long-run mean or as the demeaned factor used in estimation.

Substituting equation 2.4 in equation 2.1, we obtain:

$$R_{i,t} = \alpha_i + \beta_i \mu_f + \beta_i \tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.5}$$

From the time-series factor model, equation 2.1, taking expectations over t gives

$$\mathbb{E}[R_i] = \alpha_i + \beta_i \mu_f, \quad \forall i \tag{2.6}$$

Fama and MacBeth (1973) defined a two-step regression approach to estimate risk prices in cross-sectional asset pricing models. For the first-stage time-series regression they used the centred or demeaned factor model $\tilde{f}_t$:

$$R_{i,t} = \alpha_i + \beta_i \tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.7}$$

The second-stage is defined as a cross-sectional regression:

$$\mathbb{E}[R_i] = \lambda_0 + \lambda_f \beta_i, \quad \text{with } H_0 : \lambda_0 = 0, \quad \forall i \tag{2.8}$$

We use the cross-sectional pricing relation at equilibrium, see equation 2.3, so we drop the error term in equation 2.8. Equating the two expressions for $\mathbb{E}[R_i]$ in equations 2.6 and 2.8, we obtain:

$$\alpha_i = \lambda_0 + \lambda_f \beta_i - \beta_i \mu_f, \quad \forall i \tag{2.9}$$

Instead of using $\mathbb{E}[R_i]$, we use the first-stage alphas on the left-hand side of equation 2.8, then:
- using the demeaned factor yields

$$\text{1st stage } \alpha_i^{\text{demeaned}} = \mathbb{E}[R_i]^{3}, \quad \text{2nd stage } \mathbb{E}[R_i] = \lambda_0 + \lambda_{\tilde{f}} \beta_i.$$

- using the raw factor yields

$$\text{1st stage } \alpha_i^{\text{raw}} = \mathbb{E}[R_i] - \beta_i \mu_f, \quad \text{2nd stage } \mathbb{E}[R_i] = \lambda_0 + \lambda_f \beta_i + \beta_i \mu_f.$$

The following equivalence holds:

$$\lambda_{\tilde{f}} = \lambda_f + \mu_f, \qquad \lambda_{\tilde{f}} = \lambda_f + \bar{f} \text{ (with, } \bar{f} = \tfrac{1}{T} \sum_{t=1}^{T} f_t).$$

i.e., demeaning the factor corrects the intercept α by $\beta_i \, \mu_f$.

Substituting equation 2.9 in equation 2.5, we obtain

$$R_{i,t} = \lambda_0 + \lambda_f \beta_i + \beta_i \tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.10}$$

Note. We will use the notation λ_f in place of $\lambda_{\tilde{f}}$, since our analysis is based on the centred-factor specification. Only in the empirical section we report results under both conventions (with and without factor demeaning) for comparison.

Because the Sharpe (1964) CAPM, written in excess returns, implies a zero intercept, in this thesis, we include an intercept in equation 2.8 and equation 2.10 and test $H_0 : \lambda_0 = 0$.[4] Expression 2.10 represents a version of the beta-pricing model, in which the excess return of asset i is driven by its exposure β_i to a common factor $\tilde{f}$, along with a cross-sectional pricing relation

[3] Taking expectations of equation (2.7) gives $\mathbb{E}[R_i] = \alpha_i + \beta_i \, \mathbb{E}[\tilde{f}] = \alpha_i$, since $\mathbb{E}[\tilde{f}] = 0$.

[4] Running the model with *simple* returns $R^*_{i,t}$ instead of excess returns $R_{i,t}$ yields an equivalent reparametrisation: $\lambda^*_0 = \lambda_0 + \mu_{Rf}$, $\lambda^*_f = \lambda_f$, with $\mu_{Rf} \equiv \mathbb{E}[R_f]$.

governed by the risk prices λ_0 and λ_f. This framework builds on the linear factor pricing models pioneered by Ross (1976) in the Arbitrage Pricing Theory (APT), and later extended in empirical asset pricing studies such as Fama and French (1993) and Cochrane (2005).

2.3 Linear Approximation Methods

Many real-world problems involve product terms of variables, for which linearization simplifies the analysis, enhances computational efficiency, and allows for the application of standard linear techniques.

We propose three main linear approximation methods of the beta-pricing model. The methods will enable solving by Linear Programming (LP), Quadratic Programming (QP), and Mixed Integer Linear Programming (MILP) convex solvers.

We also compare the results with the classic solutions via two steps Fama-MacBeth cross-sectional regression, NSUR GLS-based method, and GMM.

The starting point is equation 2.10, our aim is to linearise the product terms $\lambda_f \beta_i$, namely the risk premia. In the experiments we use portfolio and market return time series where $T = 240$ months, corresponds to the number of time periods, and N=6, is the number of portfolios.

The error estimation is computed in Appendix B.

2.3.1 Taylor Product

The Taylor series expansion provides a local linear approximation of a product at a given point, which is a good approximation of the product function provided that the variables do not deviate from the centre. This type of linear approximation was first applied in beta-pricing by Gibbons (1982) and Stambaugh (1982), which deploy a Taylor expansion centred in the beta and lambda estimates $(\hat{\beta}_i, \hat{\lambda}_f)$:

$$\begin{aligned} \beta_i \lambda_f &\approx \hat{\beta}_i \hat{\lambda}_f + \hat{\beta}_i(\lambda_f - \hat{\lambda}_f) + \hat{\lambda}_f(\beta_i - \hat{\beta}_i) \\ &= \hat{\beta}_i \lambda_f + \hat{\lambda}_f \beta_i - \hat{\lambda}_f \hat{\beta}_i, \quad \forall i \end{aligned} \tag{2.11}$$

substituting equation 2.11 in equation 2.10, we obtain a system of linear equations:

$$R_{i,t} + \hat{\lambda}_f \hat{\beta}_i \approx \lambda_0 + \hat{\beta}_i \lambda_f + \beta_i(\tilde{f}_t + \hat{\lambda}_f) + \varepsilon_{i,t}, \quad \forall i, t \tag{2.12}$$

After applying the Taylor series linear approximation of the product term $\beta_i\lambda_f$ as shown in Equation 2.11, the estimation reduces to solving the following least squares problem:

$$\min_{\lambda_0,\lambda_f,\boldsymbol{\beta}} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t} - \lambda_0 - \left[\hat{\beta}_i\lambda_f + \hat{\lambda}_f\beta_i + \beta_i\tilde{f}_t\right] + \hat{\beta}_i\hat{\lambda}_f\right)^2 \tag{2.13}$$

where $\hat{\beta}_i$ and $\hat{\lambda}_f$ are the linear approximation points from the previous iteration.

This procedure is repeated iteratively, updating $(\hat{\beta}_i, \hat{\lambda}_f)$ until convergence.

The residual inside the square is an affine function of the parameters $(\lambda_0, \lambda_f, \beta_i)$, the sum of a square of an affine function is a convex unconstrained quadratic problem (specifically, a Least Squares problem), that is solved in Python cvxpy using OSQP (Operator Splitting Quadratic Program).

In the results, this method is named Taylor product.

2.3.2 Product Factor

If the following two conditions are met: (i) one of the two parameters, λ_f, is not referenced elsewhere in equation 2.10; and (ii) the product $\beta_i\lambda_f$ is non-negative for all i; we can substitute the product factor $\boldsymbol{\beta}\lambda_f$ by the vector $\boldsymbol{\gamma} = \begin{bmatrix}\gamma_1 & \gamma_2 & \cdots & \gamma_N\end{bmatrix}^\top$, then:

$$R_{i,t} = \lambda_0 + \gamma_i + \beta_i\tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.14}$$

with the extra constraints $l_{\lambda_f}\beta_i \leq \gamma_i \leq u_{\lambda_f}\beta_i$ and $\lambda_f = \frac{\gamma_i}{\beta_i}$. These constraints (we use l for the lower bound and u for the upper bound of the subscript's parameter) guarantee that $l_{\lambda_f} \leq \lambda_f \leq u_{\lambda_f}$ whenever $\beta_i \geq 0$, otherwise the product $\boldsymbol{\beta}\lambda_f$ becomes non-monotonic and non convex over its domain. However, we know that negative betas and negative risk price make economic sense. In order to stay in the frame of the Disciplined Convex Programming (DCP) rules, we remove the constraints $\boldsymbol{\beta}, \lambda_f \geq 0$ and do not enforce $\gamma_i = \beta_i\lambda_f$ directly. Instead, we introduce a convex Taylor penalty around $(\hat{\boldsymbol{\beta}}^{(0)}, \hat{\lambda}_f^{\text{cs}})$ to encourage consistency between $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and λ_f.

In the first step, we solve the unconstrained model, equation 2.14, for the following vector of estimates:

$$\hat{\boldsymbol{\theta}}^{(0)} = \begin{bmatrix}\hat{\lambda}_0^{(0)} & (\hat{\boldsymbol{\beta}}^{(0)})^\top & (\hat{\boldsymbol{\gamma}}^{(0)})^\top\end{bmatrix}^\top \in \mathbb{R}^{2N+1}$$

From this unconstrained model, we define a cross-sectional (CS) risk price signal for each portfolio

as:

$$\hat{\lambda}_{f,i}^{\text{cs}} = \frac{\hat{\gamma}_i^{(0)}}{\hat{\beta}_i^{(0)}}, \quad \forall i \tag{2.15}$$

In practice, the average $\hat{\lambda}_f^{\text{cs}}$ is computed over portfolios with sufficiently large $|\hat{\beta}_i^{(0)}|$ to avoid instability. We then compute its average:

$$\hat{\lambda}_f^{\text{cs}} = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\gamma}_i^{(0)}}{\hat{\beta}_i^{(0)}}. \tag{2.16}$$

The starting values for the second step, $\hat{\beta}_i^{(0)}, \quad \hat{\lambda}_f^{\text{cs}}$, are computed only once from the unconstrained regression, and are then kept fixed in the penalty term.

In the second step, we introduce a regularisation term in the objective function, that penalizes deviations from a common risk price $\hat{\lambda}_f^{\text{cs}}$ among equations:

$$\min_{\lambda_0, \lambda_f, \boldsymbol{\beta}, \boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(R_{i,t} - \lambda_0 - \gamma_i - \beta_i \tilde{f}_t \right)^2 + \rho \sum_{i=1}^{N} \left(\gamma_i - \beta_i \hat{\lambda}_f^{\text{cs}} \right)^2, \tag{2.17}$$

where $\rho > 0$ controls the strength of the regularisation.

This soft constraint enforces:

$$\gamma_i \approx \hat{\lambda}_f^{\text{cs}} \beta_i \quad \Rightarrow \quad R_{i,t} \approx \lambda_0 + \hat{\lambda}_f^{\text{cs}} \beta_i + \beta_i \tilde{f}_t, \quad \forall i, t \tag{2.18}$$

- When $\rho = 0$, the model is entirely unconstrained and γ_i is freely estimated.
- As ρ increases, the model enforces $\gamma_i = \hat{\lambda}_f^{\text{cs}} \, \beta_i + \hat{\beta}_i^{(0)} \, \lambda_f - \hat{\lambda}_f^{\text{cs}} \, \hat{\beta}_i^{(0)}$, recovering a linear factor pricing model.

The product term $\boldsymbol{\beta} \hat{\lambda}_f^{\text{cs}}$ is approximated using a first-order Taylor expansion.

The parameter vector $\hat{\boldsymbol{\theta}} = \left[\hat{\lambda}_0 \quad \hat{\lambda}_f \quad \hat{\boldsymbol{\beta}}^\top \quad \hat{\boldsymbol{\gamma}}^\top \right]^\top \in \mathbb{R}^{2N+2}$ is estimated in Python by minimizing the following objective function, enforcing the consistency of the γ_i elements with its Taylor-linearised approximation:

$$\min_{\lambda_0, \lambda_f, \boldsymbol{\beta}, \boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(R_{i,t} - \lambda_0 - \gamma_i - \beta_i \tilde{f}_t \right)^2 + \rho \sum_{i=1}^{N} \left(\gamma_i - \left[\hat{\lambda}_f^{\text{cs}} \, \beta_i + \hat{\beta}_i^{(0)} \, \lambda_f - \hat{\lambda}_f^{\text{cs}} \, \hat{\beta}_i^{(0)} \right] \right)^2 \tag{2.19}$$

where the second term is the quadratic penalty, and $\hat{\lambda}_f^{\text{cs}}$, $\hat{\boldsymbol{\beta}}^{(0)}$are the fixed input from the first step. This objective is convex and can be solved efficiently using convex optimisation techniques.

In the results, this variant is named Factor.

In the next method, we again use as the starting values $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\lambda}_f^{(0)} = \hat{\lambda}_f^{\text{cs}}$ in the Taylor approximation:

$$\lambda_f \beta_i \approx \hat{\lambda}_f^{(m)} \beta_i + \hat{\beta}_i^{(m)} \lambda_f - \hat{\lambda}_f^{(m)} \hat{\beta}_i^{(m)}, \quad \forall i, m \tag{2.20}$$

However, in this iterative version, the linear approximation centres $(\hat{\lambda}_f^{(m)}, \hat{\beta}_i^{(m)})$ are updated at each step (m) to the current estimates until a convergence level is reached[5]:

$$\hat{\beta}_i^{(m)} \;=\; \beta_i, \quad \forall i \qquad \hat{\lambda}_f^{(m)} \;=\; \lambda_f$$

This approach generalizes both the Factor and the Taylor product models via an iterative estimation strategy.

In the results, this variant is named Factor iter (abbreviation of iterated).

2.3.3 McCormick Envelope

In this linear programming (LP) formulation, built on equation 2.14, we link the product factor γ_i to the common risk price λ_f through the bilinear identity $\gamma_i = \beta_i \lambda_f$, which we replace by its McCormick convex envelope, McCormick (1976), over the boxes $\beta_i \in [l_i, u_i]$ and $\lambda_f \in [l_{\lambda_f}, u_{\lambda_f}]$, for each portfolio i:

$$\begin{aligned}
\gamma_i &\geq l_i\, \lambda_f + \beta_i\, l_{\lambda_f} - l_i l_{\lambda_f} \\
\gamma_i &\geq u_i\, \lambda_f + \beta_i\, u_{\lambda_f} - u_i u_{\lambda_f} \\
\gamma_i &\leq u_i\, \lambda_f + \beta_i\, l_{\lambda_f} - u_i l_{\lambda_f} \\
\gamma_i &\leq l_i\, \lambda_f + \beta_i\, u_{\lambda_f} - l_i u_{\lambda_f}
\end{aligned} \tag{2.21}$$

The returns are fitted in the same form as the product factor specification:

$$\widehat{R}_{i,t} \;=\; \lambda_0 \;+\; \gamma_i \;+\; \beta_i\, \tilde{f}_t, \qquad \forall i, t,$$

We enforce stability and tighten the relaxation by adding the following corner product bounds:

$$\min\{l_i l_{\lambda_f},\, l_i u_{\lambda_f},\, u_i l_{\lambda_f},\, u_i u_{\lambda_f}\} \;\leq\; \gamma_i \;\leq\; \max\{l_i l_{\lambda_f},\, l_i u_{\lambda_f},\, u_i l_{\lambda_f},\, u_i u_{\lambda_f}\}. \tag{2.22}$$

We estimate by least absolute deviations, introduce variables $u_{i,t} \geq 0$ for absolute residuals and

[5] We set a tolerance to exit the iteration when the residual $\leq 10^{-7}$ and alternatively a maximum number of iteration M=10.

solve the LP:

$$\begin{aligned} \min_{\lambda_0,\lambda_f,\{\beta_i,\gamma_i\},\{u_{i,t}\}} \quad & \sum_{i=1}^{N}\sum_{t=1}^{T} u_{i,t} \\ \text{s.t.} \quad & u_{i,t} \geq R_{i,t} - \left(\lambda_0 + \gamma_i + \beta_i \tilde{f}_t\right), \\ & u_{i,t} \geq -R_{i,t} + \left(\lambda_0 + \gamma_i + \beta_i \tilde{f}_t\right), \quad \forall i,t, \\ & \text{McCormick for } \gamma_i = \beta_i \lambda_f \text{ on } [l_i, u_i] \times [l_{\lambda_f}, u_{\lambda_f}] \text{ (Eq. 2.21)}, \\ & l_i \leq \beta_i \leq u_i, \quad l_{\lambda_f} \leq \lambda_f \leq u_{\lambda_f}, \quad u_{i,t} \geq 0. \end{aligned} \tag{2.23}$$

For each portfolio i, the McCormick formulation consists of: the four linear inequalities given by equation 2.21; the two box constraints for each β_i; the two λ_f bounds. Therefore, for six portfolios with a common factor λ_f, the polyhedron in $(\beta, \lambda_f, \gamma)$-space is defined by $4N+2N+2 = 6 \cdot 4 + 6 \cdot 2 + 2 = 38$ facet inequalities. The γ_i corner-product bounds are included for numerical robustness, though they are implied by the McCormick envelope and do not add facets. Tighter bounds $[l_i, u_i]$ and $[l_{\lambda_f}, u_{\lambda_f}]$ shrink the McCormick envelope, which, in the limit, collapses onto the bilinear graph $\gamma_i = \beta_i \lambda_f$.

While the sector constraints in the product factor method 2.14, $l_{\lambda_f}\beta_i \leq \gamma_i \leq u_{\lambda_f}\beta_i$, guarantee $l_{\lambda_f} \leq \gamma_i/\beta_i \leq u_{\lambda_f}$ only when $\beta_i \geq 0$, the McCormick envelope 2.21 does not require any restrictions on β_i, as long as finite bounds $[l_i, u_i]$ and $[l_{\lambda_f}, u_{\lambda_f}]$ are provided.

Problem 2.23 can be solved with a standard Guroby Dual Simplex LP solver. In the results, this variant is named LP L1.

2.3.4 Integer Programming

In this section we use a combination of Taylor approximation alternate with piecewise linear approximation associated with an integer programming method, see Asghari et al. (2022). Although these methods are well known in operational research, to our knowledge they have never been applied to beta-pricing. The method consists of defining the vectors $\mathbf{y}_1 = \begin{bmatrix} y_{1,1} & y_{2,1} & \cdots & y_{N,1} \end{bmatrix}^\top$, $\mathbf{y}_2 = \begin{bmatrix} y_{1,2} & y_{2,2} & \cdots & y_{N,2} \end{bmatrix}^\top$ and the following new variables[6]:

$$y_{i,1} = \frac{1}{2}(\beta_i + \lambda_f) \quad y_{i,2} = \frac{1}{2}(\beta_i - \lambda_f), \quad \forall i \tag{2.24}$$

[6] Under the constraints $l_{\beta_i} \leq \beta_i \leq u_{\beta_i}$, $l_{\lambda_f} \leq \lambda_f \leq u_{\lambda_f}$

Then, we obtain a result that is separable:

$$\beta_i \lambda_f = y_{i,1}^2 - y_{i,2}^2, \quad \forall i \tag{2.25}$$

The model under the new variables[7] is defined as:

$$\begin{aligned} R_{i,t} &= \lambda_0 + y_{i,1}^2 - y_{i,2}^2 + (y_{i,1} + y_{i,2})\tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t \\ y_{i,1} - y_{i,2} &= y_{j,1} - y_{j,2}, \text{ where } i \neq j \text{ with } i, j \in [1, N] \end{aligned} \tag{2.26}$$

Betas and risk price can then be derived as:

$$\beta_i = y_{i,1} + y_{i,2} \quad \lambda_f = y_{i,1} - y_{i,2}, \quad \forall i. \tag{2.27}$$

However, the term $y_{i,1}^2 - y_{i,2}^2$ is non-convex, and thus incompatible with standard convex optimisation frameworks. Then, we apply the following variants.

2.3.4.1 Taylor Convex Approximation

The nonlinear term $y_{i,1}^2 - y_{i,2}^2$ is approximated using a first-order Taylor expansion around the starting centre points a_i and b_i:

$$y_{i,1}^2 - y_{i,2}^2 \approx 2a_i y_{i,1} - 2b_i y_{i,2} - a_i^2 + b_i^2, \quad \forall i \tag{2.28}$$

The parameter vector is defined as $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\lambda}_0 & \hat{\lambda}_f & \hat{\mathbf{y}}_1 & \hat{\mathbf{y}}_2 \end{bmatrix}^\top \in \mathbb{R}^{2N+2}$. The objective function is given by:

$$\begin{aligned} \min_{\lambda_0, \lambda_f, \mathbf{y}_1, \mathbf{y}_2} & \sum_{i=1}^{N} \sum_{t=1}^{T} \left(R_{i,t} - \lambda_0 - (2a_i y_{i,1} - 2b_i y_{i,2} - a_i^2 + b_i^2) - (y_{i,1} + y_{i,2})\tilde{f}_t \right)^2 \\ \text{subject to} \quad & y_{i,1} - y_{i,2} = \lambda_f, \quad \forall i \end{aligned} \tag{2.29}$$

This linear approximation reduces the problem to a Quadratic Program (QP) with linear constraints, which enables the use of iterative convex optimisation techniques, as discussed in Boyd and Vandenberghe (2004).

After solving, we update the Taylor centre points:

$$a_i \leftarrow y_{i,1}, \quad b_i \leftarrow y_{i,2}, \quad \forall i \tag{2.30}$$

[7] The constraints under the new variables are: $\frac{1}{2}(l_{\beta_i} + l_{\lambda_f}) \leq y_{i,1} \leq \frac{1}{2}(u_{\beta_i} + u_{\lambda_f})$, and $\frac{1}{2}(l_{\beta_i} - l_{\lambda_f}) \leq y_{i,2} \leq \frac{1}{2}(u_{\beta_i} - u_{\lambda_f})$

and repeat the process until the convergence tolerance is reached.

In the results, we name this method Taylor convex

2.3.4.2 Piecewise Approximation

We use piecewise linear interpolation via hybrid Mixed-Integer Linear Programming (MILP), Vielma (2015).
First, we define the grids to approximate $y_{i,1}$ and $y_{i,2}$:

$$x_{i,1,k} \in [l_{i,1}, u_{i,1}] \quad x_{i,2,k} \in [l_{i,2}, u_{i,2}], \quad \forall i, k \tag{2.31}$$

We allocate each portfolio to an individual grid of K points each, with the respective lower $(l_{i,1}, l_{i,2})$ and upper $(u_{i,1}, u_{i,2})$ bounds.

Then, the method consists of the following formulations:

Special Ordered Set of Type 1. SOS1 constraints allow only one variable in the set to be non-zero, with the result of selecting a single grid point (piecewise constant approximation). SOS1 are related with discrete weights, and require integer programming. The trade-off is between model complexity (due to integer variables) and the approximation accuracy (dependent on K).

Special Ordered Set of Type 2. SOS2 constraints allow at most two adjacent variables to be non-zero, enabling linear interpolation between two grid points (piecewise linear approximation). SOS2 are associated with continuous Weights (Convex Combination), resulting in smoother approximations and faster solve times. However, it may lead to over-smoothing and reduced statistical significance in estimates.

In our setup we mostly use the SOS2 approach where:
- the convex combination constraints are defined as:

$$\sum_{k=1}^{K} w_{i,1,k} = 1, \quad \sum_{k=1}^{K} w_{i,2,k} = 1; \quad w_{i,1,k} \geq 0, \quad w_{i,2,k} \geq 0, \quad \forall i \tag{2.32}$$

- the variable definitions are:

$$\begin{aligned} y_{i,1} &\approx \sum_{k=1}^{K} x_{i,1,k} w_{i,1,k}, \quad y_{i,2} \approx \sum_{k=1}^{K} x_{i,2,k} w_{i,2,k}, \quad \forall i \\ y_{i,1}^2 &\approx \sum_{k=1}^{K} x_{i,1,k}^2 w_{i,1,k}, \quad y_{i,2}^2 \approx \sum_{k=1}^{K} x_{i,2,k}^2 w_{i,2,k}, \quad \forall i \end{aligned} \tag{2.33}$$

For SOS1, we replace the weights in equations 2.32 and 2.33 with:

$$
\begin{aligned}
y_{i,1} &\approx \sum_{k=1}^{K} x_{i,1,k}\delta_{i,1,k}, \quad y_{i,2} \approx \sum_{k=1}^{K} x_{i,2,k}\delta_{i,2,k} \\
y_{i,1}^2 &\approx \sum_{k=1}^{K} x_{i,1,k}^2\delta_{i,1,k}, \quad y_{i,2}^2 \approx \sum_{k=1}^{K} x_{i,2,k}^2\delta_{i,2,k} \\
\delta_{i,1,k} &\in \{0,1\}, \quad \delta_{i,2,k} \in \{0,1\}, \quad \forall i
\end{aligned}
\tag{2.34}
$$

The mixed-integer linear programming model utilizes both binary and continuous variables to achieve accurate piecewise linear approximations. Most of the solvers use the branch-and-cut (branch-and-bound augmented with cutting planes) algorithm for the solution.

2.3.4.3 MILP L1

The objective function is defined as:

$$
\min_{\lambda_0,\lambda_f,\mathbf{y}_1,\mathbf{y}_2} \sum_{i=1}^{N}\sum_{t=1}^{T} |R_{i,t} - \hat{R}_{i,t}| + \rho \sum_{i=1}^{N} |y_{i,1} - y_{i,2} - \lambda_f| \tag{2.35}
$$

where:

- $R_{i,t}$ is the actual excess return for portfolio i at time t.
- $\hat{R}_{i,t}$ is the model predicted return.

$$
\hat{R}_{i,t} = \lambda_0 + (y_{i,1}^2 - y_{i,2}^2) + (y_{i,1} + y_{i,2})f_t, \quad \forall i,t \tag{2.36}
$$

- ρ is the regularisation parameter that is introduced to penalize deviations between $y_{i,1} - y_{i,2}$ and the parameter λ_f.

Substituting the piecewise approximation of equation 2.33 in equation 2.35 and 2.36:

$$\min_{\substack{\lambda_0,\lambda_f,\\ w_{i,1,k},w_{i,2,k}}} \sum_{i=1}^{N}\sum_{t=1}^{T}\left|R_{i,t}-\lambda_0-\left(\sum_{k=1}^{K}x_{i,1,k}^2 w_{i,1,k}-\sum_{k=1}^{K}x_{i,2,k}^2 w_{i,2,k}\right)\right. \tag{2.37}$$
$$\left.-\left(\sum_{k=1}^{K}x_{i,1,k}w_{i,1,k}+\sum_{k=1}^{K}x_{i,2,k}w_{i,2,k}\right)f_t\right|$$
$$+\rho\sum_{i=1}^{N}\left|\sum_{k=1}^{K}x_{i,1,k}w_{i,1,k}-\sum_{k=1}^{K}x_{i,2,k}w_{i,2,k}-\lambda_f\right|$$

We use a hybrid approximation: SOS1 for betas and SOS2 for the other estimates.

In the results, we name this method MILP L1.

The L1-norm corresponds to the median of the conditional distribution and is more robust to outliers and heavy-tailed noise than the L2-norm.

Although L1 and L2 typically produce different estimates, they coincide under certain restrictive conditions: the factors are orthogonal and properly scaled (at the cost of losing economic interpretability); the error terms are symmetrically distributed with zero mean; there are no outliers, especially in the regressor space; the model is low-dimensional or has symmetric structure; and there is no conditional skew in the dependent variable.

In Tables 2.1 and 2.2, we report the normality tests for the EU Market factor and the residual of our portfolio equation system, equation 2.10.

Statistic	**Value**	**Interpretation**
Mean	0.667	-
Median	0.945	Left-skew
Skewness	-0.493	Moderate left skew
Excess Kurtosis	1.603	Heavy tails (leptokurtic)
Jarque-Bera Test	33.459 ($p < 0.001$)	Reject normality
Shapiro-Wilk Test	0.976 ($p < 0.001$)	Reject normality

Table 2.1: Normality and Tail Tests, EU Market Factor

Because the factors and the residuals are asymmetric and heavy-tailed, our model does not meet the restrictions necessary for the equivalence between L1 and L2 estimates.

We first estimate the parameters λ_0, λ_f, and the vector of portfolio exposures $\boldsymbol{\beta}$ using an L1 non

Statistic	Value	Interpretation
Mean	0.000	–
Median	-0.031	Asymmetry (Mean $\neq$ Median)
Skewness	0.211	Mild positive skew
Excess Kurtosis	2.315	Heavy tails
Jarque-Bera Test	328.811 (p < 0.001)	Reject normality
Shapiro-Wilk Test	0.976 (p < 0.001)	Reject normality

Table 2.2: Normality and Tail Tests of Residuals, Equities

linear estimator (equation 2.44) applied to the sample data.

These estimates serve as the "true" parameter values in a simulation experiment with Laplace-distributed (double exponential) error terms,

$$\varepsilon_{i,t} \sim \text{Laplace}(0, b_i), \quad \forall i, t,$$

where the scale parameter b_i is chosen to match the empirical residual variance,

$$b_i = \frac{\hat{\sigma}_i}{\sqrt{2}},$$

with $\hat{\sigma}_i^2$ denoting the sample variance of $\hat{\varepsilon}_{i,t}$. Thus, the simulated errors have mean zero and the same variance as the empirical residuals.

We then simulate 10,000 datasets, each consisting of $T = 500$ time periods, and re-estimate the model parameters on each simulated dataset using L1.

The same procedure is repeated using an L2 non linear estimator (equation 2.45) from the empirical data as the true parameters, and simulating data with Gaussian error terms. Then, we re-estimate the model parameters on each simulated dataset using L2.

While the true parameters obtained from L1 and L2 estimation differ, we find that in both setups the corresponding estimator (L1 for Laplace, L2 for Gaussian) is consistently recovering the true parameters on average, see Table 2.3. This confirms that both methods are well-calibrated under their respective noise assumptions.

Parameter	True (from data)	Average Estimate (simulations)
	L1 with Laplace noise (10000 simulations)	
λ_0	0.8290	0.8253
λ_f	-0.1284	-0.1245
	L2 with Gaussian noise (10000 simulations)	
λ_0	0.7483	0.7463
λ_f	0.0110	0.0134

Table 2.3: True and Simulated Estimates under the L1- and L2-norms.

The robustness[8] and efficiency[9] of the L1 and L2 estimators are then compared under heavy-tailed errors, via a Monte Carlo simulation experiment where the true parameters are derived from L2 estimation on real data: we estimate λ_0, λ_f, and the vector of portfolio exposures $\boldsymbol{\beta}$ using the Fama–MacBeth two-step procedure.

Given an estimator $\hat{\boldsymbol{\theta}}$ of the true parameter vector $\boldsymbol{\theta}$, we define:

$$\text{Bias}(\hat{\boldsymbol{\theta}}) = \mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} \tag{2.38}$$

$$\text{Var}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}} - \mathbb{E}[\hat{\boldsymbol{\theta}}]\right)\left(\hat{\boldsymbol{\theta}} - \mathbb{E}[\hat{\boldsymbol{\theta}}]\right)^\top\right] \tag{2.39}$$

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right] = \left\|\text{Bias}(\hat{\boldsymbol{\theta}})\right\|^2 + \text{Tr}\left(\text{Var}(\hat{\boldsymbol{\theta}})\right) \tag{2.40}$$

where $\text{Var}(\hat{\boldsymbol{\theta}})$ is the covariance matrix of a vector estimator $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{n+2}$, and its trace, $\text{Tr}(\text{Var}(\hat{\boldsymbol{\theta}}))$ represents the sum of the variances of the individual components of $\hat{\boldsymbol{\theta}}$.

For the test, we then simulate 10,000 datasets, each consisting of $T = 500$ time periods. In each simulation, we generate factor realizations and return data using the pricing equation 2.10, where the error term $\varepsilon_{i,t}$ is drawn independently from a standardised t-distribution with 3 degrees of freedom (i.e., heavy-tailed noise) and the noise scale is calibrated using the empirical residuals from the initial 1st step OLS regression, equation 2.7.

For each simulated dataset, we estimate the model parameters $(\lambda_0, \lambda_f, \boldsymbol{\beta})$ using both L1 and L2 methods. We then compute the empirical bias, variance and mean squared error (MSE) of each estimator with respect to the L1 and L2 derived true parameters, respectively.

[8] In the context of parameter estimation, robustness refers to the stability of the estimator $\hat{\boldsymbol{\theta}}$ in the presence of deviations from classic assumptions such as normality or homoscedasticity. A robust estimator exhibits relatively low variance and bounded bias even under heavy-tailed or heteroskedastic noise.

[9] Efficiency describes the ability of an estimator to achieve minimal variance when the model is correctly specified. In this sense, the classic L2 estimator is efficient under Gaussian noise.

In Figures 2.1 and 2.2, we show the trade-off between bias and variance when estimating parameters under fat-tailed and Gaussian distributions, and the relative performance of robust (L1) versus efficient (L2) estimation techniques. The results show that, for the portfolios[10] beta loadings

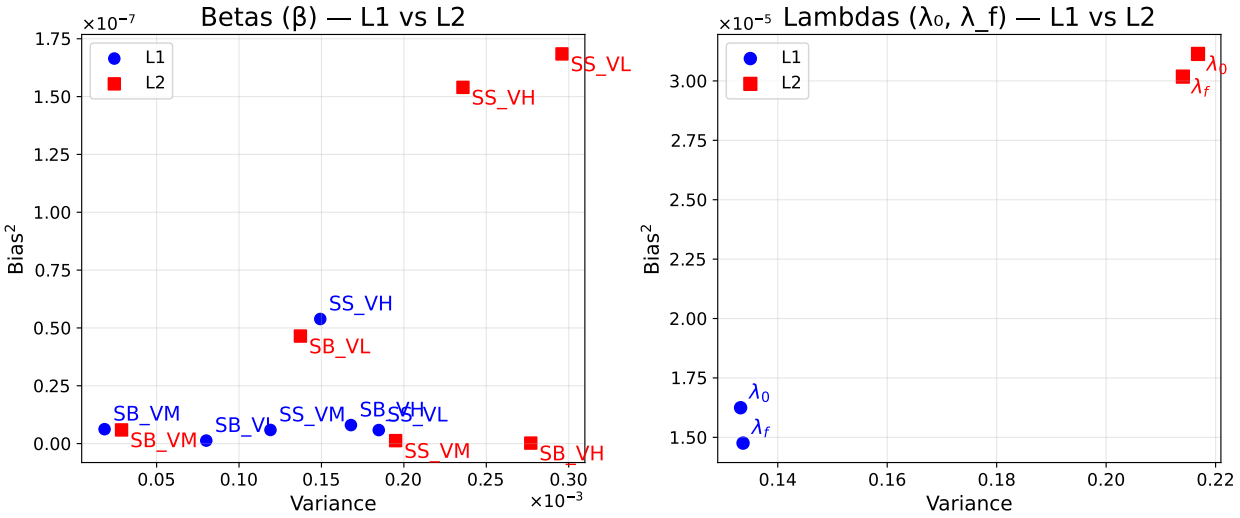

Figure 2.1: Bias–Variance Trade-off for the L1- vs L2-Norms under Heavy Tailed Errors.

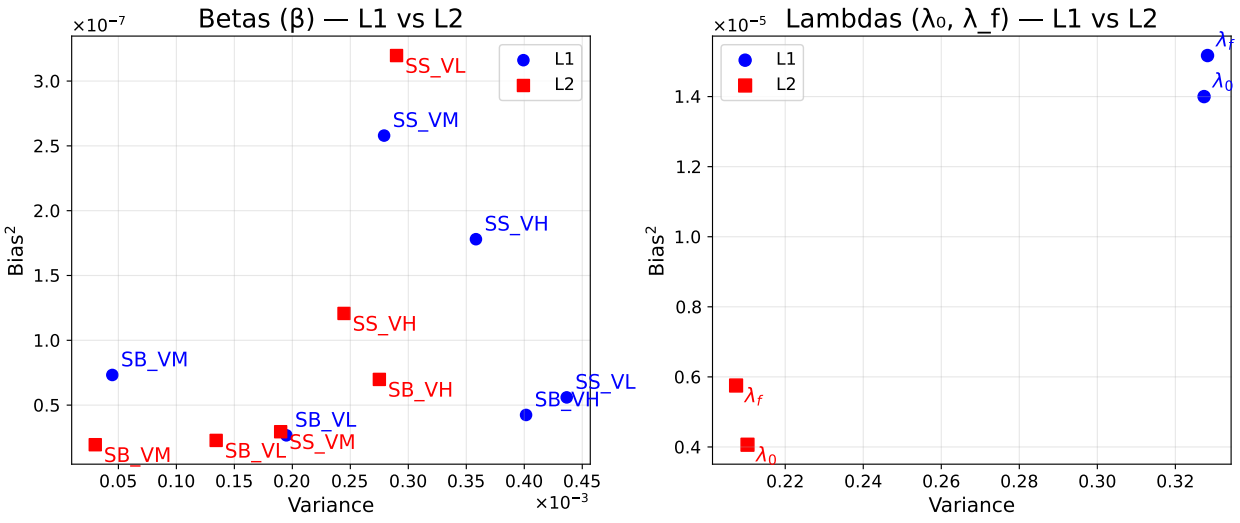

Figure 2.2: Bias–Variance Trade-off for the L1- vs L2-Norms under Normal Errors.

and the risk price parameters, the L1 estimator exhibits lower bias and variance than L2 under heavy tail, while the L2 performs better under Gaussian errors. This highlights L1 as a robust alternative when the classic assumptions of normality and homoscedasticity are violated.

[10] See Section 2.3 for the portfolios definition.

2.3.4.4 MIQP L2

We use an L2-norm, which has the advantage of more stable results, although the problem is now convex quadratic.
The objective function is:

$$\min_{\lambda_0,\lambda_f,\mathbf{y}_1,\mathbf{y}_2} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t}-\hat{R}_{i,t}\right)^2 \tag{2.41}$$

where $\hat{R}_{i,t}$ is already defined in equation 2.36, and instead of the regularisation term we use the following constraint:

$$\lambda_f = y_{i,1} - y_{i,2} \quad \forall i \tag{2.42}$$

Substituting the piecewise approximation of equation 2.33:

$$\begin{aligned}\min_{\substack{\lambda_0,\lambda_f,\\ w_{i,1,k},w_{i,2,k}}} \quad & \sum_{i=1}^{N}\sum_{t=1}^{T}\left[R_{i,t}-\lambda_0-\left(\sum_{k=1}^{K}x_{i,1,k}^2 w_{i,1,k}-\sum_{k=1}^{K}x_{i,2,k}^2 w_{i,2,k}\right)\right.\\ & \left.-\left(\sum_{k=1}^{K}x_{i,1,k}w_{i,1,k}+\sum_{k=1}^{K}x_{i,2,k}w_{i,2,k}\right)f_t\right]^2 \\ & \lambda_f = \sum_{k=1}^{K}x_{i,1,k}w_{i,1,k}-\sum_{k=1}^{K}x_{i,2,k}w_{i,2,k}\end{aligned} \tag{2.43}$$

In addition, the model is solved iteratively using refined grids. The changes in beta values are computed untill they are below a certain tolerance (10^{-7}), while the process exits if the changes started to increase.

After each step, the grid is re-centred around the latest β_i plus a span to ensure full grid coverage.

In the results, we name this method MIQP L2.

The L2-norm corresponds to the mean of the conditional distribution of errors. It is optimal when residuals are normally distributed, but it is highly sensitive to outliers (large deviations).

2.3.5 Nonlinear

Finally, we use a nonlinear solver with the following objective function.

L1-norm:

$$\min_{\lambda_0,\lambda_f,\boldsymbol{\beta}} \sum_{i=1}^{N}\sum_{t=1}^{T}\left|R_{i,t}-\lambda_0-\beta_i\lambda_f-\beta_i\tilde{f}_t\right| \tag{2.44}$$

In the results, this variant is named Nonlinear L1.

L2-norm:

$$\min_{\lambda_0,\lambda_f,\boldsymbol{\beta}} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t}-\lambda_0-\beta_i\lambda_f-\beta_i\tilde{f}_t\right)^2 \tag{2.45}$$

In the results, this variant is named Nonlinear L2.

For the optimisation, we use the quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) and the Sequential Least Squares Programming (SLSQP) to further check accuracy.

2.4 A General Nonlinear Model

The motivation for this section is to provide a general framework for a system of nonlinear equations. We follow the definition of a system of nonlinear equations reported in the SAS Model Procedure guide, page 1488, SAS (2018):

$$\begin{aligned}
&q_1(y_{1,t},y_{2,t},\ldots,y_{N,t},\ x_{1,t},x_{2,t},\ldots,x_{\bar{\ell},t},\ \theta_1,\theta_2,\ldots,\theta_{\bar{k}})=\varepsilon_{1,t}\\
&q_2(y_{1,t},y_{2,t},\ldots,y_{N,t},\ x_{1,t},x_{2,t},\ldots,x_{\bar{\ell},t},\ \theta_1,\theta_2,\ldots,\theta_{\bar{k}})=\varepsilon_{2,t}\\
&\quad\vdots\\
&q_N(y_{1,t},y_{2,t},\ldots,y_{N,t},\ x_{1,t},x_{2,t},\ldots,x_{\bar{\ell},t},\ \theta_1,\theta_2,\ldots,\theta_{\bar{k}})=\varepsilon_{N,t},\quad \forall t
\end{aligned} \tag{2.46}$$

where $\mathbf{q}\in\mathbb{R}^N$ is a real vector valued function of $\mathbf{y}_t\in\mathbb{R}^N, \mathbf{x}_t\in\mathbb{R}^{\bar{\ell}}, \boldsymbol{\theta}\in\mathbb{R}^{\bar{k}},$.

N is the number of the endogenous variables; $\bar{\ell}$ is the number of exogenous components; $\bar{k}$ is the number of parameters; t ranges from 1 to T is the number of non missing observations; $\mathbf{z}_t\in\mathbb{R}^{\bar{p}}$ is a vector of $\bar{p}$ instruments; $\boldsymbol{\varepsilon}_t$ is an unobservable disturbance vector with the following properties: $\mathbb{E}(\boldsymbol{\varepsilon})=0$; $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)=\boldsymbol{\Sigma}$.

The vector of endogenous variables at time t is defined as

$$\mathbf{y}_t=\begin{bmatrix}y_{1,t} & y_{2,t} & \cdots & y_{N,t}\end{bmatrix}^\top,\quad \forall t$$

The vector of observable variables at time t is given by the factor vector $\mathbf{f}_t$[11]:

$$\mathbf{x}_t=\begin{bmatrix}x_{1,t} & x_{2,t} & \cdots & x_{\bar{\ell},t}\end{bmatrix}^\top=\begin{bmatrix}f_{1,t} & f_{2,t} & \cdots & f_{\bar{\ell},t}\end{bmatrix}^\top,\quad \forall t$$

[11] We use bold font to denote vectors, e.g., $\mathbf{f}_t=[f_{1,t},f_{2,t},\ldots,f_{\bar{\ell},t}]^\top$.

The instrument vector $\mathbf{z}_t$ is a function of $\mathbf{x}_t$:

$$\mathbf{z}_t = \mathbf{Z}(\mathbf{x}_t) = \begin{bmatrix} Z_1(x_{1,t}) & Z_2(x_{2,t}) & \cdots & Z_{\bar{\ell}}(x_{\bar{\ell},t}) \end{bmatrix}^\top, \quad \forall t$$

where each $Z_l(x_{\bar{\ell},t})$ represents a transformation of the corresponding component of $\mathbf{x}_t$.
The normalised form of the model can be written:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{f}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t, \quad \forall t \\ \mathbf{z}_t &= \mathbf{Z}(\mathbf{x}_t), \quad \forall t \end{aligned} \tag{2.47}$$

The same nonlinear model in general vector form is shown below:

$$\begin{aligned} \boldsymbol{\varepsilon}_t &= \mathbf{q}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}), \quad \forall t \\ \mathbf{z}_t &= \mathbf{Z}(\mathbf{x}_t), \quad \forall t \end{aligned} \tag{2.48}$$

In the Generalized Method of Moments, we assume that the model implies a set of moment conditions:

$$\mathbb{E}[\mathbf{q}(\mathbf{y}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] = 0. \tag{2.49}$$

These conditions imply that, at the true parameter vector, the expectation of the moment function is zero. The following definitions apply:

$\mathbf{g}_T$ is a vector of moment functions.
$\hat{\mathbf{g}}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{q}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}) \otimes \mathbf{z}_t$ is the sample moment condition
$\mathbf{r}_i = \begin{bmatrix} q_i(\mathbf{y}_1, \mathbf{x}_1, \boldsymbol{\theta}) & q_i(\mathbf{y}_2, \mathbf{x}_2, \boldsymbol{\theta}) & \cdots & q_i(\mathbf{y}_T, \mathbf{x}_T, \boldsymbol{\theta}) \end{bmatrix}^\top \in \mathbb{R}^{T\times 1}$ is the column vector of residuals for the ithequation
$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1^\top & \mathbf{r}_2^\top & \cdots & \mathbf{r}_N^\top \end{bmatrix}^\top \in \mathbb{R}^{TN\times 1}$ is the vector of residuals of the N equations stacked together
$\mathbf{S} \in \mathbb{R}^{N\times N}$ is a matrix that estimates $\boldsymbol{\Sigma}$ the covariances of the errors across equations (referred to as the $\mathbf{S}$ matrix).
$\hat{\mathbf{V}} \in \mathbb{R}^{N\bar{p}\times N\bar{p}}$ is the matrix that represents the variance of the moment functions.[12]
$\mathbf{Z} \in \mathbb{R}^{T\times \bar{p}}$ is the matrix of instruments
$\mathbf{J}$ is the Jacobian, $\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^\top} \in \mathbb{R}^{TN\times \bar{k}}$, the partial derivatives matrix of the residual with respect to the parameters
$\mathbf{I} \in \mathbb{R}^{T\times T}$ is the identity matrix

[12] See equation 4.65 for the definition

$\mathbf{H} \in \mathbb{R}^{N\bar{p}\times TN}$ is a matrix of instruments, $\mathbf{H} = \begin{bmatrix} \mathbf{Z}^\top & 0 & \cdots & 0 \\ 0 & \mathbf{Z}^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}^\top \end{bmatrix} = \mathbf{I_N} \otimes \mathbf{Z}^\top$

$\otimes$ is the notation for a Kronecker product[13]

In the appendices, we derive the objective function and the covariance matrix for the a systems of nonlinear equations, using the following methods: Ordinary Least Squares, Seemingly Unrelated Regression, and Generalized Method of Moments. Here, we present the objective functions and covariance matrices as obtained from those derivations, with reference to the corresponding proofs in the appendices.

OLS objective function $= \mathbf{r}^\top \mathbf{r}/T$ [14]
OLS covariance of $\boldsymbol{\theta} = (\mathbf{J}^\top (diag(\mathbf{S})^{-1} \otimes \mathbf{I})\mathbf{J})^{-1}$ [15]
NSUR objective function $= \mathbf{r}^\top (\mathbf{S}^{-1} \otimes \mathbf{I})\mathbf{r}/T$ [16]
NSUR covariance of $\boldsymbol{\theta} = (\mathbf{J}^\top (\mathbf{S}^{-1} \otimes \mathbf{I})\mathbf{J})^{-1}$ [17]
GMM objective function $= [T\hat{\mathbf{g}}_T(\boldsymbol{\theta}]^\top \hat{\mathbf{V}}^{-1} [T\hat{\mathbf{g}}_T(\boldsymbol{\theta})]/T$ [18]
GMM covariance of $\boldsymbol{\theta} = ((\mathbf{HJ})^\top \hat{\mathbf{V}}^{-1} (\mathbf{HJ}))^{-1}$ [19]

We now focus on a simplified system of nonlinear equations derived from equation 2.47, the SAS-style general model notation. While equation 2.47 defines a fully general simultaneous nonlinear system of equations in terms of the parameters, a reduced specification is sufficient for the purposes of beta-pricing and risk premium estimation.

Specifically, we define:

- $\mathbf{y}_t \equiv \mathbf{R}_t \in \mathbb{R}^N$: is a column vector of excess returns across N portfolios,
- $\mathbf{x}_t = [\tilde{\mathbf{f}}_t] \in \mathbb{R}^{\bar{\ell}}$, where $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \mathbb{E}[\mathbf{f}_t]$ is the vector of demeaned factors,
- $\boldsymbol{\theta} = \begin{bmatrix} \lambda_0 & \text{vec}(\boldsymbol{B})^\top & \boldsymbol{\lambda}_f^\top \end{bmatrix}^\top \in \mathbb{R}^{1+N\bar{\ell}+\bar{\ell}}$ is the parameter vector:
  - $\lambda_0 \cdot \mathbf{1}_N \in \mathbb{R}^N$: common intercept across all assets,

[13] See definition of the matrix direct product 4.15 in the Appendices
[14] See equation 4.26 for reference. In literature, this is the nonlinear least squares (NLS) method unweighted.
[15] See equation 4.54 for reference. SAS reports this as OLS, but it corresponds to the diagonal WLS/GLS case with no cross equation off–diagonal terms.
[16] See equation 4.34 for reference
[17] See equation 4.58 for reference
[18] See equation 4.68. The objective function is also written: $[\hat{\mathbf{g}}_T(\boldsymbol{\theta}]^\top \hat{\boldsymbol{\Lambda}}^{-1} [\hat{\mathbf{g}}_T(\boldsymbol{\theta})]$ with $\hat{\mathbf{V}} = T\hat{\boldsymbol{\Lambda}}$
[19] See equation 4.92 and Wooldridge (2010) pag.193

- $\boldsymbol{B} \in \mathbb{R}^{N \times \bar{\ell}}$: matrix of factor loadings (row i is $\boldsymbol{\beta}_i^\top$), see equation 4.16,
- $\text{vec}(\boldsymbol{B}) \in \mathbb{R}^{N\bar{\ell}}$: vectorization of $\boldsymbol{B}$ by stacking its columns, see equation 4.17,
- $\boldsymbol{\lambda}_f \in \mathbb{R}^{\bar{\ell}}$: vector of factor risk prices.

Then the model becomes:

$$\boldsymbol{f}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}) = \lambda_0 \cdot \mathbf{1}_N + \left(\mathbf{I}_N \otimes (\mathbf{x}_t + \boldsymbol{\lambda}_f)^\top\right) \cdot \text{vec}(\boldsymbol{B}) \quad \forall t \tag{2.50}$$

Alternatively:

$$\mathbf{y}_t = \boldsymbol{f}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t = \lambda_0 \cdot \mathbf{1}_N + \boldsymbol{B}\,(\mathbf{x}_t + \boldsymbol{\lambda}_f) + \boldsymbol{\varepsilon}_t, \quad \forall t$$

which can be written as:

$$\mathbf{R}_t = \lambda_0 \cdot \mathbf{1}_N + \boldsymbol{B}(\tilde{\mathbf{f}}_t + \boldsymbol{\lambda}_f) + \boldsymbol{\varepsilon}_t, \quad \forall t$$

and is equivalent to the multifactor beta-pricing model:

$$R_{i,t} = \lambda_0 + \sum_{l=1}^{\bar{\ell}} \beta_{i,l}(\tilde{f}_{l,t} + \lambda_l) + \varepsilon_{i,t}, \quad \forall i, t \tag{2.51}$$

From 2.51, we recover the one-factor model ($\bar{\ell} = 1$) defined in Equation (2.10):

$$R_{i,t} = \lambda_0 + \beta_i(\tilde{f}_t + \lambda_f) + \varepsilon_{i,t}, \quad \forall i, t$$

In the next section, we analyse a model with: $\bar{\ell} = 2$ factors[20]; $N = 6$ portfolios. $\bar{k} = 1 + N\bar{\ell} + \bar{\ell} = 15$ is the number of parameters; t ranges from 1 to T, the number of data points; $\mathbf{z}_t = [1,\, x_{1,t},\, x_{2,t}]^\top \in \mathbb{R}^{\bar{p}}$ is a vector of instruments, where $\bar{p} = \bar{\ell} + 1 = 3$.

For the Generalized Method of Moments , the following orthogonality condition is desired[21]:

$$\mathbb{E}[\varepsilon_i(\boldsymbol{\theta}) \cdot z_p] = 0 \quad \forall i, p \tag{2.52}$$

that corresponds to the Kronecker product:

$$\mathbb{E}[\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \otimes \mathbf{z}] = \mathbf{0} \tag{2.53}$$

[20] The factors are the exogenous variables, $\text{Cov}(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathbf{0}$.
[21] In econometrics terms the instruments are exogenous, see Wooldridge (2010), pag. 187

where $\varepsilon(\boldsymbol{\theta}) \otimes \mathbf{z} \in \mathbb{R}^{N\bar{p}}$. The vector of sample moments $\hat{\mathbf{g}}_T$ is the empirical average of the Kronecker product between the asset-level residual vector and the instrument vector:

$$\hat{\mathbf{g}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) \otimes \mathbf{z}_t \tag{2.54}$$

It makes sense to denote the sample moment vector as $\hat{\mathbf{g}}_T(\boldsymbol{\theta})$ to make the dependence on the sample size explicit. In large-sample theory, this convergence is expressed as:

$$\hat{\mathbf{g}}_T(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{g}(\boldsymbol{\theta}) \quad \text{as } T \to \infty,$$

where $\xrightarrow{p}$ denotes convergence in probability.

We have $\bar{g}$ moment conditions, where $\bar{g} = N\bar{p}$, and $\bar{k} = 1 + N\bar{\ell} + \bar{\ell}$ parameters to estimate. The system is overidentified if $\bar{g} > \bar{k}$. For example, with two factors (i.e., $\bar{\ell} = 2$), we would need at least $\bar{p} = 3$ instruments to ensure overidentification.

2.5 Multifactor Beta-Pricing Model

We consider a multifactor beta-pricing model as specified in equation 2.51:

$$R_{i,t} = \lambda_0 + \sum_{l=1}^{\bar{\ell}} \beta_{i,l}(\tilde{f}_{l,t} + \lambda_l) + \varepsilon_{i,t} \quad \forall i, t$$

where $R_{i,t}$ is the excess return of portfolio i at time t, computed as the return minus the risk-free rate; $f_{l,t}$ is the excess return of factor l at time t; $\mu_{f,l}$ is the mean of factor l, i.e. the l-th element of $\boldsymbol{\mu}_f$; $\tilde{f}_{l,t} = f_{l,t} - \mu_{f,l}$ is the demeaned value of factor l at time t; $\lambda_0 \in \mathbb{R}$ is the model intercept; $\beta_{i,l}$ is the loading or exposure of portfolio i on factor l; $\boldsymbol{\beta}_i = \begin{bmatrix}\beta_{i,1} & \beta_{i,2} & \cdots & \beta_{i,\bar{\ell}}\end{bmatrix}^\top \in \mathbb{R}^{\bar{\ell}}$ is the vector of factor loadings for portfolio i; $\lambda_l \in \mathbb{R}$ is the risk price of factor l and $\boldsymbol{\lambda}_f = \begin{bmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_{\bar{\ell}}\end{bmatrix}^\top \in \mathbb{R}^{\bar{\ell}}$ the risk price vector; $\varepsilon_{i,t}$ is the error term for asset i at time t; N is the number of assets; T is the number of time periods; and $\bar{\ell}$ is the number of factors.

This formulation extends the linear factor pricing framework introduced by Fama and MacBeth (1973) and builds on the asset pricing literature as summarized in Cochrane (2005).

The model under study is a special case of the general nonlinear framework described in the previous section. In this setting, it reduces to a system of nonlinear equations with unknown parameters, which in econometrics is usually solved using:

- Fama-MacBeth two-step regression, FM, Fama and MacBeth (1973).

- Nonlinear Seemingly Unrelated Regression, NSUR[22], Zellner (1962).

- Generalized Method of Moments, GMM, Hansen (1982).

The one-step GMM approach jointly estimates all parameters by enforcing the moment conditions orthogonality across time and portfolios (see Appendix K for more details):

$$\mathbb{E}\left[R_i - \lambda_0 - \boldsymbol{\beta}_i^\top(\mathbf{f} + \boldsymbol{\lambda}_f - \boldsymbol{\mu}_f)\right] = 0 \qquad \text{(pricing error)} \tag{2.55a}$$

$$\mathbb{E}\left[\left(R_i - \lambda_0 - \boldsymbol{\beta}_i^\top(\mathbf{f} + \boldsymbol{\lambda}_f - \boldsymbol{\mu}_f)\right)\mathbf{f}^\top\right] = 0 \qquad \text{(error } \times \text{ instrument)} \tag{2.55b}$$

$$\mathbb{E}\left[\mathbf{f} - \boldsymbol{\mu}_f\right] = 0 \qquad \text{(factor mean)} \tag{2.55c}$$

Here, R_i denotes a scalar random variable representing the return of portfolio i at a generic time point. The observed time series, $(R_{i,1}, \ldots, R_{i,T})^\top$, consists of realizations of this random variable over T periods. The population expectation $\mathbb{E}[\cdot]$ is estimated by the sample average over these T realizations.

For GMM estimation, the conditions are stacked across all N portfolios to form the moment vector $\mathbf{g}(\boldsymbol{\theta}) \in \mathbb{R}^{N\bar{p}}$, where $\bar{p}$ is the number of the instruments (instrumental variables), and the GMM estimator is obtained by solving:

$$\hat{\boldsymbol{\theta}}_{\text{GMM}} = \arg\min_{\boldsymbol{\theta}} \ \left(\bar{\mathbf{g}}_T(\boldsymbol{\theta})^\top \mathbf{W}_T \bar{\mathbf{g}}_T(\boldsymbol{\theta})\right), \tag{2.56}$$

where

$$\bar{\mathbf{g}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_t(\boldsymbol{\theta}), \tag{2.57}$$

and $\mathbf{W}_T$ is a positive-definite weighting matrix.

In contrast, the Fama–MacBeth regression proceeds in two steps. First, betas are estimated via time-series regressions for each portfolio:

$$R_{i,t} = \alpha_i + \boldsymbol{\beta}_i^\top \tilde{\mathbf{f}}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{2.58}$$

where $\tilde{\mathbf{f}}_t \in \mathbb{R}^{\bar{\ell}}$ is the vector of $\bar{\ell}$ demeaned factor values at time t, and $\boldsymbol{\beta}_i \in \mathbb{R}^{\bar{\ell}}$ is the vector of factor loadings for portfolio i.

Then, the vector of factor risk prices is estimated in the second step by regressing the average

[22] In the rest of the document, we use NSUR and SUR interchangeably, as we deal with a nonlinear model and employ the SAS implementation of nonlinear SUR for estimation.

returns on the estimated betas:

$$\bar{R}_i = \lambda_0 + \boldsymbol{\lambda}_f^\top \boldsymbol{\beta}_i + \eta_i, \quad \forall i \tag{2.59}$$

where $\bar{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t}$, and $\boldsymbol{\lambda}_f \in \mathbb{R}^{\bar{\ell}}$ is the vector of factor risk prices.

If the factors are not demeaned, this is equivalent to regressing the adjusted returns on the estimated betas:

$$\bar{R}_i - \bar{\mathbf{f}}^\top \boldsymbol{\beta}_i = \lambda_0 + \boldsymbol{\lambda}_f^\top \boldsymbol{\beta}_i + \eta_i, \quad \forall i \tag{2.60}$$

where $\bar{\mathbf{f}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}_t \in \mathbb{R}^{\bar{\ell}}$ is the time-series mean of the factors.

While the second step in FM typically uses Ordinary Least Squares, it can also be performed using Generalized Least Squares to account for cross-sectional heteroscedasticity or correlation, making it closer to the GMM and NSUR frameworks.

For the cross-sectional regression step, let:

- $\bar{\mathbf{R}} \in \mathbb{R}^N$: vector of average excess returns across N assets.
- $\mathbf{B}^* \in \mathbb{R}^{N \times \bar{\ell}}$: matrix of estimated betas $\boldsymbol{\beta}_i$.
- $\mathbf{B} = \left[\mathbf{1},\ \mathbf{B}^*\right] \in \mathbb{R}^{N \times (\bar{\ell}+1)}$: augmented design matrix
- $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$: covariance matrix of residuals from the time-series regressions.

The GLS estimator of the risk prices is:

$$\hat{\boldsymbol{\theta}}_{\text{GLS}} = (\mathbf{B}^\top \boldsymbol{\Sigma}^{-1} \mathbf{B})^{-1} \mathbf{B}^\top \boldsymbol{\Sigma}^{-1} \bar{\mathbf{R}}, \tag{2.61}$$

where $\hat{\boldsymbol{\theta}}_{\text{GLS}} = [\hat{\lambda}_0\ \hat{\boldsymbol{\lambda}}_f^\top]^\top \in \mathbb{R}^{\bar{\ell}+1}$ is the estimated vector of the intercept and factor risk prices.

2.5.1 Integration and Segmentation Models

In this section we report the methods and modelling described extensively in the global economies integration article, Penco and Lucas (2024), which is is built upon the one-factor CAPM integration and segmentation model of Jorion and Schwartz (1986) and the three factors model of Brooks and Iorio (2009). The model is an application of the multifactor beta-pricing model described in the previous section.

For the definition of the integration model, we use the beta representation of the asset price, Cochrane (2005), and we assume that integration cannot be tested by directly running a univariate

regression on market beta, Stehle (1977), being the returns of the Local and Global market positively correlated, Bruner et al. (2008). As noted by Brooks and Iorio (2009), a world market portfolio that is mean variance efficient in a global integrated market should show that assets of different geographic areas with the same sensitivities to the world market portfolio will be traded at similar prices, irrespective of their physical location (Solnik (1974); Stultz (1984); Jorion and Schwartz (1986)).

We refer to the CAPM integration model between a local market and a global market, the extension to the three factor model and to the other markets is straightforward and not shown here. In this work, the geographic market types are: Europe (EU), the local market; North America (US), Asia (AS) and Japan (JP), the global markets; and the Commodities market types are: Oil, the local market; Gas, Aluminium and Soybean, the global markets.

For this example, we use US as the Global market and EU as the Local market and six EU portfolios built by Fama and French using the size (Market Capitalization) and the value (Book-to-Market) as group criteria.

Notation:
$R_{f,t}$ = risk-free rate at time t; we use the US zero coupon bond rate time series
$R^*_{i,t}$ = random return of the local portfolio i at time t
$R_{i,t}$ = excess random return of the local portfolio i at time t, i.e. $R_{i,t} = R^*_{i,t} - R_{f,t}$
$\mathbb{E}(R_{i,t})$ = expected excess return of the of the local portfolio i at time t
$R^*_{US,t}$ = US Global market return at time t
$R_{US,t}$ = US excess Global market return at time t, i.e. $R_{US,t} = R^*_{US,t} - R_{f,t}$
$R_{US\perp EU,t}$ = orthogonalised US return, obtained by projecting $R_{US,t}$ onto the space orthogonal to $R_{EU,t}$; equivalently, it is the residual from regressing $R_{US,t}$ on $R_{EU,t}$
$R^*_{EU,t}$ = EU Local market return at time t
$R_{EU,t}$ = EU excess Local market return at time t, i.e. $R_{EU,t} = R^*_{EU,t} - R_{f,t}$
$R_{EU\perp US,t}$ = orthogonalised EU return, obtained by projecting $R_{EU,t}$ onto the space orthogonal to $R_{US,t}$; equivalently, it is the residual from regressing $R_{EU,t}$ on $R_{US,t}$
β_i^{US} = the integration exposure related to the US Global Market returns for portfolio i
$\beta_i^{EU\perp US}$ = the integration exposure related to the EU orthogonal Local vector of Market returns for portfolio i
λ_{US} = US Global Market risk price in the integration model
λ_{EU} = EU Local Market risk price in the integration model
$\eta_{i,t}$ = the integration model error for the excess returns of the local portfolio i time t
ζ_i^{EU}= the segmentation exposure related to the EU Local Market returns for portfolio i

$\zeta_i^{US\perp EU}$ = the segmentation exposure related to the US orthogonal Global vector of Market returns for portfolio i
δ_{EU} = EU Local Market risk price in the segmentation model
δ_{US} = US Global Market risk price in the segmentation model
$\nu_{i,t}$ = the segmentation model error for the excess returns of the local portfolio i time t

Brooks and Iorio (2009) showed that if the European and US markets are integrated, the only priced factor for an EU stock is the US market return. Hence, the returns on an EU stock-based portfolio i are determined by the empirical CAPM equation below:

$$R_{i,t} = \mathbb{E}(R_{i,t}) + \beta_i^{US} R_{US,t} + \eta_{i,t}, \quad \forall i, t \tag{2.62}$$

Assuming no arbitrage opportunities and some additional conditions, Connor (1984), the expected return on portfolio *i* can be written as:

$$\mathbb{E}(R_i) = \lambda_0 + \lambda_{US}\beta_i^{US}, \quad \forall i \tag{2.63}$$

A non-zero λ_0 implies that the expected return on the zero-beta portfolio is the riskless rate plus a constant. We have already seen that equation 2.63 is the cross-sectional regression of asset returns on beta, refer to the risk price estimation in Section 2.2, which is also called the beta representation of the asset price, Cochrane (2005). In this equation, the local systematic risk, β_i^{EU} relative to the European portfolio, R_i, does not contribute to the pricing of assets. On the other hand, Stehle (1977) exposed how integration cannot be tested by directly running a univariate regression on β_i^{EU}, being the returns on the European and Global market positively correlated, Bruner et al. (2008). For testing the integration of two competing models we build the enhanced model using the local and the global factors.
The collinearity issue between the EU and US market makes a multiple regression on the two factors inadequate as well. Instead, we build the orthogonal projections of the local and global market returns using the Graham – Schmidt process, Apostol (1969).
We define $R_{EU\perp US,t}$ as the orthogonal local vector, the fitted values obtained from the projections of $R_{EU,t}$ into the line crossed by the vector $R_{US,t}$, and we use it as a measure of the local factors, in the enhanced integration model:

$$\mathbb{E}(R_i) = \lambda_0 + \lambda_{US}\beta_i^{US} + \lambda_{EU}\beta_i^{EU\perp US}, \quad \forall i \tag{2.64}$$

where λ_{US} is the risk price related to the global US market return , λ_{EU} is the risk price related to the local EU market return and λ_0 is the intercept. Now, we can write the empirical CAPM

equation for the integrated model:

$$R_{i,t} = \mathbb{E}(R_i) + \beta_i^{US} R_{US,t} + \beta_i^{EU \perp US} R_{EU \perp US,t} + \eta_{i,t}, \quad \forall i, t \tag{2.65}$$

Substituting equation 2.64 in equation 2.65 we obtain the integrated version of the CAPM model:

$$R_{i,t} = \lambda_0 + \beta_i^{US}(R_{US,t} + \lambda_{US}) + \beta_i^{EU \perp US}(R_{EU \perp US,t} + \lambda_{EU}) + \eta_{i,t}, \quad \forall i, t \tag{2.66}$$

In order to prove the complete integration hypothesis, the domestic market risk prices λ_{EU} should be equal to zero, while the global factor λ_{US} should be different from zero for integration.
The segmented model is built in a similar way and we get the following equation:

$$R_{i,t} = \delta_0 + \zeta_i^{EU}(R_{EU,t} + \delta_{EU}) + \zeta_i^{US \perp EU}(R_{US \perp EU,t} + \delta_{US}) + \nu_{i,t}, \quad \forall i, t \tag{2.67}$$

In order to prove the complete segmentation hypothesis, the global market risk prices δ_{US} should be equal to zero, while the local factor δ_{EU} should be different from zero for segmentation.
In the approach proposed by Jorion and Schwartz (1986), we can write equation 2.62 as:

$$R_{i,t} = \mathbb{E}(R_i) + \beta_i^{EU}(R_{US,t} - \mathbb{E}(R_{US})) + \eta_{i,t}, \quad \forall i, t \tag{2.68}$$

Equation 2.65 can be rewritten as:

$$R_{i,t} = \mathbb{E}(R_i) + \beta_i^{US}(R_{US,t} - \mathbb{E}(R_{US})) + \beta_i^{EU \perp US} R_{EU \perp US,t} + \eta_{i,t}, \quad \forall i, t \tag{2.69}$$

Substituting equation 2.64 in equation 2.69 we obtain the integrated version of the CAPM model:

$$R_{i,t} = \lambda_0 + \beta_i^{US}(R_{US,t} - \mathbb{E}(R_{US}) + \lambda_{US}) + \beta_i^{EU \perp US}(R_{EU \perp US,t} + \lambda_{EU}) + \eta_{i,t}, \quad \forall i, t \tag{2.70}$$

We note that the CAPM implies the restriction, $\lambda_{US} = \mathbb{E}(R_{US}) - \lambda_0$ and write:

$$R_{i,t} = \lambda_0(1 - \beta_i^{US}) + \beta_i^{US} R_{US,t} + \beta_i^{EU \perp US}(R_{EU \perp US,t} + \lambda_{EU}) + \eta_{i,t}, \quad \forall i, t \tag{2.71}$$

The segmented model is built in a similar way and we get the following equation:

$$R_{i,t} = \delta_0(1 - \alpha_i^{EU}) + \zeta_i^{EU} R_{EU,t} + \zeta_i^{US \perp EU}(R_{US \perp EU,t} + \delta_{US}) + \nu_{i,t}, \quad \forall i, t \tag{2.72}$$

As described in the results sections, we used several methods to solve the system of nonlinear

equations: Fama-MacBeth (FB) or Cross sectional regression, Section 2.2; NSUR with FGLS, Appendix I; GMM, Appendix K; and the linear approximation techniques, Section 2.3.

Below, we present the details of the Generalized Method of Moments (GMM) applied to the integration system of nonlinear equations. An analogous representation can be derived for the Nonlinear Seemingly Unrelated Regressions (NSUR) framework.

The GMM method is in general based on assigning different weights to the residuals as it is shown for the Weight Least Squares estimator (Appendix G) which is the simplest case. Our article on global economies integration use the Martingale Difference sequence for weighting the residuals and improve the weighting matrix calculation.
Starting from equation 2.66, we define $M_{i,t}$ as the measured error terms ($M_{i,t} \neq \eta_{i,t}$ since the disturbances $\eta_{i,t}$ in equation 2.66 are unobserved):

$$M_{i,t}(\theta) = R_{i,t} - (\lambda_0 + \beta_i^{US}(R_{US,t} + \lambda_{US}) + \beta_i^{EU \perp US}(R_{EU \perp US,t} + \lambda_{EU})), \quad \forall i, t \tag{2.73}$$

Comparing equation 2.50 with our integration model, for each time period t, we have: the equation vector $\mathbf{y}_t = \mathbf{R}_t \in \mathbb{R}^6$, the independent variable vector $\mathbf{x}_t^\top = (R_{US,t}, R_{EU \perp US,t})^\top \in \mathbb{R}^2$, the sample error vector $\mathbf{M}_t(\boldsymbol{\theta}) = [M_{1,t}, \ldots, M_{N,t}]^\top \in \mathbb{R}^6$, the parameter vector is $\boldsymbol{\theta} = \begin{bmatrix} \lambda_0 & \text{vec}(\boldsymbol{B})^\top & \boldsymbol{\lambda}_f^\top \end{bmatrix} \in \mathbb{R}^{15}$ with $\boldsymbol{\lambda}_f = (\lambda_{US}, \lambda_{EU}) \in \mathbb{R}^2$ and $\text{vec}(\boldsymbol{\beta}) = \begin{bmatrix} \boldsymbol{\beta}^{US^\top} & \boldsymbol{\beta}^{EU \perp US^\top} \end{bmatrix}^\top \in \mathbb{R}^{12}$, the instruments' vector $\mathbf{z}_t = [1, R_{US,t}, R_{EU \perp US,t}]^\top \in \mathbb{R}^3$. We have fifteen parameters, six equations, three instruments. Therefore the total number of moments $\bar{g}$ (vectors of nonlinear functions) is then composed of $\mathbf{g} \in \mathbb{R}^{18}$ given by $\bar{g} = N\bar{p} = 6 \times 3 = 18$. The system is over identified, $\bar{g} > \bar{k}$, as per definition of the GMM estimator.

The vector of sample moments $\hat{\mathbf{g}}_T$ is the empirical mean (first moment condition) of the Kronecker product between the asset-level residual vector and the instrument vector, as defined in equation 2.54 which is rewritten below:

$$\hat{\mathbf{g}}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{M}_t(\boldsymbol{\theta}) \otimes \mathbf{z}_t \tag{2.74}$$

Since $\bar{g} > \bar{k}$, we have more equations than parameters, and there is no solution for $\hat{\mathbf{g}}_T(\boldsymbol{\theta}) = 0$. Therefore we use a quadratic form, defined as:

$$Q_T(\boldsymbol{\theta}) = \|\hat{\mathbf{g}}_T(\boldsymbol{\theta})\|_{\mathbf{W}}^2 = \hat{\mathbf{g}}_T^\top(\boldsymbol{\theta}) \mathbf{W} \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \tag{2.75}$$

where $\mathbf{W} \in \mathbb{R}^{\bar{g} \times \bar{g}}$ is a positive definite weighting matrix. The GMM estimator of the objective

function is:

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}} \hat{\mathbf{g}}_T^{\top}(\boldsymbol{\theta}) \mathbf{W} \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \tag{2.76}$$

In the first step (1st) we use the identity matrix $\mathbf{I} \in \mathbb{R}^{18\times 18}$ as the initial weighting matrix: all the moments will have the same weight.

In the second step we estimated the weighting matrix $\mathbf{W}$ using the Outer Product of Moments estimator (OPG[23]), Hamilton (1994), to improve our weighting matrix calculation under the assumption that the moment sequence $(\boldsymbol{M}_t(\boldsymbol{\theta}))$ is uncorrelated over time (e.g., satisfies a Martingale Difference Sequence property). The weight matrix $\mathbf{W}$ is now computed using the OPG variance: the matrix product of the eighteen moments calculated using the parameters of the first step, $\mathbf{g}_T(\hat{\boldsymbol{\theta}}^{1st})$.

$$\mathbf{W} = \text{Var}_{MDS}^{-1} = (\frac{1}{T}\mathbf{g}_T^{\top}(\hat{\boldsymbol{\theta}}^{1st})\mathbf{g}_T(\hat{\boldsymbol{\theta}}^{1st}))^{-1} \tag{2.77}$$

The second step parameters estimator $\hat{\boldsymbol{\theta}}^{2nd}$ of our models minimises this nonlinear optimisation problem. The estimation can be iterated till the chosen GMM step tolerance (i.e. 10^{-7}) is reached.

2.6 Empirical Applications

This section presents the results of the risk premium estimation for the one-factor model and the two-factor models (integration and segmentation), applied to global equity markets and commodity indices.

2.6.1 Global Economies

2.6.1.1 Data Processing

For the test of market integration/segmentation, we used a set of six EU portfolios as dependent variables ($R_{i,t}$). In most of the experiments, the EU portfolios returns were built using the size (Market Capitalization) and the value (Book-to-Market) as group criteria and were found in the French (2025) website together with the Market Factors, extracting the data from January 2003 to December 2022, 240 monthly observations. We refer to the original Fama and French (2015) article for the details of the portfolio construction, and we use the following abbreviation:

- SS_VH: Small Size – High Value portfolio.

[23] The acronym comes from likelihood estimation, Outer Product of Gradients (OPG), where the moments are the gradients (the scores), and their outer product is taken.

- SS_VM: Small Size – Medium Value portfolio.

- SS_VL: Small Size – Low Value portfolio.

- SB_VH: Big Size – High Value portfolio.

- SB_VM: Big Size – Medium Value portfolio.

- SB_VL: Big Size – Low Value portfolio.

The Fama-French portfolios and factors are constructed dynamically using data that accounts for both newly listed firms and firms that have been delisted (e.g., due to liquidation or mergers). Newly listed firms enter the portfolios based on their size and value characteristics, while delisted firms are removed. Returns are adjusted to include delisting returns where available, ensuring that the factor and portfolio data accurately reflect the current market and associated risks.

The independent variables of the integration model are the excess US market return time series $(R_{US,t})$ and the corresponding orthogonalised domestic return $(R^{\top}_{EU,t})$, as shown in equation 2.66. For the segmentation model, the independent variables are the excess EU market return time series $(R_{EU,t})$ and the corresponding orthogonalised North American return $R_{US\perp EU,t}$.
The main correlation tests were run for the time series and their results can be found in Section 2.6.1.2. The difference in the results between the White and the Breusch-Pagan test is explained because, the Breusch-Pagan test only checks for the linear form of heteroscedasticity, while the White Test is more generic but it can be less efficient when the number of regressors increase. From the low p value for some portfolio, we can conclude that we have heteroscedasticity and it is worth using the NSUR and GMM methods over OLS. Finally, the following definitions are provided to compare the results for the Brooks et al. model and the Jorion, Schwartz one[24].

Brooks et al.

Integration test:

- Total Integration (TI, also named complete integration in the article): λ_{Local}, the risk price of the component orthogonal to the domestic and global market factors, should not be statistically significant ($p > 0.05$), while the global risk price λ_{Global} should be significantly ($p < 0.05$) different from zero ($|\lambda_{\text{Global}}| > 0.1$)

- Partial Integration (PI): evidence that the local factor is priced, the global risk price λ_{Global} should be significantly (p<0.05) different from zero ($|\lambda_{\text{Global}}| > 0.1$) and the orthogonal risk price should be significant and different from zero as well

- Integration Rejected (IR): evidence that only the local factor is priced, the global factor risk price λ_{Global} should not be significant and the risk price of the local orthogonal factor should be

[24] We slightly changed the definitions compared with the Penco and Lucas (2024) article

significant and different from zero.
Segmentation test:
- Total Segmentation (TS, also named complete segmentation in the article): δ_{Global}, the risk price of the component orthogonal to the global and domestic market factors, should not be significant, while the local risk price δ_{Local} should be significantly $(p < 0.05)$ different from zero $(|\delta_{Local}| > 0.1)$
- Partial Segmentation (PS): the local risk price δ_{Local} should be significantly $(p < 0.05)$ different from zero $(|\delta_{Local}| > 0.1)$ and the orthogonal risk price should be significant and different from zero as well
- Segmentation Rejected (SR): evidence that only the global factor is priced, the local factor risk price should not be statistically significant and the global orthogonal factor risk price should be significant and different from zero.

Jorion and Schwartz

Total integration: the domestic market risk price λ_{Local} should not be significantly different from zero $(p > 0.05)$; i.e., $\lambda_{\text{Local}} \approx 0$.
Total segmentation (also named complete segmentation in the article): the global market risk price δ_{Global} should not be significantly different from zero $(p > 0.05)$; i.e., $\delta_{\text{Global}} \approx 0$.

2.6.1.2 Data Analysis

Cross-Correlation and Multicollinearity

The correlation matrix between global market factors is shown in Table 2.4, together with the Pearson correlation test results. As discussed, a high correlation between the market returns is shown with a high confidence level (p below 1%).

	$R_{EU,t}$	$R_{AS,t}$	$R_{JP,t}$	$R_{US,t}$
$R_{EU,t}$	1.0000 (<0.001)	0.8694 (<0.001)	0.6740 (<0.001)	0.8761 (<0.001)
$R_{AS,t}$	0.8694 (<0.001)	1.0000 (<0.001)	0.6543 (<0.001)	0.8105 (<0.001)
$R_{JP,t}$	0.6740 (<0.001)	0.6543 (<0.001)	1.0000 (<0.001)	0.6359 (<0.001)
$R_{US,t}$	0.8761 (<0.001)	0.8105 (<0.001)	0.6359 (<0.001)	1.0000 (<0.001)

Table 2.4: Correlation Matrix with p-values, Global Market Factors

The results show multicollinearity, which was the reason to use the orthogonal projection of the returns. As shown in Table 2.5, applying the projections will fix the multicollinearity problem.

	$R_{EU,t}$	$R_{AS\perp EU,t}$	$R_{US\perp EU,t}$	$R_{JP\perp EU,t}$
$R_{EU,t}$	1.0000 (0.0000)	-0.0098 (0.8802)	-0.0178 (0.7832)	-0.0048 (0.9413)
$R_{AS\perp EU,t}$	-0.0098 (0.8802)	1.0000 (0.0000)	0.2052 (0.0014)	0.1872 (0.0036)
$R_{US\perp EU,t}$	-0.0178 (0.7832)	0.2052 (0.0014)	1.0000 (0.0000)	0.1276 (0.0484)
$R_{JP\perp EU,t}$	-0.0048 (0.9413)	0.1872 (0.0036)	0.1276 (0.0484)	1.0000 (0.0000)

Table 2.5: Correlation Matrix with p-values, EU vs Orthogonalised Global Market Factors

Autocorrelation and Residual Diagnostics

Autocorrelation was tried via the Durbin Watson test, Table 2.6, where we report the result between the Big size and High Value Portfolio, the European market return and the US/EU market return projection. The ordinary least square results are also reported.

	Coefficient	**Std. Error**	**t-Statistic**
Intercept	-0.0866	0.1270	-0.6818
$R_{EU,t}$	1.1750	0.0234	50.1320
$R_{US\perp EU,t}$	-0.1024	0.0582	-1.7586
R^2		0.914	
Durbin-Watson		1.983	

Table 2.6: OLS Regression, EU Big Size High Value Portfolio on $R_{EU,t}$ and $R_{US\perp EU,t}$

Table 2.7 extends this diagnostic across all six portfolios. The values range from 1.78 to 1.98, indicating weak or no first order serial correlation in the residuals: $\text{Cov}(r_t, r_{t-1}) \neq 0$.

Portfolio	**Durbin-Watson**
r_{SS_VH}	1.8000
r_{SS_VM}	1.9067
r_{SS_VL}	1.7890
r_{SB_VH}	1.9833
r_{SB_VM}	1.7771
r_{SB_VL}	1.8945

Table 2.7: Durbin-Watson Statistics, EU Portfolio Residuals on $R_{EU,t}$ and $R_{US\perp EU,t}$

The Ljung–Box test in Table 2.8 indicates that two portfolios (e.g., SS_VL and SB_VL) exhibit significant residual autocorrelation within the first ten lags.

In Table 2.9, we report the residual correlation matrix between the European Size and Value portfolios and the EU market factor.

Portfolio	Statistic	p-value
r_{SS_VH}	18.661	0.045
r_{SS_VM}	14.201	0.164
r_{SS_VL}	27.747	0.002
r_{SB_VH}	18.143	0.053
r_{SB_VM}	14.400	0.156
r_{SB_VL}	18.893	0.042

Table 2.8: Ljung–Box Autocorrelation Test, EU Portfolio Residuals on $R_{EU,t}$

	r_{SS_VH}	r_{SS_VM}	r_{SS_VL}	r_{SB_VH}	r_{SB_VM}	r_{SB_VL}
r_{SS_VH}	1.000 (<0.0001)	0.776 (<0.0001)	0.443 (<0.0001)	0.377 (<0.0001)	-0.654 (<0.0001)	-0.610 (<0.0001)
r_{SS_VM}	0.776 (<0.0001)	1.000 (<0.0001)	0.902 (<0.0001)	-0.256 (0.015)	-0.333 (0.005)	-0.012 (0.059)
r_{SS_VL}	0.443 (<0.0001)	0.902 (<0.0001)	1.000 (<0.0001)	-0.594 (<0.0001)	-0.094 (0.006)	0.374 (0.004)
r_{SB_VH}	0.377 (<0.0001)	-0.256 (0.015)	-0.594 (<0.0001)	1.000 (<0.0001)	-0.620 (<0.0001)	-0.953 (<0.0001)
r_{SB_VM}	-0.654 (<0.0001)	-0.333 (0.005)	-0.094 (0.006)	-0.620 (<0.0001)	1.000 (<0.0001)	0.642 (0.207)
r_{SB_VL}	-0.610 (<0.0001)	-0.012 (0.059)	0.374 (0.004)	-0.953 (<0.0001)	0.642 (0.207)	1.000 (<0.0001)

Table 2.9: Residual Correlation Matrix with p-values, EU Portfolios on $R_{EU,t}$

The results show evidence of cross-correlation between residuals.

Heteroscedasticity Tests

Below, we also tested for heteroscedasticity Var$(r_t) \neq \sigma^2$, which causes the OLS estimates to be inefficient as it assumes constant error variance, while GLS and GMM that take into account the changing variance can make more efficient use of the data. Both White's test and the Breusch-Pagan (BP) test based on the residuals of the fitted model were run. The White test is robust to general forms of heteroscedasticity, including nonlinearity, while the BP test is sensitive to linear forms. For systems of equations, these tests are computed separately for the residuals of each equation.

These results suggest that smaller firms, particularly those with high book-to-market ratios, are more prone to heteroscedasticity in their return-generating process. This may reflect time-varying volatility or other nonlinear dynamics. In such cases, using heteroscedasticity robust standard errors or Generalized Least Squares may improve inference reliability.

The results of the test are shown in Table 2.10 for the European market return and US/EU market return projection model (segmentation model).

Equation	Test	Statistic	df	p-value
r_{SS_VH}	White	51.092	6	0.0000
r_{SS_VH}	Breusch-Pagan	6.501	2	0.0387
r_{SS_VM}	White	47.180	6	0.0000
r_{SS_VM}	Breusch-Pagan	3.086	2	0.2137
r_{SS_VL}	White	23.951	6	0.0002
r_{SS_VL}	Breusch-Pagan	2.049	2	0.3589
r_{SB_VH}	White	19.254	6	0.0017
r_{SB_VH}	Breusch-Pagan	0.774	2	0.6791
r_{SB_VM}	White	6.086	6	0.2980
r_{SB_VM}	Breusch-Pagan	1.294	2	0.5235
r_{SB_VL}	White	10.909	6	0.0532
r_{SB_VL}	Breusch-Pagan	1.832	2	0.4002

Table 2.10: White and BP Heteroscedasticity Test, EU Portfolios on $R_{EU,t}$ and $R_{US\perp EU,t}$

The Autoregressive Conditional Heteroskedasticity Lagrange Multiplier (ARCH LM) test results for the one-factor model with the European Market are shown in Table 2.11.

Portfolio	Statistic	p-value
r_{SS_VH}	1.651	0.895
r_{SS_VM}	6.042	0.302
r_{SS_VL}	25.375	0.000
r_{SB_VH}	21.653	0.001
r_{SB_VM}	16.410	0.006
r_{SB_VL}	8.635	0.125

Table 2.11: ARCH LM Test for Conditional Heteroscedasticity, EU Portfolios on $R_{EU,t}$

The White and Breusch-Pagan tests primarily detect cross-sectional heteroscedasticity—variance that depends on the level of fitted values or regressors. In contrast, the ARCH LM test targets time-series heteroscedasticity (conditional heteroscedasticity or volatility clustering). Here, SS_VL, SB_VH, and SB_VM show clear evidence of ARCH effects ($p < 0.01$), indicating that their variance also depends on lagged shocks up to five lags.

2.6.1.3 One-Factor Model

The first experiment is to estimate a one-factor beta-pricing model of six European portfolios excess returns regressed against the European Market factor excess returns. In the experiment

we compare the classic methods and the linear approximation results.

Risk Premium Estimation via Linear Approximation

We present the results for six EU portfolios grouped on size and book-to-market against the EU market excess returns using the different linear approximation methods:

- Taylor product, that refers to the first-order Taylor approximation of the product, proposed in equation 2.11, which is a convex unconstrained quadratic problem (specifically, a Least Squares problem).
- Taylor convex, which is the convex approximation proposed in equation 2.28, a Quadratic Program (QP) with linear constraints.
- MILP L1, hybrid piecewise linear approximation solved with Linear Programming (LP).
- MIQP L2, piecewise linear approximations via convex continuous combination that is based on L2-norm.
- Factor, the ratio substitution proposed in equation 2.14, which becomes a Convex Quadratic Program with quadratic constraints.
- Factor iter, the iterated version of Factor as per equation 2.20, which is again a Convex QP with quadratic constraints.
- LP, the linear program L1 method with McCormick envelope, equation 2.23.
- Nonlinear L1 , the nonlinear L1 method, equation 2.44.
- Nonlinear L2 , the nonlinear L2 method, equation 2.45.

We use Python packages (`cvxpy`, `gurobipy`, and `pulp`) together with custom code for all the methods implemented.

In Table 2.12 we show the result of the centred factor regression as formulated in equation 2.10. In Table 2.13 the results of the raw factor regression.

The results of the centred regression using linearised methods indicate that the estimated risk price λ_f are consistently close to zero across all approaches, with high standard errors and p-values near one, suggesting no significant factor risk price. The intercept estimates λ_0 are stable across methods but also statistically not significant, reflecting the limited explanatory power of the single-factor specification under the linearised frameworks. Despite an R^2 around 0.9 across all portfolios, the intercept estimates λ_0 are statistically insignificant, indicating limited pricing

Method	Parameter	Estimate	Std. Error[1]	t-Statistic	*p*-value
Taylor product	Intercept	0.7483	0.4674	1.6011	0.1096
Taylor product	Risk price	0.0111	0.4571	0.0242	0.9807
Taylor convex	Intercept	0.7483	0.4674	1.6011	0.1096
Taylor convex	Risk price	0.0111	0.4571	0.0242	0.9807
MILP L1	Intercept	0.8223	0.5160	1.5937	0.1112
MILP L1	Risk price	-0.1216	0.5089	-0.2390	0.8112
MIQP L2	Intercept	0.7478	0.4674	1.6000	0.1098
MIQP L2	Risk price	0.0116	0.4571	0.0254	0.9797
Nonlinear L1	Intercept	0.8290	0.5159	1.6070	0.1083
Nonlinear L1	Risk price	-0.1284	0.5088	-0.2523	0.8009
Nonlinear L2	Intercept	0.7483	0.4674	1.6012	0.1096
Nonlinear L2	Risk price	0.0110	0.4571	0.0242	0.9807
Factor	Intercept	0.7487	0.4674	1.6020	0.1094
Factor	Risk price	0.0107	0.4571	0.0233	0.9814
Factor iter	Intercept	0.7487	0.4674	1.6019	0.1094
Factor iter	Risk price	0.0107	0.4571	0.0234	0.9814
LP L1	Intercept	0.8252	0.5262	1.5683	0.1170
LP L1	Risk premium	-0.1284	0.5184	-0.2477	0.8044
FM	Intercept	0.7487	0.5060	1.480	0.2131
FM	Risk price	0.0107	0.4948	0.022	0.9838
OLS	Intercept	0.7487	0.4422	1.69	0.0918
OLS	Risk price	0.0107	0.4434	0.02	0.9808
FM GLS	Intercept	1.3615	0.4937	2.758	0.0063
FM GLS	Risk price	-0.6636	0.4945	-1.342	0.1809
SUR	Intercept	1.3782	0.4997	2.760	0.0063
SUR	Risk price	-0.6790	0.5009	-1.360	0.1766
GMM	Intercept	1.3962	0.5036	2.580	0.0061
GMM	Risk price	-0.6973	0.5067	-1.380	0.1700

[1] In all the variants we use the design matrix (dm) method for the inference estimation; see Appendix B

Table 2.12: One-Factor Risk Price Estimates, Centred Factor

error; however, the slope coefficients λ_f are also not significant, suggesting that the price of risk is small or uncertain. As a result, the product $\lambda_f \hat{\beta}_i$, which determines the magnitude of expected returns across portfolios, is negligible.

For the centred factor $\tilde{f}$ (also shown as cent.), which sample average is zero (i.e., demeaned), the following equivalence identity holds:

$$\lambda_{\tilde{f}} = \lambda_f + \bar{f} \tag{2.78}$$

In Appendix L, we show the GMM equivalence of raw and demeaned factors.

In the example below, the risk price $\lambda_{\tilde{f}}$, in the second last column of Table 2.13 is derived applying the conversion formula to the Taylor prod estimate: $\lambda_{\tilde{f}} = \lambda_f + \bar{f} = -0.656 + 0.667 = 0.011$, where λ_{f} is the risk price estimated using the raw factor; $\lambda_{\tilde{f}}$ is the derived equivalent risk price estimated with a demeaned factor; $\bar{f}$ is the time-series mean of the factor (e.g., 0.6671).

The adjusted test statistic under the centred normalization is computed as:

$$t_{\lambda_{\tilde{f}}} = t_{\lambda_f} \cdot \frac{\lambda_{\tilde{f}}}{\lambda_{\mathsf{f}}} = -1.435 \cdot \frac{0.011}{-0.656} \approx 0.024$$

Then the two-sided p-value is:

$$p_{\lambda_{\tilde{f}}} = 2\left(1 - \mathcal{T}_\nu\big(|t_{\lambda_{\tilde{f}}}|\big)\right) \approx 0.9807 \tag{2.79}$$

where $\mathcal{T}_\nu(\cdot)$ denotes the CDF of the Student's t-distribution with ν degrees of freedom (refer to Appendix B), and $|t_{\lambda_{\tilde{f}}}|$ is the absolute value of the adjusted statistic under the centred parametrisation.

For the two steps regression (Fama-MacBeth) centring the factor f (i.e., subtracting its sample mean so that $E[\tilde{f}] = 0$) does not affect the estimated factor loadings $\hat{\beta}_i$ in the first-step (time-series regressions), but it shifts the dependent variable in the second-step (cross-sectional regression). This removes the influence of the factor mean from the returns and clarifies the interpretation of the risk price λ_f as capturing compensation for zero-mean variation in factor exposure.

In contrast, when using the uncentred (raw) factor, the intercept λ_0 partially absorbs the effect of the factor mean, that can obscure the interpretation of λ_f and the contribution of factor exposure to expected returns.

However, the true economic value of the risk price is associated with the raw (uncentred) factor, as shown in the simulation results.

For the FM method, the average return in the second step are demeaned using the betas of the first step multiplied by the factor mean. In short, for FM OLS, the raw factor regression in the second step corresponds to demeaned factor regression of the moments method. In fact, in order to produce the FM raw equivalent we have to add the factor mean on the right-hand side, i.e., demean the average returns. In both the centred and raw risk price estimates tables, we use no rolling in the second step.

Method	Parameter	Estimate	Std. Error[1]	t-Statistic	*p*-value	$\lambda_{\tilde{f}}$	*p*-value$_{\tilde{f}}$
Taylor product	Intercept	0.7483	0.4674	1.6011	0.1096	–	–
Taylor product	Risk price	-0.6561	0.4571	-1.4354	0.1514	0.0111	0.9807
Taylor convex	Intercept	0.7483	0.4674	1.6011	0.1096	–	–
Taylor convex	Risk price	-0.6561	0.4571	-1.4354	0.1514	0.0111	0.9807
MILP L1	Intercept	0.8272	0.5158	1.6036	0.1090	–	–
MILP L1	Risk price	-0.7937	0.5088	-1.5598	0.1190	-0.1265	0.8036
MIQP L2	Intercept	0.7468	0.4673	1.5980	0.1103	–	–
MIQP L2	Risk price	-0.6546	0.4570	-1.4323	0.1523	0.0125	0.9781
Nonlinear L1	Intercept	0.8290	0.5159	1.6070	0.1083	–	–
Nonlinear L1	Risk price	-0.7955	0.5088	-1.5634	0.1182	-0.1284	0.8009
Nonlinear L2	Intercept	0.7483	0.4674	1.6011	0.1096	–	–
Nonlinear L2	Risk price	-0.6561	0.4571	-1.4354	0.1514	0.0111	0.9807
Factor	Intercept	0.7487	0.4674	1.6020	0.1094	–	–
Factor	Risk price	-0.6565	0.4571	-1.4363	0.1511	0.0107	0.9814
Factor iter	Intercept	0.7487	0.4674	1.6019	0.1094	–	–
Factor iter	Risk price	-0.6564	0.4571	-1.4362	0.1512	0.0107	0.9814
LP L1	Intercept	0.8252	0.5262	1.5683	0.1170	–	–
LP L1	Risk premium	-0.7956	0.5184	-1.5345	0.1251	-0.1284	0.8044
FM	Intercept	0.7487	0.2837	2.6396	0.0088		0.0088
FM	Risk price	-0.6565	0.2774	-2.3665	0.0187	0.0106	0.9692
OLS	Intercept	0.7483	0.4420	1.690	0.0919		0.0918
OLS	Risk price	-0.6561	0.4432	-1.480	0.1402	0.0106	0.9808
FM GLS	Intercept	1.3615	0.4937	2.758	0.0063		0.0063
FM GLS	Risk price	-1.3307	0.4945	-2.691	0.0076	-0.6636	0.1820
SUR	Intercept	1.3783	0.4997	2.760	0.0063		0.0063
SUR	Risk price	-1.3462	0.5009	-2.690	0.0077	-0.6791	0.1761
GMM	Intercept	1.3962	0.5051	2.760	0.0061		0.0061
GMM	Risk price	-1.3644	0.5067	-2.690	0.0076	-0.6973	0.1705

[1] In all the variants, we use the design matrix for the inference estimation; see Appendix B

Table 2.13: One-Factor Risk Price Estimates, Raw Factor

When using raw factors, the estimates of λ_f are negative for most methods, but remain statistically insignificant (large standard errors with high p-values). A negative λ_f suggests that higher exposure to the factor would be associated with lower expected returns, which is counter intuitive in typical risk-return frameworks where risk price are expected to be positive. After centring the factor, the estimates of λ_f shift close to zero across all approaches, reflecting that the centring process removes the influence of the factor mean and clarifies the interpretation of the risk price.

The results are consistent across the different convex approximation methods, which suggests that the different linear approximation techniques all provide a valid estimation of the parameters.

The LP L1, MILP L1 and Nonlinear L1 methods lead to different estimates than the other approaches, both for the intercept λ_0 and the risk price λ_f. However, their estimates are consistent within the L1 methods: $\lambda_0 \approx 0.829$ and $\lambda_f \approx -0.795$. In the earlier simulation, we have also shown how the L1-norm risk price estimate will show lower bias and variance than the L2 one, being more robust to outliers.

Risk Premium Estimation via Classical Methods

In this section, we estimate the cross-sectional asset pricing model using Fama-MacBeth (FM), NSUR and GMM approaches. We use SAS proprietary software `PROC MODEL` to implement the OLS, SUR, and GMM methods, and custom Python code for the Fama-MacBeth procedure with both OLS and GLS estimators. We consider a one-factor asset pricing model, a system of nonlinear equations, as defined in equation 2.10.

We show the different regression results in Table 2.12 for the demeaned factor and in Table 2.13 for the raw factor: the classic methods can be found below the double horizontal lines.

The GLS approach finds a statistically significant intercept, suggesting potential mispricing or omitted factors. However, the estimated risk price on the EU market factor is not statistically significant under either method. OLS and GLS yield substantially different estimates for both parameters, highlighting the importance of accounting for cross-sectional error correlation in asset pricing tests. The SAS GMM and SUR estimators account for the cross-sectional covariance of residuals across portfolios through the contemporaneous weighting matrix $\mathbf{\Sigma}^{-1}$. In addition, time-series heteroskedasticity and autocorrelation in the stacked moment conditions are addressed using the Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) variance estimator. Therefore in the FM GLS we implement the same estimator in the second step cross-sectional regression: the GLS also takes into account the autocorrelation as shown below.

For the cross-sectional regression step of Fama-MacBeth, we already defined:

- $\bar{\mathbf{R}}$: vector of average returns $(N \times 1)$.
- $\mathbf{B}$: matrix of estimated betas with intercept column $(N \times 2)$.
- $\mathbf{\Sigma}$: covariance matrix of residuals from the first-step time-series regressions $(N \times N)$.

The GLS estimator of the risk prices is:

$$\hat{\boldsymbol{\theta}}_{\text{GLS}} = \left(\mathbf{B}^\top \mathbf{\Sigma}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^\top \mathbf{\Sigma}^{-1} \bar{\mathbf{R}},$$

where $\hat{\boldsymbol{\theta}}_{\text{GLS}} = \begin{bmatrix}\hat{\lambda}_0 & \hat{\lambda}_f\end{bmatrix}^\top$. To account for autocorrelation, we define the time-t moment vector:

$$\mathbf{g}_t = \mathbf{B}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t, \quad \forall t$$

with components:

$$g_{t,1} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\Sigma^{-1}\right)_{ij}\varepsilon_{j,t}, \quad g_{t,2} = \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\left(\Sigma^{-1}\right)_{ij}\varepsilon_{j,t},$$

and where

$$\boldsymbol{\varepsilon}_t = \left(\varepsilon_{1,t},\, \varepsilon_{2,t},\, \ldots,\, \varepsilon_{N,t}\right)^\top$$

are the time-series residuals defined in equation (2.1).

The long-run covariance matrix of these moments is estimated by:

$$\widehat{\boldsymbol{\Omega}}_{HAC} = \Gamma_0 + \sum_{h=1}^{H} k(h)\left(\Gamma_h + \Gamma_h^\top\right),$$

with:

$$\Gamma_0 = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\,\mathbf{g}_t^\top, \quad \Gamma_h = \frac{1}{T}\sum_{t=1+h}^{T}\mathbf{g}_t\,\mathbf{g}_{t-h}^\top,$$

and $k(h)$ is a kernel weighting function. In this work we use the Bartlett kernel:

$$k(h) = 1 - \frac{h}{H+1}, \quad \text{for } h = 0, \ldots, H.$$

with lag order $H = 4$:

$$k(0) = 1, \quad k(1) = 1 - \frac{1}{5} = 0.8,, \; k(2) = 0.6, \; k(3) = 0.4, \; k(4) = 0.2.$$

Thus, the estimator incorporates the instantaneous cross-sectional covariance, Γ_0, and the autocovariances up to lag H. The instantaneous cross-sectional covariance corresponds to the heteroscedasticity-robust covariance estimator, White (1980)

The HAC variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{\text{GLS}}$ is:

$$\widehat{\text{Var}}\left(\hat{\boldsymbol{\theta}}_{\text{GLS}}\right) = \frac{1}{T}\left(\mathbf{B}^\top\boldsymbol{\Sigma}^{-1}\mathbf{B}\right)^{-1}\widehat{\boldsymbol{\Omega}}_{HAC}\left(\mathbf{B}^\top\boldsymbol{\Sigma}^{-1}\mathbf{B}\right)^{-1}. \tag{2.80}$$

Standard errors are computed as the square roots of the diagonal entries of this matrix: autocorrelation of the residuals affects only the estimated standard errors, the GLS estimator itself

remains unchanged.

Although Fama-MacBeth (OLS) estimation minimizes the sum of squared residuals (SSR) independently for each portfolio, it ignores cross-sectional correlations in the residuals. In contrast, its GLS version, as well as SAS GMM (or other GLS-based methods like NSUR), solve a system of equations by minimizing a quadratic form of moment conditions, which accounts for cross-equation residual covariance. As a result, GLS estimation does not directly minimize SSR, but instead seeks efficiency and pricing consistency (same λ_f across portfolios) by balancing fit across all equations. This explains why OLS may yield a lower total SSR, while GMM and GLS provide more reliable inference results.

In GMM, SUR, and even OLS SAS, which is estimated via moments conditions, when the factor f is centred (demeaned), equivalent to zero mean, the equivalence between centred and raw factors holds (equation 2.78). This is because the GMM moment conditions explicitly account for the factor mean $\bar{f}$, so centring f and setting zero mean, shifts the estimate of the risk price λ_f by the factor mean $\bar{f}$. The two-sided p-value given a test statistic is computed as before using the formula: $t_{\lambda_{\tilde{f}}} = t_{\lambda_f} \cdot \frac{\lambda_{\tilde{f}}}{\lambda_{\mathrm{f}}}$.

We have already seen that when using an uncentred factor, the intercept picks up part of the variation in returns that would otherwise be attributed to the slope, see Cochrane (2005). In contrast, centring the factor removes this overlap, forcing the slope to explain only the cross-sectional spread in returns. If the data are such that high-β portfolios do not earn high average returns (maybe even reverse), then the slope will be negative, even though the factor mean is positive. In this case, as we observe from our data, the raw (uncentred) estimate of the risk price is more significant (lower p-value), because its value is larger in magnitude, while the standard error does not change. In fact, adding a positive mean to a negative factor will decrease the magnitude of the factor, decreasing the statistical significance. However, in the general case, if returns increase with beta, and the factor mean is positive, then centring the factor improves the estimation and increases the statistical significance of the risk price.

For what concerns the equivalence identity, equation 2.78, we have already noticed that using the demeaned return $\tilde{R}_i = \bar{R}_i - \beta_i \bar{f}$, i.e., demeaning the left-hand side, in the second step of the FM regression is equivalent to using the raw data with OLS SAS. This is the reason we reported FM demeaned average return in the raw factor table.
The risk price adjusted for common drift implies the best economic interpretation, for this reason we do not report results without drift. The true risk price (adjusted for common drift) calculated using the centred factor and the intercept in the regression leads to a high p-value, while the raw factor regression without intercept (total average return per unit β, no drift split) gives a

significant estimate, but the true risk price is hidden.

Solver Performance

For comparing performance, we tested each method with the centred and raw factor. As expected, raw specification runs were consistently slower: when using the raw factor, the intercept and factor are more collinear, which worsens conditioning and increases iterations. The computational performance and accuracy are reported, sorted by Sum of Square Residuals (SSR) metric, in Table 2.14.

Method	Package	Solver	CPU Time	Iter	Avg Time	SSR	RMSE
Nonlinear L2 cent.	scipy	SLSQP	0.0218	17	0.0013	3887.82	1.6431
Nonlinear L2 raw	scipy	SLSQP	0.0547	16	0.0034	3887.82	1.6431
Factor cent.	cvxpy	OSQP	0.0297	75	0.0004	3887.82	1.6431
Factor raw	cvxpy	OSQP	0.2601	75	0.0035	3887.82	1.6431
Taylor product cent.	cvxpy	OSQP	0.0576	8	0.0072	3887.82	1.6431
Taylor product raw	cvxpy	OSQP	0.3873	8	0.0484	3887.82	1.6431
Taylor-convex cent.	cvxpy	ECOS	0.0863	5	0.0173	3887.82	1.6431
Taylor-convex raw	cvxpy	ECOS	1.1918	8	0.1490	3887.82	1.6431
MIQP L2 cent.	gurobipy	GRB	1.2165	1	1.2165	3887.83	1.6431
MIQP L2 raw	gurobipy	GRB	1.0102	1	1.0102	3887.83	1.6431
FM OLS raw	numpy	LU	0.0021	1	0.0021	3888.57	1.6433
FM OLS cent.	numpy	LU	0.0029	1	0.0029	3888.57	1.6433
LP L1 cent.	gurobipy	GRB	1.5421	1558	0.0010	3912.63	1.6484
LP L1 raw	gurobipy	GRB	1.5593	1579	0.0010	3912.63	1.6484
Nonlinear L1 cent.	scipy	SLSQP	0.0178	50	0.0004	3919.10	1.6497
Nonlinear L1 raw	scipy	SLSQP	0.1375	52	0.0026	3919.10	1.6497
MILP L1 cent.	gurobipy	GRB	3.7612	1	3.7612	3919.13	1.6497
MILP L1 raw	gurobipy	GRB	6.0705	1	6.0705	3919.09	1.6497

Table 2.14: Solver CPU Time and Accuracy

The python packages used are: cvxpy with OSQP (Operator Splitting Quadratic Program Solver); gurobipy with GRB (Gurobi Optimizer) using the Dual Simplex algorithm for the LP method; scipy with SLSQP (Sequential Least Squares Programming); numpy with LU (Lower Upper decomposition LAPACK's dgesv direct linear equation solving).

CPU time. The two-step Fama-MacBeth formulation that use LU decomposition is the fastest method; however its estimation is affected by the errors-in-variables problem, and the accuracy is not aligned with the other methods, reporting higher SSR and RMSE. The convex formulations based on CVXPY quadratic programming outperforms MILP approaches. On average, the Factor

Product requires only 0.03 seconds, that is half of the time compared to 0.06 seconds required for Taylor-product. Nonlinear optimisation problems are traditionally considered more computationally intensive; however, it seems that for this problem class, the MILP solvers introduce significant CPU time overhead, despite the convexity of the underlying objective (where any local minimum is also a global minimum). It is worth noting that the Product factor and Taylor convex approximation, which are novel approximation methods for risk premium estimation, result in being as fast as the Nonlinear L2 and the Taylor approximation, respectively. While the MILP L1 Linear Program is the slowest method, the fully linear programming (LP L1) approach via McCormick relaxation delivers performance comparable to nonlinear methods. However, whereas the nonlinear convex methods rely on smooth L2 objectives that can be sensitive to heavy tails, LP and MILP piecewise linear discretizations of the product are exact at the chosen breakpoints (given tightened bounds and chosen grid tolerances), and in general more accurate.

Accuracy. Beyond the Fama-MacBeth results, which are impacted by the two-step approach, in general, we observe that the sum of squared residuals (SSR) and root mean squared error (RMSE) are virtually identical across all approximation methods. Larger deviations are observed for the L1 methods, which produce higher SSR and RMSE (3919.1 and 1.649).

The results bring evidence that smooth quadratic and convex methods implemented with CVXPY and OSQP provide the right balance of speed and accuracy for this problem.

However, for the L1-norm methods, we should compare the results using the L1 metric, which is the sum of the absolute errors (SAE):

$$\text{SAE} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left| R_{i,t} - \hat{R}_{i,t} \right| \tag{2.81}$$

The results are reported in Table 2.15.

We have already noted that from the simulations (Figure 2.1), the L1 risk price estimate is closer to the true value estimates (lower bias and variance) than the L2 one. From Table 2.15, we see that in terms of the SAE metric, L1 methods outperform L2 methods. The result is as expected, since in the presence of outliers,when the standard assumptions of normality and homoscedasticity do not hold, L1 provides a more robust alternative. However, while the sum of squared residuals (SSR) naturally favours L2-norm methods and the sum of absolute errors (SAE) favours L1-norm methods, we find that in the presence of heavy tails, L1-norm approaches yield lower bias and variance in the full set of parameters estimation (see Figure 2.1). Especially, L1-norm methods appear more appropriate for risk price estimation under heavy-tailed distributions, as observed in the case of the EU market factor (see Table 2.1) and its associated residual distribution (see

Method	Parameter	Estimate	Std. Error	t-Stat	*p*-value	SAE
LP L1	Intercept	0.8252	0.5262	1.57	0.1170	1734.98
LP L1	Risk premium	-0.7956	0.5184	-1.53	0.1251	1734.98
MILP L1	Intercept	0.8272	0.5158	1.60	0.1090	1734.99
MILP L1	Risk price	-0.7937	0.5088	-1.56	0.1190	1734.99
Nonlinear L1	Intercept	0.8290	0.5159	1.61	0.1083	1734.99
Nonlinear L1	Risk price	-0.7955	0.5088	-1.56	0.1182	1734.99
MIQP L2	Intercept	0.7484	0.5035	1.49	0.1372	1740.47
MIQP L2	Risk price	-0.6562	0.4921	-1.33	0.1823	1740.47
Nonlinear L2	Intercept	0.7487	0.4674	1.60	0.1093	1740.41
Nonlinear L2	Risk price	-0.6561	0.4571	-1.44	0.1512	1740.41

Table 2.15: L1 Metric Raw Estimation Results

Table 2.2).

Results Comparison

The risk price estimates obtained through the linear approximation methods are consistent with the estimates computed with classic approaches, such as Fama-MacBeth, SUR, and GMM. In both the centred and raw factor regressions, λ_f remains statistically insignificant across all methods. The intercept estimates have similar magnitudes and statistical significance among the linear and classic methods, under the homoscedasticity assumption (OLS), that confirms the validity of the linear approximation approach as an alternative estimation technique. It is worth noting that, the linear approximation method represents a novel contribution to the literature, capable of handling nonlinearities through Taylor or piecewise approximations while preserving comparability with traditional econometric estimators. Especially the Taylor Convex and the Product Factor approximation methods have the merit of matching the accuracy of other approaches while offering superior CPU time performance.

2.6.1.4 Integration and Segmentation

We now run the integration and segmentation analysis following the approaches of Jorion and Schwartz (1986) and Brooks and Iorio (2009).

Here, we use six EU portfolios, grouped by size and book-to-market, to estimate a system of non-linear equations that regress portfolio returns on market factors from North America, Asia-Pacific, and Japan[25]. For verifying the results, different techniques have been used (Fama-MacBeth,

[25] In the article Penco and Lucas (2024), given the importance of the US economy in the international financial

Nonlinear Seemingly Unrelated Regression and the General Method of Moments[26]. As previously explained, we proceed as follows:
1. Model construction. We build two competing models for asset pricing: the integrated and the segmented model. This requires the orthogonalization of the local factors: orthogonal projections are taken when building the integrated model because the local factors can be some non-significant proportion of the international factors.
2. Estimation via NSUR. We use the Nonlinear Seemingly Unrelated Regression NSUR, Zellner (1962), to estimate the parameters of the equations defined in the first step: if the errors are normally distributed the NSUR estimator is also a maximum likelihood estimator .
3. Fama–MacBeth procedure. We apply the Fama-MacBeth method to obtain cross-sectional estimates of the model parameters.
4. Generalized Method of Moments, Finally, we estimate the parameters using the Generalized Method of Moments GMM, Richardson and MacKinlay (1991), to verify the integration or segmentation of the markets, relaxing the assumption of normalization of the assets returns.

The values of estimates coincide among different methods (FM, SUR and GMM). This seems consistent with other literature results, see: Shi and Li (2019), Sarisoy et al. (2024), Anatolyev and Mikusheva (2022). The significance can be improved enlarging the number of portfolios and the range of beta, the six portfolio used are all highly correlated with the market factors.

The integration results are reported in Table 2.16, drawing from the methodology discussed in the article Brooks and Iorio (2009), a structured interpretation of the EU portfolio integration shows:

- $\lambda_{US} = -1.955$ is negative and statistically significant at the 1% level (p = 0.0011), indicating that global risk is priced in the cross-section of returns.

- $\lambda_{EU} = 0.733$ is not statistically significant (p = 0.1539), suggesting that EU orthogonal specific risk is not priced.

- All portfolios show strong and statistically significant exposure to both factors: for the global US factor, the estimated betas range from 0.91 to 1.20, all with p-values < 0.0001; for the orthogonal EU factor, the estimated betas range from 0.78 to 1.25, all with p-values < 0.0001.

market, we also perform the nonlinear regression of European, Asian and Japanese of six size and book-to-market portfolios against the North America market factors. Finally, we also use the European economy as global market, and we run the regression of North America, Asian and Japanese of six size and book-to-market portfolios against the European market factors. These results are not reported here for brevity.

[26] In the article we used Maximum Likelihood Estimation (MLE) instead of Fama-MacBeth. Reviewing the results, we found some typos reported in the article that we corrected.

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
US_EU	SUR	λ_{US}	-1.955	0.001	***	Statistically significant and economically large; consistent with integration	TI
US_EU	SUR	λ_{EU}	0.733	0.154		Estimate is not statistically significant; no evidence of segmentation	
US_EU	SUR, JS	λ_{EU}	0.098	$<.001$	***	Statistically significant and close to zero; consistent with total integration	TI
US_EU	GLS	λ_{US}	-1.85	0.001	***	Consistent with SUR; significant global factor supports integration	TI
US_EU	GLS	λ_{EU}	0.623	0.1911		Estimate is not statistically significant; no evidence of segmentation	
US_EU	GMM	λ_{US}	-2.102	0.002	***	Consistent with SUR; significant global factor supports integration	TI
US_EU	GMM	λ_{EU}	0.805	0.121		Estimate not close to zero, but not statistically significant; integration not rejected	
AS_EU	SUR	λ_{AS}	0.303	0.785		Estimate not statistically significant; no evidence of integration	
AS_EU	SUR	λ_{EU}	-1.200	0.049	**	Statistically significant and economically large; integration rejected	IR
AS_EU	SUR, JS	λ_{EU}	-0.201	0.028	**	Statistically significant, but not close to zero; suggests segmented pricing	
AS_EU	GLS	λ_{AS}	0.07	0.941		Estimate is not statistically significant; no evidence of integration	
AS_EU	GLS	λ_{EU}	-1.071	0.046	**	Statistically significant and economically large; inconsistent with integration	IR
AS_EU	GMM	λ_{AS}	0.229	0.832		Estimate not statistically significant; no evidence of integration	
AS_EU	GMM	λ_{EU}	-1.190	0.042	**	Statistically significant and economically large; inconsistent with integration	IR
JP_EU	SUR	λ_{JP}	1.935	0.367		Estimate not statistically significant; no evidence of integration	
JP_EU	SUR	λ_{EU}	-3.025	0.105		Estimate is not statistically significant; no evidence of segmentation	
JP_EU	SUR, JS	λ_{EU}	-0.176	0.028	**	Statistically significant, but not close to zero; suggests segmented pricing	
JP_EU	GLS	λ_{JP}	0.127	0.923		Estimate not significant; no evidence of integration	
JP_EU	GLS	λ_{EU}	-1.237	0.283		Estimate not statistically significant; segmentation not supported	
JP_EU	GMM	λ_{JP}	0.841	0.580		Estimate not statistically significant; no evidence of integration	
JP_EU	GMM	λ_{EU}	-2.093	0.118		Estimate not close to zero, but not significant; inconclusive regarding integration	

Table 2.16: Equities Integration Test Results

- We use 2σ and 3σ confidence intervals, marking with ** for p-values below 0.05, with *** for p-values below 0.003, and with * for p-values below 0.1.

This result supports the idea that, despite the strong exposure to both global and local risks, only the global risk is rewarded in expected returns, while local EU-specific shocks do not command a premium. This pattern is consistent with full financial integration.

The segmentation results are reported in Table 2.17. The interpretation is as follows:

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
US_EU	SUR	δ_{EU}	-1.306	0.014	**	Significant and different from zero, which shows segmentation	
US_EU	SUR	δ_{US}	-0.990	0.036	**	Significant, and different from zero to determine partial segmentation	PS
US_EU	SUR, JS	δ_{US}	-2.538	0.002	***	However significant, it is not close to zero to determine segmentation	
US_EU	GLS	δ_{EU}	-1.298	0.009	**	Consistent with SUR; significant local factor supports segmentation	
US_EU	GLS	δ_{US}	-0.890	0.044	**	Consistent with SUR, there is no total segmentation	PS
US_EU	GMM	δ_{EU}	-1.388	0.015	**	Consistent with SUR; significant local factor supports segmentation	
US_EU	GMM	δ_{US}	-1.078	0.028	**	Consistent with SUR, there is no total segmentation	PS
AS_EU	SUR	δ_{EU}	-0.951	0.112		Estimate not statistically significant; no evidence of segmentation	
AS_EU	SUR	δ_{AS}	1.180	0.132		Estimate not statistically significant; no evidence of segmentation rejection	
AS_EU	SUR, JS	δ_{AS}	4.772	0.026	**	However significant, it is not close to zero to determine segmentation	
AS_EU	GLS	δ_{EU}	-1.012	0.056		Consistent with SUR	
AS_EU	GLS	δ_{AS}	1.005	0.144		Consistent with SUR	
AS_EU	GMM	δ_{EU}	-1.00	0.102		Consistent with SUR	
AS_EU	GMM	δ_{AS}	1.15	0.125		Consistent with SUR	
JP_EU	SUR	δ_{EU}	-1.397	0.039	**	Significant and different from zero, which shows segmentation	TS
JP_EU	SUR	δ_{JP}	2.706	0.208		Not Significant, which shows segmentation	
JP_EU	SUR, JS	δ_{JP}	35.901	0.658		It is not close to zero, nor it is significant to determine segmentation	
JP_EU	GLS	δ_{EU}	-1.345	0.007	**	Significant and different from zero, which shows segmentation	TS
JP_EU	GLS	δ_{JP}	0.608	0.644		Not Significant, which shows segmentation	
JP_EU	GMM	δ_{EU}	-1.386	0.014	**	Significant and different from zero, which shows segmentation	TS
JP_EU	GMM	δ_{JP}	1.60	0.291		Not Significant, which shows segmentation	

Table 2.17: Equities Segmentation Test Results

- λ_{EU} (local EU market risk price) is strongly negative and highly significant ($p < 0.001$), implying that the local EU factor is priced in the cross-section of returns. This supports partial segmentation, as the EU factor drives expected returns.

- λ_{US} (orthogonal component of US and EU factors is also negative and significant at the 1% level, suggesting that this component—though orthogonal to the EU factor—also carries pricing information. This points to a non-trivial influence from a global dimension (i.e., integrated information beyond the local EU factor), though not necessarily full integration.

- $\beta_{i,US}$ (exposure to global orthogonal factor, shown in the Apendix) are small and mostly insignificant, meaning EU portfolios have little sensitivity to the orthogonal global factor. This weakens the case for full integration—if the global factor were priced and portfolios

were exposed to it, we would expect significant $\beta_{i,US}$ estimates.

- $\beta_{i,EU}$ (exposure to local EU factor) are statistically significant across all portfolios, consistent with segmentation, where local risk is a primary driver of expected returns.

Our results show that partial integration and partial segmentation can be reported simultaneously as our model is not able to provide a cut off value for partial integration nor for partial segmentation.
From the integration model, total integration is confirmed: the global US factor is priced while the local EU factor is not significantly different from zero. From the segmentation model, the significance of δ_{US} (orthogonal US-EU) suggests that purely local pricing is not sufficient—some influence from globalized markets exists. The result falls somewhere between full integration and partial segmentation, which is most consistent with a hybrid scenario, where the EU market prices its own risk (local factor), with some influence from global dynamics, but not enough exposure to consider the markets fully integrated. This aligns well with Brooks et al.'s interpretation framework, where significant domestic premia and insignificant or weak international factor exposures point to segmentation, while shared significant premia with high exposure would imply integration.

When we consider as global market Asia Pacific, the integration is rejected, while there is no significant result that can be drawn for Japan as a global integrated market. About the segmentation, there is not significant result for the Asian Pacific market while the Japan global market is fully segmented in the period considered.

2.6.1.5 Rolling Regression

We have extended the data set period, and have now run a 20 years window rolling regression (240 data points), starting from January 1998 to December 2022 (25 years period). In order to achieve a sufficient statistical significance we need a time series of at least 240 points, therefore we have only 61 data points which corresponds to 5 years rolling by month (January 1998 - December 2017, February 1998 - January 2018, ..., January 2002 - December 2022). The rolling windows are independent although partially overlapping due to the month rolling, we aim to monitor the trend every 12 months: 1999-2019, 2000-2020, 2001-2021, 2002-2022. We use the start date of the 240 points time series in the plots. The results are reported below, Figures 2.3 - 2.6. The left axis shows the risk price and the right axis shows the corresponding p-values.

Tables 2.18 and 2.19 contain the test hypotheses.

The factor is significantly priced ($p < 0.05$), with an economically meaningful estimate, indicating

$H_0^{\text{Integration}}$	$\lambda_{\text{US}} \neq 0$ and $\lambda_{\text{EU}\perp\text{US}} = 0$ (EU portfolios are fully integrated with the global market; only global risk is priced)
$H_1^{\text{Integration}}$	$\lambda_{\text{US}} \neq 0$ and $\lambda_{\text{EU}\perp\text{US}} \neq 0$ (EU portfolios are not fully integrated; EU-specific risk is also priced)

Table 2.18: Equities Integration Hypotheses

$H_0^{\text{Segmentation}}$	$\lambda_{\text{EU}} \neq 0$ and $\lambda_{\text{US}\perp\text{EU}} = 0$ (EU portfolios are fully segmented; only local risk is priced)
$H_1^{\text{Segmentation}}$	$\lambda_{\text{US}} \neq 0$ and $\lambda_{\text{US}\perp\text{EU}} \neq 0$ (EU portfolios are not fully segmented; orthogonal global risk is also priced)

Table 2.19: Equities Segmentation Hypotheses

$\lambda \neq 0$. When the p-value is not significant, the factor is not significantly priced; we fail to reject the null hypothesis that $\lambda = 0$.

From the integration rolling, Figure 2.3, the global US factor is always different from zero and significant at 1% confidence level, which shows full integration. The orthogonal EU US component, Figure 2.4, is different from zero and significant in the first half of the 5 year period considered till 2001 circa (1998-2021), indicating partial integration, and then it is not significant at the 5% level (2001-2022) with a couple of exceptions. It seems that EU and US market are partially integrated in the first 3 years roll (1998-2021) and then switch to full integration in the next two years (2001-2022).

From the segmentation rolling, the local EU factor is always different from zero and significant, see Figure 2.5. The orthogonal US EU component, Figure 2.6, is different from zero and significant at the 5% confidence level in the second half of the period considered from 2001 circa (2001-2022), while in the first half it is not significant at 5% level (1998-2021) with few exceptions. The EU and US market looks to be fully segmented in the first three years (1998-2021) and then switch to partial segmentation in the last two year roll (2001-2022).

The overlapping rolling windows, reflects a smoothed dynamics rather than discrete shifts in the estimated risk prices, that provide evidence of regime classification.

The integration test suggests that EU portfolios are fully integrated with the global market from 2001 to 2022, as only the US factor is significantly priced. However, the segmentation test

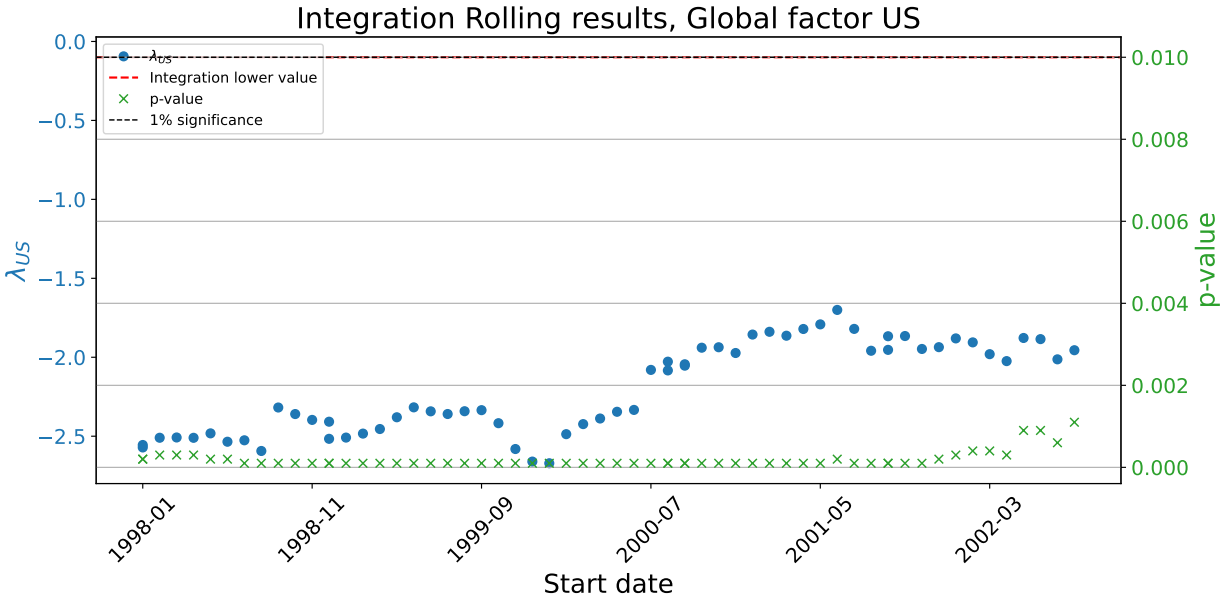

Figure 2.3: Equities Integration Rolling Regression Results from 1998 to 2022, Global factor

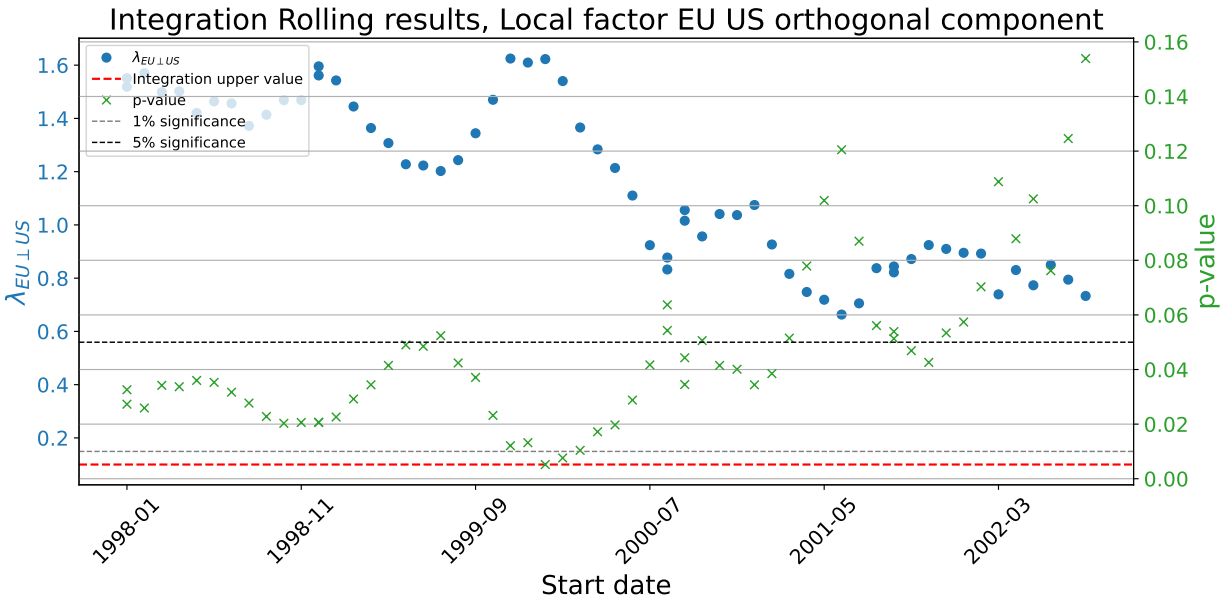

Figure 2.4: Equities Integration Rolling Regression Results from 1998 to 2022, Local factor

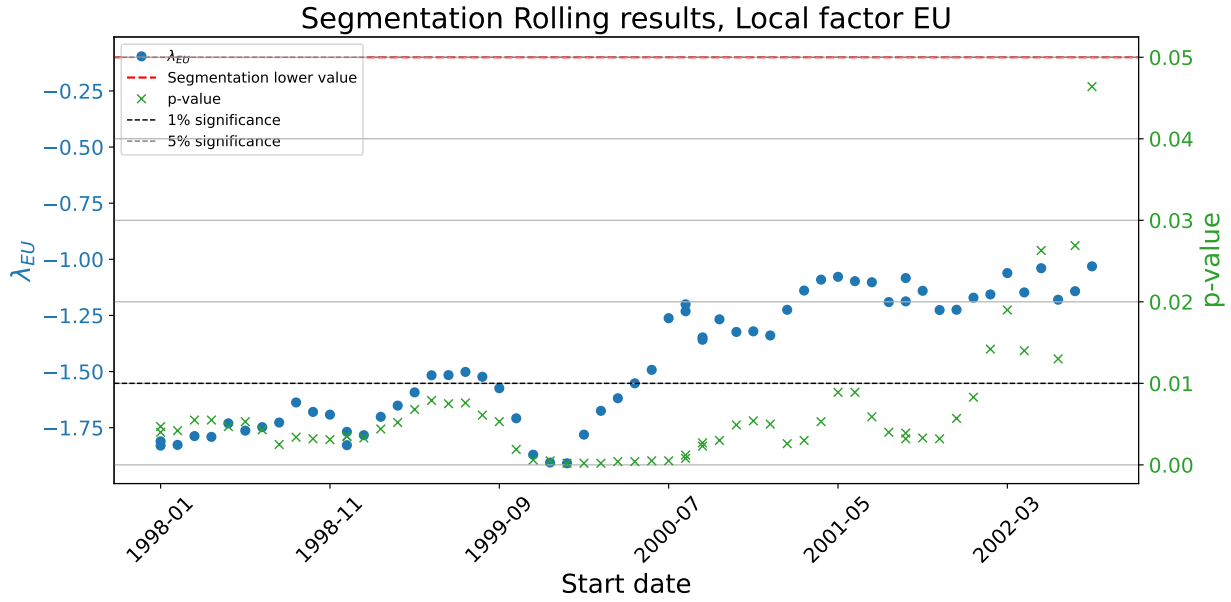

Figure 2.5: Equities Segmentation Rolling Regression Results from 1998 to 2022, Local factor

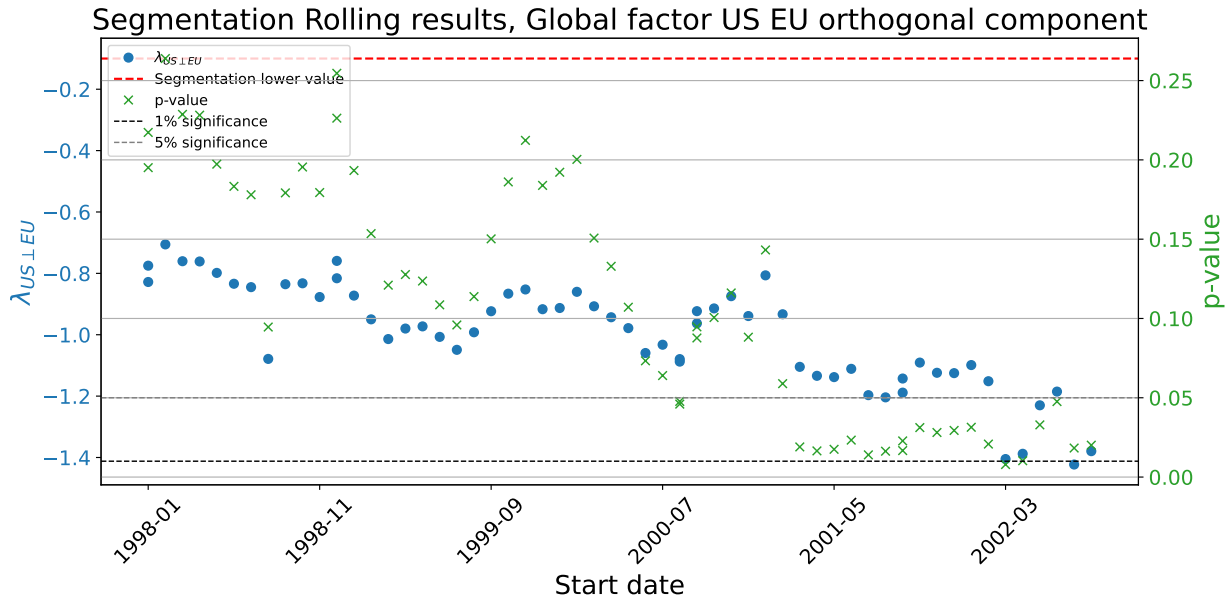

Figure 2.6: Equities Segmentation Rolling Regression Results from 1998 to 2022, Global factor

over the same period shows that EU portfolios price both local EU and orthogonalised US risk. However, from 1998 to 2021, the segmentation test suggest that the markets are fully segmented, while the integration test suggest that the global factor is also priced in the same period.

This inconsistency indicates either model instability or overlapping explanatory power between the orthogonalised and local components. Therefore, the results should be interpreted with precaution: there is some evidence of stronger integration in the second half, starting from 2001, and more segmentation in the first half starting from 1998. This transition from segmentation to integration beginning around 2001, with estimates stabilizing by 2003 coincides with post-dot-com recovery and Euro adoption in EU.

2.6.2 Commodities Indices

Similar to the global stock market integration study, this section aims to estimate the risk premium and analyse the degree of integration between Oil-related exchange-traded funds (ETFs) and global commodity markets, namely natural gas (GAS), aluminum (AL), and soybean (SOY). We use the OIL commodity as the local market. All returns are computed as excess percentage returns, we use the US Coupon bond risk-free rate as the benchmark.

2.6.2.1 Data Processing

The Oil ETFs ($N = 6$) analysed include:

- United States Oil Fund (USO): this ETF tracks the Oil commodity price as it triggers daily changes in WTI crude oil spot prices via next-month futures, it is designed for short-term crude oil hedging.
- Invesco DB Oil Fund (DBO): this ETF protects from crude oil exposure via an optimized futures roll strategy to minimize contango loss, maximise backwardation gain. In contango, storage costs, insurance, or financing drive futures prices up, the futures price F_t is above the spot price S_t: $F_t > S_t$. When a next month rolling strategy is in place (like for USO), the investor sells low and buys high, a loss is triggered. DBO follows an optimised rolling strategy picking WTI contracts along the WTI futures curve with the least contango or most backwardation.
- iShares US Oil & Gas Exploration & Production ETF (IEO): invests in US oil and gas exploration and production companies.

- iShares Global Energy ETF (IXC): includes large-capitalisation (cap) US energy firms and international companies.
- iShares U.S. Energy ETF (IYE): includes large- and mid-cap US energy sector companies, including oil producers and energy services.
- Energy Select Sector SPDR Fund (XLE): invests in S&P 500 energy-sector companies.

The reference commodity factors are:

- Crude Oil (OIL).
- Natural Gas (GAS).
- Aluminum (AL).
- Soybean (SOY).

The analysis covers $T = 180$ monthly observations from January 2007 to December 2022. The ETF price data are downloaded from Yahoo Finance (2025) while the commodity prices are extracted from the International Monetary Fund, IMF (2025), historical price database.

2.6.2.2 Data Analysis

Cross-Correlation and Multicollinearity

In the case of global market economies, there was strong multicollinearity among the factors, and we had to use the orthogonal components in the integrated model. In Table 2.20, we report the Variance Inflation Factor (VIF) for all two-factor combinations which quantifies how much a factor's coefficient variance is inflated due to collinearity, and it is computed as:

$$\text{VIF}_l = \frac{1}{1 - R_l^2} \quad \text{for } l = 1, \ldots, \bar{\ell}$$

where R_l^2 is the R^2 from regressing one-factor on the other factor in the integrated model and $\bar{\ell} = 4$.

- VIF < 4: satisfactory (low multicollinearity).
- VIF > 4: needs transformation (medium to high collinearity).
- VIF > 10: problematic (very high collinearity).

	EU	US	AS	JP
EU	-	4.300970	4.094873	1.832412
US	4.300970	-	2.914434	1.678825
AS	4.094873	2.914434	-	1.748523
JP	1.832412	1.678825	1.748523	-

Table 2.20: VIF Global Markets

The global market factors exhibit moderate multicollinearity, particularly between European and US market excess returns. This justifies the use of orthogonalised factors in the integration models.

In Table 2.21, we show the VIFs for the commodity factors.

	OIL	GAS	AL	SOY
OIL	-	1.011982	1.294316	1.041972
GAS	1.011982	-	1.039341	1.043415
AL	1.294316	1.039341	-	1.043357
SOY	1.041972	1.043415	1.043357	-

Table 2.21: VIF Commodity Indices

All VIFs values are near 1, confirming negligible collinearity for the commodity factors (OIL, GAS, AL, SOY): the factors are independent and there is no real need for the orthogonalization.

Autocorrelation and Residual Diagnostics

In Figure 2.7, we compare the residual variance across portfolios (Fama French, Equities vs ETFs).

The commodity ETFs have much higher and dispersed residual variances, from 40 to 90, compared to the Fama-French size and value portfolios, that are mostly clustered between 2 and 4.

As we will see later, in the one-factor regression, the large heteroscedasticity in the commodity data makes the estimated covariance matrix ill-conditioned, so feasible GLS becomes unstable or underperforms OLS, whereas it works well for the Equities data.

Figure 2.8 shows the residual autocorrelation for both time series: Fama French equities portfolios and commodities ETFs. For Equities, AR(1) residual autocorrelation is very low (from 0.01 to 0.10), which shows no time dependence: residuals are effectively white noise. The residual variance is low and fairly balanced across portfolios. OLS and GLS give similar results because the

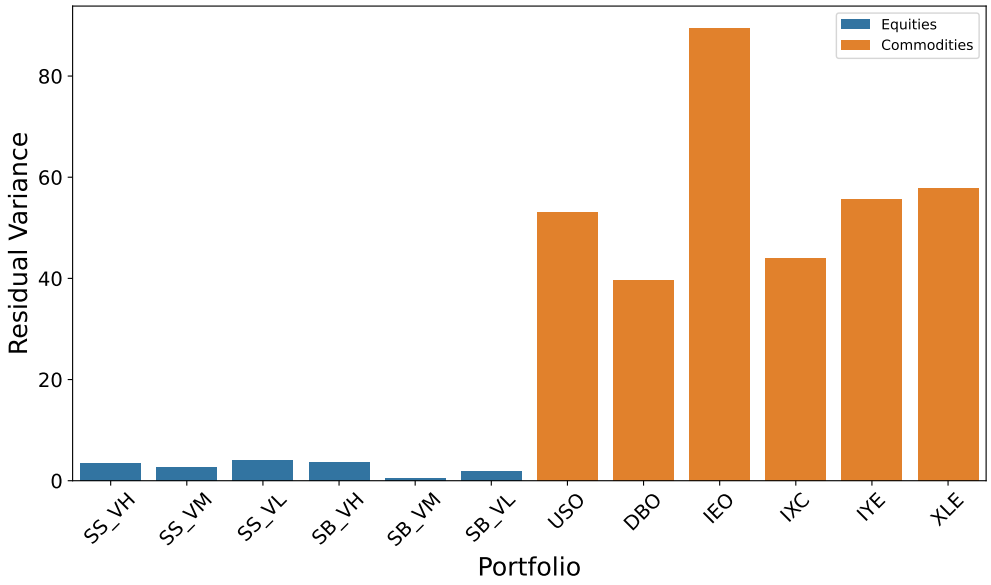

Figure 2.7: Residual Variance Across Portfolios, Equities vs ETFs

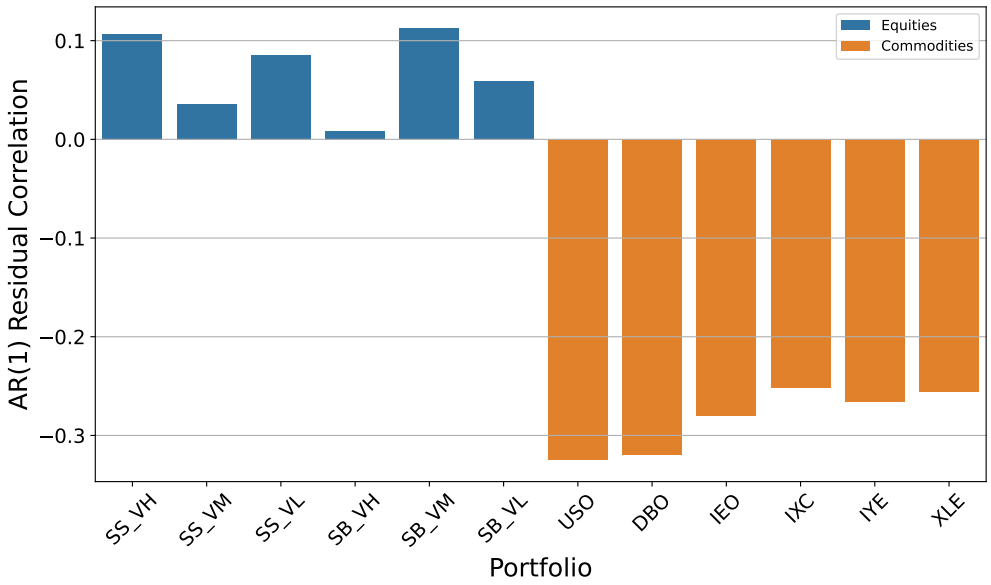

Figure 2.8: AR(1) Residual Autocorrelation Across Portfolios, Equities vs ETFs

residual structure is close to spherical (independent residuals and heteroscedasticity not present). For ETFs, AR(1) residual autocorrelation is strongly negative (from –0.25 to –0.30), which indicates mean-reversion and oscillation over time. The residual variance is very high, showing cross-sectional heteroskedasticy. OLS ignores both heteroscedasticity and correlation, resulting in underestimated standard errors. The GLS model misspecifies the residual covariance structure if autocorrelation is ignored, leading to inflated standard errors. Heteroscedasticity- and autocorrelation-consistent (HAC) adjustment is used to compensate for heteroscedasticity and

autocorrelation in both OLS and GLS methods, but it does not change the estimated coefficients.

Finally, in Table 2.22, we show the ETFs cross-sectional correlation matrix of Residuals with the Oil factor regression. In paragraph 2.6.1.2, we developed the cross-sectional correlation analysis for the global economies market factor independently.

	USO	DBO	IEO	IXC	IYE	XLE
USO	1.000	0.855	0.470	0.547	0.487	0.493
DBO	0.855	1.000	0.646	0.682	0.643	0.646
IEO	0.470	0.646	1.000	0.923	0.965	0.965
IXC	0.547	0.682	0.923	1.000	0.969	0.971
IYE	0.487	0.643	0.965	0.969	1.000	0.997
XLE	0.493	0.646	0.965	0.971	0.997	1.000

Table 2.22: ETFs Cross-Sectional Correlation Matrix of Residuals with the Oil factor

From Table 2.22, we can see that the residuals are strongly correlated, especially among the ETFs linked to the Oil firm prices: IEO, IXC, IYE, XLE. This represents a violation of OLS assumptions. GLS performs better than OLS when residuals are heteroskedastic but not autocorrelated. For commodity portfolios, the strong time-series dependence in residuals can lead OLS to misestimate the covariance structure. This explains why OLS often reports smaller p-values: it ignores serial correlation, whereas GLS explicitly account for it. For Equities, the covariance matrix is well-behaved, which helps GLS to reweight without numerical instability. We will use GLS combined with the HAC covariance estimator to address cross-sectional correlation, heteroscedasticity, and autocorrelation in the residuals.

Factor Testing

Table 2.23 reports the Harvey–Liu test, Harvey and Liu (2021), for incremental factor significance. The test compares pricing errors (intercepts of the cross-sectional regression) before/after adding a factor.

For the pair Aluminium Oil the mean reduction in pricing errors is statistically significant, that suggests the Aluminium factors adds explanatory power to the OIL ETFs. The median reduction (robust evidence) is not significant however, indicating that the effect may not be across all ETFs.

Factor Pair	Mean	p (mean)	Signif.	Median	p (median)	Signif.
AL-OIL	0.0238	0.0450	**	0.0451	0.1980	
AL-OIL_AL_ort	0.0238	0.0540		0.0451	0.2000	
GAS-OIL	0.0387	0.1990		0.0457	0.2470	
GAS-OIL_GAS_ort	0.0387	0.1820		0.0457	0.2280	
SOY-OIL	0.0813	0.1240		0.0878	0.2090	
SOY-OIL_SOY_ort	0.0813	0.1250		0.0878	0.2020	
OIL-AL	0.0085	0.0440	**	0.0154	0.2000	
OIL-AL_OIL_ort	0.0085	0.0370	**	0.0154	0.1780	
OIL-GAS	-0.0084	0.8230		-0.0295	0.8630	
OIL-GAS_OIL_ort	-0.0084	0.8510		-0.0295	0.8940	
OIL-SOY	-0.0004	0.8310		-0.0014	0.9300	
OIL-SOY_OIL_ort	-0.0004	0.8520		-0.0014	0.9450	

Table 2.23: Harvey–Liu Bootstrap Test for Factor Significance, Commodities

2.6.2.3 One-Factor Model

We refer to equation 2.7 and 2.8 for the Fama-MacBeth model and to equation 2.10 for the nonlinear model. In this section, $R_{i,t}$ is the excess percentage return of Oil ETF i at time t, f_t is the Oil factor excess percentage return at time t, $\tilde{f}_t = f_t - \mathbb{E}[f]$ is the demeaned factor excess percentage returns.

We estimate the cross-sectional asset pricing model using Fama-MacBeth (FM), SUR and GMM approaches. In Table 2.24 and 2.25, we show the different regression results for the centred factor and the raw factor, using the equivalence formula (equation 2.78) as we did for the Equities regression: the Oil factor mean is 0.4451.

Method	Parameter	Estimate	Std. Error	t-Statistic	p-value
FM	Intercept	0.9242	0.8388	1.102	0.2706
FM	Risk price	-1.6276	1.3058	-1.246	0.2126
OLS	Intercept	0.9242	0.5947	1.550	0.1220
OLS	Risk price	-1.6276	1.0927	-1.490	0.1381
GLS	Intercept	-0.1203	0.5216	-0.231	0.8179
GLS	Risk price	-0.6631	0.8416	-0.788	0.4318
SUR	Intercept	-0.0803	0.5291	-0.150	0.8795
SUR	Risk price	-0.7169	0.8479	-0.850	0.3990
GMM	Intercept	-0.1945	0.5436	-0.360	0.7210
GMM	Risk price	-0.4453	0.8890	-0.500	0.6170

Table 2.24: ETFs One-Factor Risk Price Estimates, Centred Factor

The fit of the one-factor model is good: $R^2 = 0.837$, adjusted $R^2_{adj} = 0.796$, with a statistically significant risk price for the OIL factor: $\hat{\lambda}_{\text{OIL}} = -2.07,\ p\text{-value} = 0.11$ for the centred factor with the two steps Fama-MacBeth regression, as shown in the FM item of Table 2.25.

The Oil risk is negatively priced: ETFs with higher exposure to the OIL factor (higher β_i) earn lower expected returns. This suggests that investors value the factor for its insurance or hedging benefits as compensation for the lower return.

Method	**Parameter**	**Estimate**	**Std. Error**	**t-Statistic**	***p*-value**	$\lambda_{\tilde{f}}$	***p*-value**$_{\tilde{f}}$
FM	Intercept	0.9242	0.8388	1.102	0.2706		0.2706
FM	Risk price	-2.0721	1.3058	-1.587	0.1126	-1.6276	0.2126
OLS	Intercept	0.9222	0.5947	1.550	0.1227		0.1227
OLS	Risk price	-2.0680	1.0926	-1.890	0.0600	-1.6234	0.1381
GLS	Intercept	-0.1203	0.5216	-0.231	0.8179		0.8179
GLS	Risk price	-1.1077	0.8416	-1.316	0.1898	-0.6631	0.4318
SUR	Intercept	-0.0804	0.5291	-0.150	0.8795		0.8795
SUR	Risk price	-1.1613	0.8478	-1.370	0.1725	-0.7167	0.3990
GMM	Intercept	-0.1945	0.5436	-0.360	0.7210		0.7210
GMM	Risk price	-0.8899	0.8890	-1.000	0.3182	-0.4453	0.6170

Table 2.25: ETFs One-Factor Risk Price Estimates, Raw Factor

We already noticed that the FM GLS significance is higher than the OLS one, which we have already explained with the strong autocorrelation of the ETFs OIL residuals. OLS is computed in SAS with PROC MODEL, which uses standard least squares for point estimates, but a WLS/diagonal covariance for inferences, producing a lower standard error compared to a two-step Fama–MacBeth OLS when there is substantial cross-equation heteroscedasticity/correlation.

2.6.2.4 Integration and Segmentation

In this section, we apply the CAPM integrated model, with the six ETFs, OIL as the local factor and the other commodities as global factors. Below, we show the integrated model for the GAS global factor and the OIL local factor. We use the orthogonal components, although we have already shown that there is low collinearity between the local and the global factors.

$$R_{i,t} = \lambda_0 + \beta_i^{GAS}(R_{GAS,t} + \lambda_{GAS}) + \beta_i^{OIL\perp GAS}(R_{OIL\perp GAS,t} + \lambda_{OIL}) + \eta_{i,t}, \quad \forall i,t \qquad (2.82)$$

In order to prove the complete integration hypothesis, the domestic risk prices λ_{OIL} should be equal to zero, while, for integration, the global factor λ_{GAS} should be different from zero.

The segmented model is built in a similar way and we get the following general equation:

$$R_{i,t} = \delta_0 + \zeta_i^{OIL}(R_{OIL,t} + \delta_{OIL}) + \zeta_i^{GAS \perp OIL}(R_{GAS \perp OIL,t} + \delta_{GAS}) + \nu_{i,t}, \quad \forall i, t \qquad (2.83)$$

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
AL_OIL	FM OLS	λ_{AL}	5.0244	0.2796		Estimate not statistically significant; no evidence of integration	
AL_OIL	FM OLS	λ_{OIL}	-5.2168	0.1463		Estimate is not statistically significant; no evidence of segmentation	
GAS_OIL	FM OLS	λ_{GAS}	-3.7277	0.5036		Estimate not statistically significant; no evidence of integration	
GAS_OIL	FM OLS	λ_{OIL}	-1.3138	0.3757		Estimate is not statistically significant; no evidence of segmentation	
SOY_OIL	FM OLS	λ_{SOY}	-1.4113	0.6730		Estimate not statistically significant; no evidence of integration	
SOY_OIL	FM OLS	λ_{OIL}	-1.6383	0.1928		Estimate is not statistically significant; no evidence of segmentation	
AL_OIL	FM HAC	λ_{AL}	5.0244	0.1336		Estimate not statistically significant; no evidence of integration	
AL_OIL	FM HAC	λ_{OIL}	-5.2168	0.0240	**	Estimate is statistically significant; integration rejected	IR
GAS_OIL	FM HAC	λ_{GAS}	-3.7277	0.4822		Estimate not statistically significant; no evidence of integration	
GAS_OIL	FM HAC	λ_{OIL}	-1.3138	0.4007		Estimate is not statistically significant; no evidence of segmentation	
SOY_OIL	FM HAC	λ_{SOY}	-1.4113	0.6457		Estimate not statistically significant; no evidence of integration	
SOY_OIL	FM HAC	λ_{OIL}	-1.6383	0.2038		Estimate is not statistically significant; no evidence of segmentation	
AL_OIL	GLS	λ_{AL}	1.0087	0.7433		Estimate not statistically significant; no evidence of integration	
AL_OIL	GLS	λ_{OIL}	-1.6975	0.4121		Estimate is not statistically significant; no evidence of segmentation	
GAS_OIL	GLS	λ_{GAS}	-3.5249	0.1640		Estimate not statistically significant; no evidence of integration	
GAS_OIL	GLS	λ_{OIL}	-0.5471	0.5792		Estimate is not statistically significant; no evidence of segmentation	
SOY_OIL	GLS	λ_{SOY}	1.1259	0.6695		Estimate not statistically significant; no evidence of integration	
SOY_OIL	GLS	λ_{OIL}	-1.3363	0.1841		Estimate is not statistically significant; no evidence of segmentation	
AL_OIL	GLS HAC	λ_{AL}	1.0087	0.6612		Estimate not statistically significant; no evidence of integration	
AL_OIL	GLS HAC	λ_{OIL}	-1.6975	0.2846		Estimate not statistically significant; no evidence of segmentation	
GAS_OIL	GLS HAC	λ_{GAS}	-3.5249	0.1762		Estimate not statistically significant; no evidence of integration	
GAS_OIL	GLS HAC	λ_{OIL}	-0.5471	0.5090		Estimate is not statistically significant; no evidence of segmentation	
SOY_OIL	GLS HAC	λ_{SOY}	1.1259	0.6376		Estimate not statistically significant; no evidence of integration	
SOY_OIL	GLS HAC	λ_{OIL}	-1.3363	0.1179		Estimate is not statistically significant; no evidence of segmentation	

Table 2.26: ETFs Integration Test Results, Centred Factors

In Tables 2.26 and 2.27, we show the results from the Fama-MacBeth regression with OLS, GLS and GLS with HAC methods. The estimated risk prices are not statistically significant.

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
OIL_AL	FM OLS	δ_{OIL}	-0.9369	0.4719		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	FM OLS	δ_{AL}	5.2747	0.2419		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	FM OLS	δ_{OIL}	-1.6376	0.1911		Estimate not statistically significant; no evidence of segmentation	
OIL_GAS	FM OLS	δ_{GAS}	-3.4890	0.5358		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	FM OLS	δ_{OIL}	-2.0469	0.1064		Estimate not statistically significant; no evidence of segmentation	
OIL_SOY	FM OLS	δ_{SOY}	-1.1211	0.7324		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	FM HAC	δ_{OIL}	-0.9369	0.4966		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	FM HAC	δ_{AL}	5.2747	0.0933		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	FM HAC	δ_{OIL}	-1.6376	0.2059		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	FM HAC	δ_{GAS}	-3.4890	0.5178		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	FM HAC	δ_{OIL}	-2.0469	0.0771	*	Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	FM HAC	δ_{SOY}	-1.1211	0.7120		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	GLS	δ_{OIL}	-0.8382	0.4310		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	GLS	δ_{AL}	1.2326	0.6706		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS	δ_{OIL}	-0.8533	0.3410		Estimate not statistically significant; no evidence of segmentation	
OIL_GAS	GLS	δ_{GAS}	-3.4005	0.1868		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	GLS	δ_{OIL}	-1.0103	0.2403		Estimate not statistically significant; no evidence of segmentation	
OIL_SOY	GLS	δ_{SOY}	1.2692	0.6270		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	GLS HAC	δ_{OIL}	-0.8382	0.2931		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	GLS HAC	δ_{AL}	1.2326	0.5708		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS HAC	δ_{OIL}	-0.8533	0.2259		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS HAC	δ_{GAS}	-3.4005	0.1998		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	GLS HAC	δ_{OIL}	-1.0103	0.1209		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	GLS HAC	δ_{SOY}	1.2692	0.5929		Estimate not statistically significant; no evidence of segmentation rejection	

Table 2.27: ETFs Segmentation Test Results, Centred Factors

However, under HAC inference, we reject the hypothesis of full integration between the Oil and Aluminium commodity markets. The OLS two-factor segmented model (OIL and GAS orthogonal component) has a slightly higher $R^2 = 0.915$ compared with the one-factor model ($R^2 = 0.909$, Adjusted $R^2_{adj} = 0.886$,), but a lower $R^2_{adj} = 0.859$, reflecting that introducing the global orthogonal component increases the model complexity but does not provide better explanatory power. We use the GLS asymptotic Shanken variance formulation, Shanken (1992), together with the two-pass Fama–MacBeth procedure: betas estimated from time-series, risk

prices estimated from cross correlation, with a moment style HAC variance estimator (Newey-West) manually built. GLS accounts for cross-sectional correlation, while HAC adjusts for serial correlation and heteroscedasticity. In the presence of autocorrelation or volatility clustering, GLS HAC-based and GLS standard errors can differ substantially: HAC-adjusted standard errors tend to be smaller as in our case, while the parameter estimates will always be the same.

2.6.2.5 Factor-Mimicking GLS Estimation

Mimicking portfolios are at the base of Fama–French portfolio construction, which provides more robust inference by securing factor exposures in the actual return space. Following Cochrane (2005), we estimate the risk price vector $\boldsymbol{\lambda}$ using Generalized Least Squares with factor-mimicking portfolios, which involves: constructing factor-mimicking portfolios; estimating betas from time-series regressions using the constructed factor-mimicked in the first initial step, and applying GLS on the cross-section of average returns.

The idea of projecting the factors into the asset space can improve statistical efficiency by reducing noise and mitigating multicollinearity. This projection is achieved through mimicking portfolios, as introduced by Cochrane (2005) and earlier formalized under the conditional projection theorem of Hansen and Richard (1987).

We define $\mathbf{X} = (X_{t,i}) \in \mathbb{R}^{T\times N}$ as the matrix of excess returns over $t = 1, \ldots, T$ periods for $i = 1, \ldots, N$ assets, and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{\bar{\ell}}] \in \mathbb{R}^{T\times \bar{\ell}}$, the matrix of $\bar{\ell}$ factor realizations with elements $y_{t,l}$ for $l = 1, \ldots, \bar{\ell}$.

The demeaned matrices are:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{e}_T \bar{\mathbf{X}}, \qquad \tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{e}_T \bar{\mathbf{Y}},$$

where $\mathbf{e}_T$ is a $T \times 1$ vector of ones, $\mathbf{e}_T = [1, \ldots, 1]^\top$.

The model is:

$$\mathbf{Y} = \mathbf{XW} + \mathbf{E},$$

where $\mathbf{W} \in \mathbb{R}^{N\times \bar{\ell}}$ is the weight matrix for the mimicking portfolios, and $\mathbf{E} \in \mathbb{R}^{T\times \bar{\ell}}$ is the error matrix. The mimicking problem is a joint least squares projection:

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}\in\mathbb{R}^{N\times\bar{\ell}}} \sum_{t=1}^{T}\sum_{l=1}^{\bar{\ell}} \left(\tilde{y}_{t,l} - \sum_{i=1}^{N} \tilde{X}_{t,i} W_{i,l}\right)^2. \tag{2.84}$$

subject to the normalization constraint:

$$\mathbf{e}_N^\top \mathbf{W} = \mathbf{e}_{\bar{\ell}}^\top, \tag{2.85}$$

where $\mathbf{e}_N$ is an $N \times 1$ vector of ones and $\mathbf{e}_{\bar{\ell}}$ is an $\bar{\ell} \times 1$ vector of ones.

The solution without the constraint is the OLS solution:

$$\widehat{\mathbf{W}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}}. \tag{2.86}$$

The solution with the constraint is the Lagrange multiplier solution:

$$\widehat{\mathbf{W}} = \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}} - \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \mathbf{e}_N \Big(\mathbf{e}_N^\top \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \mathbf{e}_N\Big)^{-1} \Big(\mathbf{e}_N^\top \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}} - \mathbf{e}_{\bar{\ell}}^\top\Big). \tag{2.87}$$

Each column of $\tilde{\mathbf{F}} = \tilde{\mathbf{X}}\widehat{\mathbf{W}}$, the mimicked demeaned factor vector, represents the return series of a mimicking portfolio corresponding to one of the $\bar{\ell}$ original factors:

$$\tilde{F}_{t,l} = \sum_{i=1}^{N} \tilde{X}_{t,i}\, \widehat{W}_{i,l}, \quad \forall t, l. \tag{2.88}$$

We then apply Fama-Macbeth two-step regression using the mimicked factors: $\tilde{\mathbf{F}}_t = (\tilde{F}_{t,1}, \ldots, \tilde{F}_{t,\bar{\ell}})^\top$. First, we estimate factor loadings matrix $\mathbf{B}$ via time-series regression

$$X_{t,i} = \alpha_i + \sum_{l=1}^{\bar{\ell}} \beta_{i,l} \tilde{F}_{t,l} + \varepsilon_{t,i}, \quad \forall t, i \tag{2.89}$$

where the asset exposure vector is $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,\bar{\ell}})^\top$. Stacking gives $\mathbf{B} \in \mathbb{R}^{N \times \bar{\ell}}$.

Then, we run a cross-sectional (CS) GLS regression with HAC inference, using the mimicked factor beta matrix and $\bar{\mathbf{X}} = (\bar{X}_1, \ldots, \bar{X}_N)^\top$, the $N \times 1$ vector of average returns, where $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{t,i}$:

$$\bar{X}_i = \lambda_0 + \sum_{l=1}^{\bar{\ell}} \beta_{i,l}\, \lambda_l + \eta_i, \quad \forall i \tag{2.90}$$

In matrix form:

$$\bar{\mathbf{X}} = \lambda_0 \mathbf{e} + \mathbf{B}\boldsymbol{\lambda} + \boldsymbol{\eta},$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{\bar{\ell}})^\top$ are factor risk prices. We define $\mathbf{X}_{\text{cs}} = [\mathbf{1}, \mathbf{B}]$.

The GLS estimator is:

$$\widehat{\boldsymbol{\theta}}_{\text{GLS}} = (\mathbf{X}_{\text{cs}}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_{\text{cs}})^{-1}\mathbf{X}_{\text{cs}}^{\top}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{X}}, \tag{2.91}$$

where $\boldsymbol{\Sigma}$ is the residual covariance matrix from the first-step regressions.

The HAC robust covariance is defined as in equation 2.80, replacing $\mathbf{B}$ with $\mathbf{X}_{\text{cs}}$:

$$\widehat{\text{Var}}(\widehat{\boldsymbol{\theta}}_{\text{GLS}}) = \frac{1}{T}\left(\mathbf{X}_{\text{cs}}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_{\text{cs}}\right)^{-1}\widehat{\Omega}_{\text{HAC}}\left(\mathbf{X}_{\text{cs}}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{X}_{\text{cs}}\right)^{-1}. \tag{2.92}$$

While orthogonalization techniques (e.g., Gram–Schmidt) reduce collinearity initially, the OLS-based factor-mimicking step reintroduces correlation because it projects onto the correlated return space of ETFs. Regularisation is therefore necessary to maintain numerical stability in GLS estimation.

The ARPM framework (Advanced Risk and Portfolio Management) constructs factor-mimicking portfolios using the same projection principle but introduces two key differences:

1. Returns and factors are explicitly centred prior to projection.
2. Mimicking portfolios are rescaled to achieve unit variance or to meet exposure targets.

Our implementation follows Cochrane (2005) without variance normalization.

The empirical evidence in Sakkas (2024) demonstrates that mimicking portfolios based on Principal Component Analysis (PCA) extracted latent factors explaining over three-quarters of the time-series variation in commodity returns. Our approach uses raw and orthogonalised fundamental factors, addressing multicollinearity by orthogonalizing factors via Gram–Schmidt transformations. This ensures that the second-pass GLS estimation operates on linearly independent factors, mitigating instability, although it differs from PCA in that it preserves economic interpretation of individual factors.

The Fama-MacBeth regression for the integrated and segmented models are shown below. The excess percentage return now refers to the factor mimicking with weight constraint; we discard the unconstrained model, as it tends to produce unrealistically large beta estimates.

As expected, the results in Tables 2.28 and 2.29 show that the mimicking portfolio factor regressions compared with factor regressions, exhibit lower residual variance, and stronger pricing of key factors under GLS and GLS with HAC.

The integration test results show that the Gas factor is priced with Oil ETFs under both models, total integration is detected. The Gas orthogonal component is priced in the segmentation model,

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
AL_OIL	GLS	λ_{AL}	+0.5523	0.8512		Estimate not statistically significant; no evidence of integration	
AL_OIL	GLS	λ_{OIL}	+0.0915	0.9752		Estimate not statistically significant; no evidence of segmentation	
GAS_OIL	GLS	λ_{GAS}	-2.1404	0.0049	**	Statistically significant and economically large; integration detected	TI
GAS_OIL	GLS	λ_{OIL}	-1.2169	0.1096		Estimate not statistically significant; no evidence of segmentation	
SOY_OIL	GLS	λ_{SOY}	+0.4325	0.8284		Estimate not statistically significant; no evidence of integration	
SOY_OIL	GLS	λ_{OIL}	+0.1556	0.9379		Estimate not statistically significant; no evidence of segmentation	
AL_OIL	GLS HAC	λ_{AL}	+0.5523	0.8149		Estimate not statistically significant; no evidence of integration	
AL_OIL	GLS HAC	λ_{OIL}	+0.0915	0.9691		Estimate not statistically significant; no evidence of segmentation	
GAS_OIL	GLS HAC	λ_{GAS}	-2.1404	0.0014	***	Statistically significant and economically large; integration detected	TI
GAS_OIL	GLS HAC	λ_{OIL}	-1.2169	0.0691	*	Estimate not statistically significant; no evidence of segmentation	
SOY_OIL	GLS HAC	λ_{SOY}	+0.4325	0.8074		Estimate not statistically significant; no evidence of integration	
SOY_OIL	GLS HAC	λ_{OIL}	+0.1556	0.9301		Estimate not statistically significant; no evidence of segmentation	

Table 2.28: ETFs Mimicking Factor Integration Test Results

which confirms rejection of segmentation.

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
OIL_AL	GLS	δ_{OIL}	+1.0557	0.8301		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	GLS	δ_{AL}	+1.8277	0.7103		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS	δ_{OIL}	-1.4197	0.0782	*	Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS	δ_{GAS}	-2.1071	0.0090	**	Estimate statistically significant; evidence of segmentation rejection	SR
OIL_SOY	GLS	δ_{OIL}	+0.3549	0.8789		Estimate not statistically significant; no evidence of segmentation	
OIL_SOY	GLS	δ_{SOY}	+0.8333	0.7205		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	GLS HAC	δ_{OIL}	+1.0557	0.7811		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	GLS HAC	δ_{AL}	+1.8277	0.6304		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	GLS HAC	δ_{OIL}	-1.4197	0.0523	*	Estimate statistically significant; weak evidence of segmentation	
OIL_GAS	GLS HAC	δ_{GAS}	-2.1071	0.0040	**	Estimate statistically significant; evidence of segmentation rejection	SR
OIL_SOY	GLS HAC	δ_{OIL}	+0.3549	0.8643		Estimate not statistically significant; no evidence of segmentation	
OIL_SOY	GLS HAC	δ_{SOY}	+0.8333	0.6882		Estimate not statistically significant; no evidence of segmentation rejection	

Table 2.29: ETFs Mimicking Factor Segmentation Test Results

For the integration model, the estimated price of risk of the mimicked Gas factor in the pair with Oil Gas orthogonal components is -2.14%, with a highly statistical significant p-value. The Gas Oil orthogonal component in the segmentation model has a similar risk penalty, being as well statistically significant. This implies that an OIL ETF with unit exposure to the mimicked GAS factor earns, on average, 2.14% lower monthly excess return. Annualized, this corresponds to a risk penalty of approximately: $2.14 \times 12 = 25.7\%$ per year. The large negative price of risk suggests that Gas behaves as an hedging instrument: investors accept lower returns for exposure to it, possibly due to its diversification benefit.

In the Appendix 4.2, we present an alternative approach using synthetic rolling yields (RYS), motivated by the contango and backwardation dynamics of futures-based ETFs such as USO and DBO. However, this method does not conform to the SDF framework because the RYS is a signal constructed from returns rather than a tradable payoff, so using it as the dependent variable of the regression violates the fundamental asset pricing restriction that expected returns should be explained by factor loadings on tradable portfolios.

2.7 Contribution

In the classic factor model analysis, we provide robust evidence of equivalence of the centred and raw factor regression under Ordinary Least Squares and Generalized Least Squares. We noted that, as expected, Fama-MacBeth GLS regression, results in estimates of the risk price that are equivalent to the SUR and GMM methods.

In the linear approximation section, we introduce several convex approximation and linear approximation methods, which to our knowledge, are applied for first time to beta-pricing modelling. In the related results section, we show how the estimates and p -value are consistent with the results obtained with more classic estimation techniques.

Our results suggest that the L1-norm methods, proposed in this chapter, are more appropriate for risk price estimation under heavy-tails, as observed in the case of the EU market factor and its associated residual distributions.

Although the standard methods applied in the global economies integration research, have been already introduced in the literature, this is the first time that they are applied systematically to compare the integration and segmentation between different economies and a given portfolio set. This systematic approach helps to establish the conclusiveness of their forecasts and corroborates the validity of the CAPM integrated model, as shown in Penco and Lucas (2024) article.

Finally, we extend the integration model to commodity markets. To better capture the cross-sectional pricing of commodity risk, we use a factor mimicking approach, that, as expected, outperforms standard factors in explaining cross-sectional commodity risk. The integration and segmentation model between Oil (as a local factor) and other commodities (Aluminium, Gas, and Soybean, modelled as global factors), lead to lower residual variance, more stable and interpretable factor loadings, and stronger pricing of key factors under both GLS and GLS HAC methods. The results show total integration of the Oil ETFs with Gas, which makes economically sense. However, other commodities like Soybean — despite the fact that circa 5% of global Soybean production is used to produce biofuel — do not exhibit signs of integration with Oil.

Chapter 3

Dependence, Copula, Regularisation

This chapter investigates the risk price estimation and the integration and segmentation analysis of global equity markets and commodity indices with alternative copula-based dependence methods.

We have the following objectives:

- Examine whether alternative methods provide consistent outcomes.
- Confirm the added value of the beta-pricing integration and segmentation models.
- Show that copula dependence can assist in the integration and segmentation analysis.

One of the main implementations of copula is the aggregation of the single risk drivers $X_{1,t}, \ldots, X_{\bar{\ell},t}$ dynamics into a joint multivariate model for the whole risk drivers process $\mathbf{X}_t \equiv (X_{1,t}, \ldots, X_{\bar{\ell},t})$. For each set of data, global economies and commodities, we apply a copula marginal model, fitting elliptical copulas, for the joint integration analysis with the related portfolios and ETFs. Via examples, we show how the copula density together with the maximum likelihood estimator can successfully reproduce the exposure (betas) and risk price (lambda) of the moments based standard methods (GMM, FGLS).

In a second experiment, we generate two different simulations: one scenario where both the global and the local orthogonal factors are priced and another where only the global factor is priced. We then apply our integration and segmentation analysis. The results show that correlation alone does not imply pricing. It is the cross-sectional covariance with the expected returns that determines whether a factor is priced. To better visualise this result we define two measures: the sum of the factors copula with the returns and the stochastic discount factor copula with the returns. We apply these measures to our simulation data and to the historical experimental data. The results illustrate the difference between exposure and pricing:

- The sum of the factors and returns copula measures the extent to which factors explain the time-series variability of returns.
- The copula of the SDF and returns captures how much payoffs in different states are

discounted, which depends directly on the risk price λ_l.

This distinction is central in asset pricing: factors can generate return variation and co-movement even if they do not earn a risk price.

For the definitions and methods used in this chapter we refer the reader to the Advance Risk and Portfolio Management material, ARPM (2025), which is available online for subscribers.

3.1 Literature Review

Copulas have been widely applied in finance to model the dependence among asset returns, particularly for risk management and tail dependence. A classic example is Embrechts et al. (2002), who discuss copulas in the context of credit risk and value-at-risk applications. For the integration experiment, our main reference is Patton (2006), who develops asymmetric copula models for exchange rate dependence, illustrating how flexible dependence structures can improve modelling of joint returns.

However, in most of this literature, the focus is on modelling the joint distribution of raw returns, rather than residuals from a factor model. Copulas are typically used to describe co-movements in levels or volatilities, not to estimate risk prices.

A parallel literature develops copula-GARCH models, in which each marginal series follows a GARCH process and the copula captures the evolving dependence. Notable examples include Jondeau and Rockinger (2006), who estimate copula-GARCH models for international stock indices, and Patton (2013), who reviews copula models for economic time series. These approaches are primarily designed to capture dynamic tail dependence for forecasting and risk measurement.

While such models embed both heteroscedasticity and time-varying dependence, they are not typically formulated as estimators of factor model parameters or as alternatives to GMM or Fama-MacBeth estimation, which is the goal of our first experiment using the one-factor model.

Shanken and Roll (1985) and Cochrane (2005) discuss maximum likelihood estimation of linear factor models under multivariate normality. In that setting, the full covariance matrix of residuals plays a role analogous to the copula correlation matrix in our approach. Under the assumption of multivariate Gaussian residuals, MLE and GLS estimators coincide, and efficient estimation is possible without GMM.

This framework is closest in spirit to the Gaussian copula likelihood presented here, as both rely on specifying a parametric dependence structure. However, even in this literature, estimation

typically proceeds by specifying a multivariate normal model, rather than expressing the likelihood in copula form.

We have already seen that the standard two-pass Fama-MacBeth estimator, Fama and MacBeth (1973), and GMM approaches, Hansen (1982), remain the dominant methodologies in empirical asset pricing. These methods rely on moment conditions rather than specifying the full likelihood. In particular:

- Fama-MacBeth OLS treats residuals as independent across assets and over time.
- Fama-MacBeth with GLS weights the cross-sectional regressions by the estimated covariance matrix of residuals.
- GMM estimates both risk prices and standard errors by consistently estimating the covariance of the moment conditions, including serial and cross-sectional dependence.

The copula likelihood approach assumes a parametric dependence structure rather than estimating the covariance nonparametrically as is done in GMM with optimal weighting. For this reason, if the copula is misspecified, standard errors and inference may be invalid.

Most applications of copulas in finance address dependence among returns themselves, not the estimation of factor risk prices. The methodology documented in this study contributes to the literature by demonstrating how copula density likelihood estimation can serve as a parametric analogue to GMM or FGLS methods, and by providing a unified framework to compare these approaches empirically.

3.2 Copula

3.2.1 Copula Density

This section is derived from the corresponding unit of the Advance Risk and Portfolio Management, ARPM (2025) course.

Copula extends linear correlation to capture more general forms of dependence across variables, including nonlinear relationships, tail co-movements, and higher-moment features. Traditional covariance-based measures cannot fully characterize such dependencies. Instead, copulas provide a rigorous framework to separate marginal distributions from the joint dependence structure. For comprehensive treatments, see Joe (1997) and Rayens and Nelsen (2000).

Any random vector $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ with continuous marginal cumulative distribution functions F_{X_i} can be transformed via the probability integral transform:

$$U_i = F_{X_i}(X_i), \qquad \text{for } i = 1, \ldots, N. \tag{3.1}$$

The transformation is applied separately to each marginal, producing a vector $\boldsymbol{U}$ whose components are uniform on $[0, 1]$. However, while each marginal U_n is uniform, the vector $\boldsymbol{U}$ as a whole is generally not uniformly distributed on the unit hypercube. Its joint distribution, called the *copula*, encodes all dependence among the components. For each i, let $u_i \in [0, 1]$ be the uniform probabilities, a real number representing the argument of the copula cdf ($C_{\boldsymbol{X}}$). Then:

$$C_{\boldsymbol{X}}(u_1, \ldots, u_N) = \mathbb{P}\big(U_1 \leq u_1, \ \ldots, \ U_N \leq u_N\big). \tag{3.2}$$

This defines a valid multivariate cumulative distribution function with uniform marginals.

From Sklar's theorem, Sklar (1959), the original joint cumulative distribution function $F_{\boldsymbol{X}}$ can be uniquely decomposed (if the marginals are continuous) as:

$$F_{\boldsymbol{X}}(x_1, \ldots, x_N) = C_{\boldsymbol{X}}\big(F_{X_1}(x_1), \ \ldots, \ F_{X_N}(x_N)\big). \tag{3.3}$$

We use uppercase letters such as X_i to denote random variables, and lowercase letters x_i to denote realizations or arguments at which the functions are evaluated. Here:

- $F_{\boldsymbol{X}}$ denotes the joint cumulative distribution function of the vector $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$.
- F_{X_i} denotes the marginal cumulative distribution function of X_i.
- $C_{\boldsymbol{X}}$ denotes the copula function, i.e., the multivariate CDF on $[0, 1]^N$ capturing the dependence structure.

If the joint distribution is absolutely continuous with a density $f_{\boldsymbol{X}}$ and marginal densities f_{X_i}, the decomposition can also be expressed in the density form:

$$f_{\boldsymbol{X}}(x_1, \ldots, x_N) = c_{\boldsymbol{X}}\big(F_{X_1}(x_1), \ \ldots, \ F_{X_N}(x_N)\big) \prod_{i=1}^{N} f_{X_i}(x_i), \tag{3.4}$$

where:

- $f_{\boldsymbol{X}}$ denotes the joint probability density function of $\boldsymbol{X}$.
- f_{X_i} denotes the marginal density function of X_i.

- $c_{\boldsymbol{X}}$ denotes the copula density, defined as:

$$c_{\boldsymbol{X}}(u_1, \ldots, u_N) = \frac{\partial^N}{\partial u_1 \cdots \partial u_N} C_{\boldsymbol{X}}(u_1, \ldots, u_N).$$

Given any copula C and any set of marginals, formula 3.3 reconstructs a valid joint distribution. This separation and combination of the copula and marginals enables a two-stage modelling approach:

1. Estimate the marginal distributions, for example via empirical cumulative distribution functions (ECDFs).
2. Estimate the copula that captures dependence beyond marginal behaviour.

Using the empirical copula provides a nonparametric way to retain rank dependence and tail association observed in the data. Practically, in our simulation, this allows resampling residuals or in general generating observations with preserved marginal properties (including higher moments) and flexible dependence structure, rather than assuming linear Gaussian correlation.

3.2.2 Elliptical Copula Density Likelihood

We consider a one-factor model according to equation 2.10:

$$R_{i,t} = \lambda_0 + \lambda_f \beta_i + \beta_i \tilde{f}_t + \varepsilon_{i,t}, \quad \forall i, t$$

where: $R_{i,t}$ is the return of portfolio i at time t; $\tilde{f}_t$ is the demeaned factor; λ_0, λ_f are the intercept and risk price; β_i is the exposure of portfolio i and $\varepsilon_{i,t}$ are residuals.

We present a unified framework to estimate the risk price using elliptical copulas, which generalizes the Gaussian copula and extends naturally to the Student-t copula and other elliptical distributions.

The residuals are:

$$\varepsilon_{i,t} = R_{i,t} - (\lambda_0 + \lambda_f \beta_i + \beta_i \tilde{f}_t), \quad \forall i, t$$

The risk drivers are defined as the residual random variables E_i, with observed time series

$$(E_{i,1}, \ldots, E_{i,T})^\top = \begin{bmatrix} \varepsilon_{i,1} & \varepsilon_{i,2} & \cdots & \varepsilon_{i,T} \end{bmatrix}^\top \in \mathbb{R}^T, \quad \text{for all } i,$$

where $\varepsilon_{i,t}$ denotes the realization of E_i at time t. Then:

$$X_i \equiv E_i, \quad \text{for all } i$$

Applying the marginal cumulative distribution function:

$$F_{X_i}(x) = \mathbb{P}\big(X_i \leq x\big), \quad \forall i$$

we compute the probability integral transform (PIT):

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

Depending on the application, F_{X_i} can be a parametric CDF (e.g., Gaussian or Student-t) or an Empirical Cumulative Distribution Function, ECDF, see the example below for the Gaussian case:

$$u_{i,t}(\text{Gaussian}) \;=\; \Phi\Big(\frac{\varepsilon_{i,t}}{\sigma_i}\Big), \quad \text{where} \quad \sigma_i = \sqrt{\Sigma_{ii}}, \quad \text{and} \quad \Sigma_{ii} = \mathsf{Var}(\varepsilon_{i,t})$$

$$u_{i,t}^{(\text{empirical})} \;=\; \hat{F}_{X_i}\big(\varepsilon_{i,t}\big), \quad \text{where} \quad \hat{F}_{X_i}(x) = \frac{1}{T}\sum_{s=1}^{T} \mathbf{1}\big\{\varepsilon_{i,s} \leq x\big\}.$$

Here, $\mathbf{1}\{\cdot\}$ denotes the indicator function:

$$\mathbf{1}\big\{\varepsilon_{i,s} \leq x\big\} = \begin{cases} 1, & \text{if } \varepsilon_{i,s} \leq x, \\ 0, & \text{otherwise.} \end{cases}$$

The empirical PIT is defined via ranks, which is equivalent to computing the empirical grade of the residuals[1], defined via their ranks:

$$u_{i,t} = \frac{\mathsf{rank}\big(\varepsilon_{i,t} \text{ among } \{\varepsilon_{i,s}\}_{s=1}^{T}\big)}{T}, \quad \forall i, t$$

where the rank is defined as the ordinal position in the sorted list of residuals performed separately for each residual series i (i.e., the smallest residual receives rank 1, the largest receives rank T)[2].

[1] Using a more rigorous formalism, we compute the empirical grade (probability integral transform, PIT) of the residuals:

$$u_{i,t} = \hat{F}_{E_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

where $\hat{F}_{E_i}(x)$ denotes the empirical cumulative distribution function of the residuals $\{\varepsilon_{i,s}\}_{s=1}^{T}$, which is the entire sample used to estimate $\hat{F}_{E_i}$.

[2] In the implementation, we use the finite sample rank transformation $u_{i,t} = \frac{\mathsf{rank}(\varepsilon_{i,t})-0.5}{T+1}$, which avoids values

In the parametric approach, we compute the distribution grade of the residuals via the univariate marginal CDF:

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

where F_{X_i} denotes the CDF of the univariate marginal distribution.

For any elliptical copula characterized by:

- a univariate standardised CDF $G(\cdot)$,
- a univariate standardised density $g(\cdot)$,
- a multivariate standardised density $g_N(\cdot; 0, \mathbf{P})$ with correlation matrix $\mathbf{P}$,

the copula density is[3]:

$$c_{\boldsymbol{X}}(u_1, \ldots, u_N) = \frac{g_N\big(G^{-1}(u_1), \ldots, G^{-1}(u_N);\ 0, \mathbf{P}\big)}{\prod_{i=1}^{N} g\big(G^{-1}(u_i)\big)}. \tag{3.5}$$

The copula correlation matrix $\mathbf{P}$ is estimated by first transforming the PITs:

$$u_{i,t} = \hat{F}_{E_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

into latent variables whose marginal distributions match the distribution, for example:

$$q_{i,t} = \begin{cases} \Phi^{-1}\big(u_{i,t}\big), & \text{for the Gaussian copula,} \\ t^{-1}\big(u_{i,t};\, \nu\big), & \text{for the } t \text{ copula with } \nu \text{ degrees of freedom,} \end{cases}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse standard normal CDF and $t^{-1}(\cdot; \nu)$ denotes the inverse CDF of the univariate t distribution with ν degrees of freedom.

The estimated correlation matrix is:

$$\widehat{\mathbf{P}} = \mathsf{Corr}\big(q_{1,t},\ \ldots,\ q_{N,t}\big)$$

which fully determines the dependence structure of the copula, controls the implied tail dependence and rank correlation among the margins via rank-based measures.

exactly equal to 0 or 1 in the probability integral transform.

[3] The result can be derived by means of the Sklar's theorem 3.4

Then the log-likelihood is:

$$\ln \mathcal{L}(\theta) = \sum_{t=1}^{T} \Big[\ln c\big(u_{1,t}, \ldots, u_{N,t}\big) \; + \; \sum_{i=1}^{N} \ln f_{X_i}\big(\varepsilon_{i,t}\big) \Big].$$

In Appendix N we derive the closed form of the Gaussian and Student-t copula density: the methods described above are named Gauss Multi (abbreviation of multivariate) and Student-t Multi in the application results.

We refer to Appendix N.3 for the details of the optimisation steps and the definition of the parameter vector $\boldsymbol{\theta}$ below:

$$\boldsymbol{\theta} = \big(\lambda_0, \, \lambda_f, \, \beta_1, \ldots, \beta_N, \, \mathsf{vech}(\mathbf{L}), \, \nu \big),$$

where ν are the degrees of freedom, and $\mathbf{L}$ is the lower triangular Cholesky factor of a covariance matrix $\mathbf{S}$ (i.e. $\mathbf{S} = \mathbf{L}\mathbf{L}^{\top}$). The vech operator, obtained by stacking in a single column vector the elements of the lower triangular matrix. $\mathbf{L}$, is defined in Appendix M.

In the application results, we estimate the same model assuming that the residuals are independent across portfolios and over time. This corresponds to specifying the independent copula:

$$C^{\perp\!\!\!\perp}\big(u_1, \ldots, u_N\big) = \prod_{i=1}^{N} u_i,$$

whose copula density is identically equal to 1:

$$c^{\perp\!\!\!\perp}\big(u_1, \ldots, u_N\big) = 1.$$

Taking logarithms, we find:

$$\ln c^{\perp\!\!\!\perp}(u) = 0.$$

Therefore, the total log-likelihood simplifies to summing the marginal log-densities:

$$\mathcal{L}_{\mathsf{indep}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \ln f_{X_i}\big(\varepsilon_{i,t}\big),$$

where f_{X_i} denotes the marginal density of the residuals.

In the example below, we assume standard Gaussian marginals:

$$f_{X_i}(\varepsilon_{i,t}) = \phi\big(\varepsilon_{i,t}\big), \quad \forall i,t$$

where ϕ denotes the standard normal density, we have:

$$\ln \phi(\varepsilon_{i,t}) = -\frac{1}{2}\varepsilon_{i,t}^2 \; - \; \frac{1}{2}\ln(2\pi), \quad \forall i,t$$

Thus:

$$\mathcal{L}_{\text{indep}}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{i,t}^2 \; + \; \text{constant},$$

where the constant is irrelevant for optimisation.

Maximizing this likelihood is equivalent to minimizing the sum of squared residuals:

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{T}\sum_{i=1}^{N}\Big[R_{i,t} \; - \; \lambda_0 \; - \; \lambda_f\,\beta_i \; - \; \beta_i\,\tilde{f}_t\Big]^2.$$

This is exactly the objective function minimized in FM OLS. These methods are named Gauss Uni (abbreviation of univariate) and Student-t Uni in the application results.

3.2.3 Stochastic Discount Factor and Copula

For a two-factor model, using equation 2.51, the return $R_{i,t}$ of asset i at time t is specified as

$$R_{i,t} = \beta_{i,1}\left(\tilde{f}_{1,t} + \lambda_1\right) \; + \; \beta_{i,2}\left(\tilde{f}_{2,t} + \lambda_2\right) + \varepsilon_{i,t}, \quad \forall i,t$$

The exposures to systematic factors determine expected returns through their associated prices of risk λ_l:

$$\mathbb{E}\big[R_i\big] = \beta_{i,1}\,\lambda_1 \; + \; \beta_{i,2}\,\lambda_2 + \eta_i, \quad \forall i$$

The stochastic discount factor[4] consistent with the linear factor pricing model above is derived

[4] According to Cochrane (2005), the stochastic discount factor is a state-dependent weighting function that prices assets by adjusting future payoffs for both time and risk, and assigns higher value to payoffs in bad times, when marginal utility is high and consumption (C_t) is low:

$$m_{t+1} = \gamma \cdot \frac{u'(C_{t+1})}{u'(C_t)}$$

where $u(\cdot)$ is a utility function and $\gamma \in (0,1)$ is the time discount factor.

in Cochrane (2005):

$$m_t = 1 \; - \; \lambda_1 \, \tilde{f}_{1,t} \; - \; \lambda_2 \, \tilde{f}_{2,t}, \quad \forall t \tag{3.6}$$

The fundamental pricing equation, no-arbitrage pricing condition, states:

$$\mathbb{E}\big[\, m \, R_i\big] = 0, \quad \forall i$$

The decomposition below shows how the mean return is offset by the discounted co-movement with the factors, scaled by their prices of risk:

$$\mathbb{E}\big[m \, R_i\big] = \mathbb{E}\big[R_i\big] \; - \; \lambda_1 \, \mathbb{E}\big[\tilde{f}_1 \, R_i\big] \; - \; \lambda_2 \, \mathbb{E}\big[\tilde{f}_2 \, R_i\big], \quad \forall i$$

If $\lambda_l = 0$, factor l is unpriced: it contributes to variability but does not affect expected returns.

For our two-factor model, we use orthogonalised factors ($\mathsf{Cov}(\tilde{f}_1, \tilde{f}_2) = 0$):

$$\mathsf{Var}(R_i) = \beta_{i,1}^2 \, \mathsf{Var}(\tilde{f}_1) \; + \; \beta_{i,2}^2 \, \mathsf{Var}(\tilde{f}_2) \; + \; \mathsf{Var}(\varepsilon_i), \quad \forall i$$

The factor loadings $\beta_{i,l}$ explain return variation and co-movement of returns, even though the factor does not contribute to expected returns when $\lambda_l = 0$.

The price of risk λ_l determines the compensation investors require for bearing exposure to factor l. The sign indicates whether investors view the factor as desirable or undesirable risk.

The dependence structure between the SDF and returns is analysed using copulas:

1. The SDF series is computed according to equation 3.6
2. The returns $R_{i,t}$ are our portfolio data (simulated or historical)
3. Both series are transformed to uniform ranks:

$$u_{m,t} = \frac{\mathsf{rank}(m_t)}{T+1}, \quad u_{R,t} = \frac{\mathsf{rank}(R_{i,t})}{T+1}.$$

4. A Gaussian copula is fitted to $(u_{m,t}, u_{R,t})$.

The estimated copula correlation provides a description of dependence beyond linear correlation. Scenarios where both factors are priced exhibit stronger dependence between the SDF and returns. In scenarios where only one-factor is priced, dependence is weaker but still present because factor exposures continue to drive return variability.

3.3 Regularisation

Initially, we apply regularisation techniques to estimate loadings (exposure) in a multifactor regression where the observable inputs are regional market factor returns and the observable outputs are regional portfolio returns. We use the following regularisation methods:

- Lasso (L1), Tibshirani (2011)
- Ridge (L2), Hoerl and Kennard (1970)
- Elastic Net, Zou and Hastie (2005)

Starting from a model in which all factors are considered relevant for all portfolios, we identify and discard spurious factors.

- $R_{i,t}$: excess return of portfolio $i = 1, \ldots, N$ at time $t = 1, \ldots, T$,
- $\tilde{f}_{l,t} = f_{l,t} - \bar{f}_l$: centred value of factor $l = 1, \ldots, \bar{\ell}$, with $\bar{f}_l = \frac{1}{T}\sum_{t=1}^{T} f_{l,t}$,
- $\beta_{i,l}$: loading of factor l on portfolio i,
- $\varepsilon_{i,t}$: residual term.

The regression model is:

$$R_{i,t} = \sum_{l=1}^{\bar{\ell}} \beta_{i,l}\tilde{f}_{l,t} + \varepsilon_{i,t}, \quad \forall i, t \tag{3.7}$$

Regularisation imposes constraints on the regression coefficients:

- Lasso (L1 penalty) encourages factor selection (sparsity) by setting some loadings exactly to zero. The objective function combines the sum of squared residuals (SSR), which decreases as more factors are included, and an L1 penalty that grows linearly with the absolute value of the coefficients: a factor is retained only if the SSR gain outweighs the additional penalty. The Lasso penalty $\rho|\beta|$ has a constant gradient $(\pm\rho)$ near zero, that forces small coefficients toward zero:

$$\min_{\{\beta_{i,l}\}_{l=1}^{L}} \sum_{t=1}^{T} \left(R_{i,t} - \sum_{l=1}^{\bar{\ell}} \beta_{i,l}\tilde{f}_{l,t} \right)^2 + \rho \sum_{l=1}^{\bar{\ell}} |\beta_{i,l}| \tag{3.8}$$

- Ridge (L2 penalty) shrinks coefficients continuously but retains all factors, which helps when predictors are correlated and OLS estimates of betas become unstable. In Ridge regression,

the L2 penalty $\rho\beta^2$ has a gradient of $2\rho\beta$, which approaches zero as β approaches zero. As a result, the shrinkage weakens near zero, the L2 penalty avoids large coefficients and shrinks them smoothly towards zero: the solution is still dense (non-sparse), but less variable:

$$\min_{\{\beta_{i,l}\}_{l=1}^{L}} \sum_{t=1}^{T} \left(R_{i,t} - \sum_{l=1}^{\bar{\ell}} \beta_{i,l} \tilde{f}_{l,t} \right)^2 + \rho \sum_{l=1}^{\bar{\ell}} \beta_{i,l}^2 \tag{3.9}$$

- Elastic Net combines L1 and L2, and is known for enabling group selection (of related factors) and stability in high-dimensional settings thanks to the L2 component:

$$\min_{\{\beta_{i,l}\}_{l=1}^{L}} \sum_{t=1}^{T} \left(R_{i,t} - \sum_{l=1}^{\bar{\ell}} \beta_{i,l} \tilde{f}_{l,t} \right)^2 + \rho \left[\alpha \sum_{l=1}^{\bar{\ell}} |\beta_{i,l}| + (1-\alpha) \sum_{l=1}^{\bar{\ell}} \beta_{i,l}^2 \right] \tag{3.10}$$

The penalty parameter $\rho > 0$ controls the overall strength of the regularisation, while $\alpha \in [0, 1]$ controls the mix between Lasso, $\alpha = 1$, and Ridge, $\alpha = 0$. In all the models, as the penalty parameter increases, more coefficients are driven toward zero, the model becomes simpler but typically the in-sample R^2 is reduced.

While Lasso, Ridge, and Elastic Net are powerful tools for identifying statistically significant factor exposures, the simple OLS multifactor regression, presents two fundamental limitations, see Harvey and Bekaert (1994), Bruner et al. (2008):

- Collinearity. Global market factors tend to be highly correlated: the U.S. and European market returns co-move significantly. A straight regularisation OLS multifactor model is not identifying additional explanatory power of each factor. Ridge regularisation mitigates the instability through shrinkage, and Lasso may perform variable selection, but neither fully clarifies the marginal value of one-factor over another. When two factors are highly collinear, even though both are informative, OLS chooses a combination that best fits the data, and this often favours the factor with the slightly better raw correlation with $R_{i,t}$, which makes it difficult to assess phenomena such as regional (marginal) integration.

- No Pricing of Risk. The model focuses exclusively on exposures, the estimated loadings $\beta_{i,j}$, without testing whether the corresponding factors are priced. It implicitly assumes that all factors carry a nonzero risk price, which is a limitation: a factor might explain return variation (i.e., have high exposure) but still be unpriced in equilibrium. What matters for pricing is not the exposure $\beta_{i,j}$ alone, but the product $\beta_{i,j}\lambda_j$, which determines the factor's contribution to the SDF.

To resolve these issues, in the first chapter we constructed two complementary models based on orthogonalised factors and cross-sectional risk pricing:

- The integration model, in which local portfolios are priced using global factors orthogonalised with respect to the dominant (e.g., U.S.) market factor, see equation 2.66.
- The segmentation model, in which we test whether local returns are explained mainly by local factors, even after accounting for orthogonalised global influences, equation 2.67.

In this section, we extend the two-factor integrated model to multifactor, we estimate risk prices λ_l and, in addition, apply regularisation not to exposures but to the cross-sectional pricing model, incorporating p-value-based threshold. The procedure builds on the approach introduced by Bryzgalova (2015), who proposed Lasso penalised beta estimation of risk prices in a two-step cross-sectional asset pricing model to identify spurious factors. We extend that framework to incorporate both economic and statistical criteria in the pricing of risk factors, and regularise the FM second step using the Lasso penalty based on beta for the risk prices together with a soft constraint to shrink factors with high p-values, encouraging factor selection (sparsity) both in economic relevance and statistical significance.

Let $R_{i,t}$ denote the excess return of portfolio $i \in \{1, \ldots, N\}$ at time $t \in \{1, \ldots, T\}$, and let $f_{l,t}$ be the return of factor $l \in \{1, \ldots, \bar{\ell}\}$. In the first step, we estimate exposures $\beta_{i,l}$ from time-series regressions:

$$R_{i,t} = \sum_{j=1}^{\bar{\ell}} \beta_{i,l} f_{l,t} + \varepsilon_{i,t}, \quad \forall i, t \tag{3.11}$$

We orthogonalise global components with respect to local markets, enabling identification of marginal explanatory power (e.g., Asia or US factors orthogonal to the EU market).

In the second step of Fama-MacBeth, we subtract $\beta_i^\top \bar{f}$ from $\bar{R}_i$ and obtain the demeaned factor return $\tilde{R}_i$, ensuring the estimation targets the risk prices λ rather than factor mean effects: see Table 2.13, where the same method is implemented for one-factor. Finally, for each factor l we estimate its risk prices λ_l via a penalised cross-sectional GLS regression, minimizing the following objective function:

$$\begin{aligned} \min_{\lambda_0, \boldsymbol{\lambda}} \quad & \left(\tilde{\boldsymbol{R}} - \lambda_0 \mathbf{1} - \mathbf{B}\boldsymbol{\lambda}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\tilde{\boldsymbol{R}} - \lambda_0 \mathbf{1} - \mathbf{B}\boldsymbol{\lambda}\right) \\ & + \rho \sum_{l=1}^{\bar{\ell}} w_l \left|\lambda_l\right| + \rho_{\text{pval}} \sum_{l=1}^{\bar{\ell}} \mathsf{max}(0, p_l - \tau) \left|\lambda_l\right|, \end{aligned} \tag{3.12}$$

where:

- $\tilde{\boldsymbol{R}} \in \mathbb{R}^N$ is the vector of average portfolio excess returns adjusted for time-series fit, i.e., $\tilde{R}_i = \bar{R}_i - \beta_i^\top \bar{f}$.

- $\mathbf{B} \in \mathbb{R}^{N \times \bar{\ell}}$ is the matrix of time-series betas.

- $\boldsymbol{\Sigma}$ is the cross-sectional residual covariance matrix estimated from first-step residuals.

- $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{\bar{\ell}})^\top$ are the risk prices to be estimated.

- $w_l = 1/|\beta_{\cdot,l}|^d$ are adaptive weights, based on the L1-norm across portfolios. $d > 0$ is a tuning parameter in the adaptive weights; we use $w_l = (\|\widehat{\boldsymbol{\beta}}_{\cdot,l}\|_1 + \varepsilon)^{-d}$, where $\|\widehat{\boldsymbol{\beta}}_{\cdot,l}\|_1 = \sum_{i=1}^N |\widehat{\beta}_{i,l}|$ (L1 across portfolios) and $\varepsilon > 0$ avoids division by zero. Larger d penalizes weak loadings more strongly (more sparsity): $d = 1$ corresponds to adaptive Lasso, and $d = 0$ to standard Lasso.

- ρ is the Lasso regularisation strength.

- p_l is the p-value for λ_l based on GLS inference.

- $\tau \in (0, 1)$ is the significance threshold (set to 0.05 in the program).

- ρ_{pval} penalizes coefficients with p-values above threshold.

Equation 3.12 represents a heuristic one-step penalised GLS estimator, where the weights and p-values are computed from a fixed preliminary stage. A better optimisation would require iteratively updating both the GLS weights and the p-value penalties until convergence.

The GLS covariance matrix for the estimated risk prices is given by:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\lambda}}) = \frac{1}{T} \left(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1}. \tag{3.13}$$

where $X = [\mathbf{1}, \mathbf{B}]$ is the matrix of regressors including an intercept and the time-series betas. This expression accounts for heteroscedasticity and cross-sectional correlation of errors. For each risk price λ_l, the standard error is given by:

$$\text{SE}(\hat{\lambda}_l) = \sqrt{\left[\frac{1}{T} \left(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} \right]_{ll}}.$$

We compute the t-statistic:

$$t_l = \frac{\hat{\lambda}_l}{\text{SE}(\hat{\lambda}_l)}, \qquad p_l = 2\left(1 - \Phi\left(|t_l|\right)\right), \tag{3.14}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The penalty term $\sum_l \mathsf{max}(0, p_l - \tau)|\lambda_l|$ will shrink coefficients with weak statistical significance, p-value $= p_l > \tau$, while retaining factors that are still significant for marginal exposures near the threshold.

3.4 Empirical Applications

3.4.1 Global Economies

3.4.1.1 Data Processing

In the experiments, we use the market returns adjusted by the risk-free rate for Europe (EU), North America (US), Asia (AS), and Japan (JP), and the Europe six size and value Portfolios from the French (2025) website from January 2003 to December 2022, 240 monthly observations.

3.4.1.2 Data Analysis

We use location-dispersion ellipsoids to visualize the joint residuals of two portfolios under each exposure regularisation method, equations 3.8, 3.9, 3.10. Ellipsoids are constructed from the covariance matrix of residuals and visualize the remaining (idiosyncratic) risk after accounting for common risk drivers (factors). A circular shape implies low or no correlation between residuals, while an elongated ellipse indicates significant correlation. The orientation of the ellipse reflects the direction of maximal joint variability, i.e., the first principal component of the residuals. The ellipsoid summarizes the joint residual distribution between portfolios; in an ideal case of perfect invariance, the ellipsoid is a small circle aligned with the axes.

Although each ellipsoid in Figure 3.1 compares the residuals of just two portfolios, these residuals are derived from a global multifactor model that is estimated jointly across all portfolios and all factors.

For the EU portfolios, such as `SSLV_EU` versus `SSMV_EU`, the ellipsoids are elongated, indicating a strong residual correlation. This suggests that despite regularisation, these portfolios share exposure to common factors that are not fully captured by the model. When comparing `SSLV_EU` with non-EU portfolios, such as `SSLV_AS` or `SSMV_US`, the ellipsoids are more circular and tighter, reflecting weaker and more idiosyncratic residual relationships.

Across all comparisons, the ellipsoids from the Elastic Net estimates with $\alpha = 0.5$ lie between those from Lasso and Ridge.

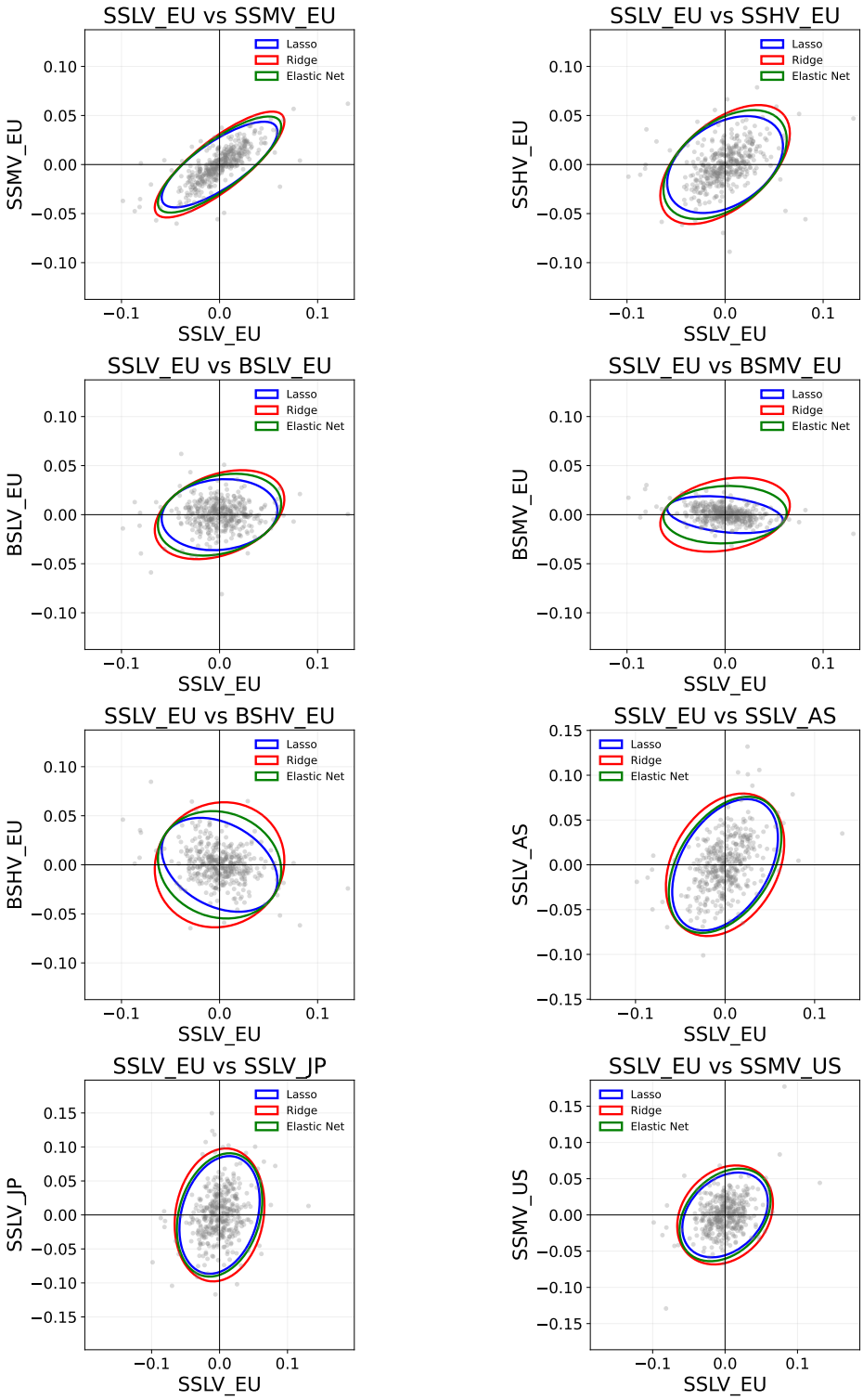

Figure 3.1: Cross Sectional Ellipsoids of Residuals, EU pairs and EU non-EU portfolios

3.4.1.3 One-Factor Model

We consider a one-factor asset pricing model, defined as a system of nonlinear equations in equation 2.10.

We use the copula density log-likelihood optimisation to estimate the parameters of the cross-sectional asset pricing model, and compare the results to the Fama-MacBeth (FM), SUR, and GMM approaches. We use SAS proprietary software `PROC MODEL` to implement the OLS, SUR, and GMM methods, and custom Python code for the copula optimisation and the Fama-MacBeth procedure with both OLS and GLS estimators.

We show the different regression results in Table 3.1 for the demeaned factor. The results are divided into four different sections, each separated by a double line.

In the first section, we present the Gauss copula density likelihood method results (univariate, uni) and multivariate (multi), followed by the same result for the Student-t copula density likelihood where the degrees of freedom (abbreviated with d.o.f. and shown with ν, nu in the table) are optimised in the objective function. The univariate and multivariate Student-t optimisations estimate slightly different degrees of freedom: 2.72 vs 7.23.

In the second section, we present the results of different fixed degrees of freedom Student-t univariate and multivariate optimisations.

The last two sections contain the results of the previous chapter experiments, respectively: the non linear L1 approximation methods and the standard methods.

In the Gauss multivariate (abbreviated multi) approach, the copula correlation matrix $\mathbf{P}$ captures the cross-sectional dependence of residuals, similar to the GLS SUR covariance matrix in equation 4.50:

$$\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t^\top .$$

However, instead of using only second moments of the residuals in the original scale, the copula likelihood models the dependence structure implied by the joint correlation of transformed ranks (probability integral transforms). This parametric approach is fully determined by the correlation matrix $\mathbf{P}$ for the Gaussian copula, but will capture fatter tail dependence or asymmetric relationships, with different copula for example: the t-copula for heavy tails, the Skew t-copula for asymmetric tails.

Copula multivariate likelihood estimation assumes that standardised residuals are i.i.d. across time, while allowing for dependence across equations (cross-sectional correlation), which is mod-

elled by the multivariate copula itself. HAC adjusts the estimated variance-covariance matrix of the parameter estimates to remain consistent under autocorrelation and heteroscedasticity while cross-sectional dependence remains in the error terms[5], so it is ideal to be used together with copula likelihood.

The intercept and risk price estimated with the multivariate copula likelihood have similar values of the parameters obtained by GMM or GLS shown at the bottom of the table. Both methods aim to achieve efficient weighting of the moment conditions: the copula density uses all available information about the dependence structure of the residuals, while GMM achieves robustness by consistently estimating the covariance of the moment conditions without assuming a parametric model. Only if the copula is correctly specified, the copula density maximum likelihood estimator is asymptotically efficient.

On the other hand, the univariate methods' parameters are close to the intercept and risk price estimated with FM OLS, where standard errors are computed assuming homoscedasticity and independence. In the univariate methods, the likelihood treats all residuals as independent over time and across assets: the estimation does not weight residuals to account for cross-sectional correlation or cross-sectional dependence.

In the second section of the table, where we apply the univariate and multivariate Student-t copula density likelihood for a fixed range of degrees of freedom ν, the convergence of the Student-t distribution to the Gaussian is quite slow in numerical estimation: ν in the order of $10{,}000$ to $100{,}000$ [6], analogue results have been reported in the literature, see Lange et al. (1989) and Fernandez and Steel (1998).

For the copula PIT we use the notation Empirical (E) and Parametric (P). The univariate Gaussian with a risk price of 0.0111 (p-value 0.99), closely matching FM and OLS estimates, with equivalent high p-value. As expected, the Gauss multivariate values, -0.698 (p-value 0.21), are close to the FIML, GLS, SUR standard methods. Under multivariate copula specifications, standard error and estimates are comparable with the classic methods.

Employing a Student-t likelihood with a small value of ν (e.g., $\nu = 3$ or $\nu = 7$) is not equivalent to performing an L1-norm regression. In our experiment an equivalent L1 risk price estimate is reached only with $\nu > 100$ in the Student-t Univariate. While both approaches reduce the influence of extreme residuals compared with the Gaussian, they are not equivalent. The Student-

[5] In contrast, prewhitening explicitly transforms the residuals to remove autocorrelation.

[6] Although the density of the standardised Student-t approaches the standard normal density in distribution as $\nu \to \infty$, in practice the log-likelihood retains differences even for moderately large ν: the tail decay rate remains distinctly heavier than the normal for ν up to several hundreds, which affects the curvature of the objective function and therefore the parameter estimates.

Method	Parameter	Estimate	Std. Error	t-Statistic	*p*-value
Gauss Uni	Intercept	0.7483	0.6898	1.0848	0.2780
Gauss Uni	Risk price	0.0111	0.6837	0.0162	0.9871
Gauss Copula Uni E P	Intercept	0.7483	0.6901	1.0844	0.2782
Gauss Copula Uni E P	Risk price	0.0111	0.6839	0.0162	0.9871
Gauss Multi	Intercept	1.3961	0.5595	2.4953	0.0126
Gauss Multi	Risk price	-0.6982	0.5609	-1.2448	0.2132
Gauss Copula Multi P	Intercept	1.3961	0.5600	2.4930	0.0127
Gauss Copula Multi P	Risk price	-0.6982	0.5616	-1.2433	0.2138
Gauss Copula Multi E	Intercept	1.3760	0.5624	2.4467	0.0144
Gauss Copula Multi E	Risk price	-0.6776	0.5640	-1.2015	0.2296
t Uni nu = 2.72	Intercept	1.3038	0.7433	1.7540	0.0794
t Uni nu = 2.72	Risk price	-0.5948	0.7461	-0.7972	0.4253
t Copula Uni E [1] P [2]	Intercept	1.3038	0.7433	1.7540	0.0794
t Copula Uni E [1] P [2]	Risk price	-0.5948	0.7461	-0.7972	0.4253
t Multi nu = 7.23	Intercept	1.7852	0.5854	3.0493	0.0023
t Multi nu = 7.23	Risk price	-1.0949	0.5858	-1.8691	0.0616
t Copula Multi P [3]	Intercept	1.7973	0.6476	2.7755	0.0055
t Copula Multi P [3]	Risk price	-1.1065	0.6503	-1.7014	0.0889
t Copula Multi E [4]	Intercept	1.7291	0.6342	2.7262	0.0064
t Copula Multi E [4]	Risk price	-1.0380	0.6367	-1.6304	0.1030
t Uni nu = 2.72	Intercept	1.3037	0.7432	1.7543	0.0794
t Uni nu = 2.72	Risk price	-0.5947	0.7461	-0.7971	0.4254
t Uni nu = 25	Intercept	1.0773	0.6818	1.5800	0.1141
t Uni nu = 25	Risk price	-0.3332	0.6795	-0.4903	0.6239
t Uni nu = 100	Intercept	0.8946	0.6767	1.3219	0.1862
t Uni nu = 100	Risk price	-0.1405	0.6720	-0.2091	0.8344
t Uni nu = 10000	Intercept	0.7504	0.6897	1.0880	0.2766
t Uni nu = 10000	Risk price	0.0089	0.6836	0.0130	0.9896
t Multi nu = 7.23	Intercept	1.7851	0.5858	3.0475	0.0023
t Multi nu = 7.23	Risk price	-1.0948	0.5861	-1.8681	0.0618
t Multi nu = 25	Intercept	1.6128	0.5665	2.8468	0.0044
t Multi nu = 25	Risk price	-0.9188	0.5674	-1.6191	0.1054
t Multi nu = 100	Intercept	1.4718	0.5679	2.5918	0.0095
t Multi nu = 100	Risk price	-0.7752	0.5691	-1.3621	0.1732
t Multi nu = 10000	Intercept	1.3969	0.5696	2.4526	0.0142
t Multi nu = 10000	Risk price	-0.6991	0.5710	-1.2244	0.2208
Nonlinear L1	Intercept	0.8290	0.5159	1.6070	0.1083
Nonlinear L1	Risk price	-0.1284	0.5088	-0.2523	0.8009
FM	Intercept	0.7487	0.6514	1.149	0.2504
FM	Risk price	0.0107	0.7332	0.015	0.9884
OLS	Intercept	0.7487	0.4422	1.69	0.0918
OLS	Risk price	0.0107	0.4434	0.02	0.9808
FIML	Intercept	1.3932	0.5155	2.700	0.0074
FIML	Risk price	-0.6954	0.5164	-1.350	0.1794
GLS	Intercept	1.3615	0.4937	2.758	0.0063
GLS	Risk price	-0.6636	0.4945	-1.342	0.1809
SUR	Intercept	1.3786	0.4997	2.760	0.0062
SUR	Risk price	-0.6793	0.5009	-1.360	0.1763
GMM	Intercept	1.3962	0.5051	2.760	0.0061
GMM	Risk price	-0.6973	0.5067	-1.380	0.1700

[1] The estimated d.o.f. of the marginals are 2.01 and the d.o.f. of the copula are 8.01

[2] The estimated d.o.f. of the marginals are 2.01 and the d.o.f. of the copula are 18

[3] The estimated d.o.f. of the marginals are 2.01 and the d.o.f. of the copula are 7.01

[4] The estimated d.o.f. of the marginals are 2.01 and the d.o.f. of the copula are 8.01

Table 3.1: One-Factor Risk Price Estimates, Copula, Centred Factor

t log-likelihood remains a smooth and differentiable function, penalizing squared residuals scaled by a heavy-tailed variance factor. In contrast, the L1 regression corresponds to the minimization of the sum of absolute errors, which yields a piecewise linear (non-differentiable at zero residual) objective function. As a result, small-ν Student-t estimators can behave similarly to robust regression but do not possess the same properties as quantile regression or L1 loss minimization: see Koenker and Hallock (2001) Lange et al. (1989) for a detailed comparison.

Under a different modelling approach, our results make it clear that student-t copula leverages robustness in the estimation of the risk price, while the multivariate copula density recovers the cross-sectional GLS results. Possibly, the most robust estimates are computed with the Student-t multivariate method where both properties are present.

Although the computation of the multivariate copula density might be less efficient than the standard methods, the technique is more robust when applied to skewed or heavy-tailed distributions, while its estimates are consistent with the classic approaches.

3.4.1.4 Integration and Segmentation Simulation

This section presents a Monte Carlo simulation of the cross-sectional expected returns integration and segmentation models.

The initial dataset contains 240 monthly observations of European and US Market factors, from January 2003 to December 2022, and their orthogonal components as described in the previous chapter.

We construct synthetic time series ($T = 500$) by resampling. Each factor is centred and orthogonalised so that: $\mathsf{E}[\tilde{f}_1] = \mathsf{E}[\tilde{f}_2] = 0, \quad \mathsf{Corr}(\tilde{f}_1, \tilde{f}_2) = 0.$

For each scenario, we simulate $N = 6$ portfolio returns using equation 2.51:

$$R_{i,t} = \beta_{i,1}(\tilde{f}_{1,t} + \lambda_1) + \beta_{i,2}(\tilde{f}_{2,t} + \lambda_2) + \varepsilon_{i,t}, \quad \forall i, t$$

where: $\beta_{i,1}, \beta_{i,2} \sim \mathcal{U}(0.5, 1.0)$ i.i.d., and $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.5\%$.

We consider two scenarios with two models for integration and segmentation each, with different risk price that define how factors are priced. In Scenario 1, both the global (US) and local (EU) factors are priced in both the integration and the segmentation models. In Scenario 2, only the

global (US) factor is priced in both models:

$$\begin{array}{lllll}
\text{Integration Scenario 1} & \tilde{f}_1 = \text{US}, & \tilde{f}_2 = \text{EU}_\perp\text{US}, & (\lambda_1 = 0.6\%, \lambda_2 = 0.4\%) \\
\text{Segmentation Scenario 1} & \tilde{f}_1 = \text{EU}, & \tilde{f}_2 = \text{US}_\perp\text{EU}, & (\lambda_1 = 0.6\%, \lambda_2 = 0.4\%) \\
\text{Integration Scenario 2} & \tilde{f}_1 = \text{US}, & \tilde{f}_2 = \text{EU}_\perp\text{US}, & (\lambda_1 = 0.6\%, \lambda_2 = 0) \\
\text{Segmentation Scenario 2} & \tilde{f}_1 = \text{EU}, & \tilde{f}_2 = \text{US}_\perp\text{EU}, & (\lambda_1 = 0, \lambda_2 = 0.6\%)
\end{array}$$

Each simulation is run with 500 repetitions. For each replication, we estimate factor risk prices using the standard two-pass Fama–MacBeth approach, as in equations 2.7 and 2.8. We report the average estimates, standard errors, t-statistics, and p-values in Table 3.2.

Scenario	**Parameter**	**Estimate**	**Std. Error**	**t-Statistic**	***p*-value**
Integration Scenario 1	Intercept	0.0057	0.0952	0.0280	0.4909
Integration Scenario 1	λ_1	0.6013	0.2163	2.7972	0.0135
Integration Scenario 1	λ_2	0.3909	0.1469	2.7299	0.0387
Segmentation Scenario 1	Intercept	-0.0147	0.0935	-0.1169	0.4848
Segmentation Scenario 1	λ_1	0.6133	0.2495	2.4646	0.0228
Segmentation Scenario 1	λ_2	0.4072	0.1340	3.1368	0.0233
Integration Scenario 2	Intercept	-0.0077	0.0933	-0.0348	0.4812
Integration Scenario 2	λ_1	0.6052	0.2160	2.8244	0.0139
Integration Scenario 2	λ_2	0.0063	0.1473	0.0261	0.6443
Segmentation Scenario 2	Intercept	-0.0063	0.0956	-0.0354	0.5049
Segmentation Scenario 2	λ_1	0.0037	0.2506	0.0104	0.7889
Segmentation Scenario 2	λ_2	0.6040	0.1332	4.6687	0.0043

Table 3.2: Estimated Risk Prices, Simulated Data

As expected:

- In Scenario 1, both factors are priced (significant λ_1 and λ_2).
- In Scenario 2, only the global factor (f_1) and its orthogonal component, both with nonzero λ are significantly priced.
- The intercept is always close to zero.

The average correlation matrices between portfolios and factors, $\text{Corr}(R_i, \tilde{f}_l)$, for $i = 1, \ldots, N,\ l = 1, \ldots, \bar{\ell}$, are highly similar across scenarios, $\beta_{i,1}, \beta_{i,2} \sim \mathcal{U}(0.5, 1.0)$ i.i.d., and only the expected risk prices (λ_l) differ.

The simulation shows that correlation alone does not imply pricing. It is the cross-sectional covariance with expected returns that determines whether a factor is priced. In Table 3.3, we report

the average cross-sectional covariance and correlation between estimated betas and expected returns over 500 simulations.

Scenario	Factor	Mean Covariance	Std. Dev.	Mean Correlation
Integration Scenario 1	Factor 1	0.01264	0.00732	0.76975
Integration Scenario 1	Factor 2	0.00841	0.00668	0.48962
Segmentation Scenario 1	Factor 1	0.01246	0.00694	0.77449
Segmentation Scenario 1	Factor 2	0.00848	0.00701	0.51363
Integration Scenario 2	Factor 1	0.01295	0.00604	0.95390
Integration Scenario 2	Factor 2	-0.00004	0.00608	-0.01251
Segmentation Scenario 2	Factor 1	-0.00041	0.00559	-0.04212
Segmentation Scenario 2	Factor 2	0.01248	0.00583	0.95350

Table 3.3: Cross-sectional Covariance and Correlation, Simulated Data

In order to demonstrate the distinction between exposure and pricing, for each of the four scenarios previously simulated:

1. Integration Scenario 1: both factors priced,
2. Segmentation Scenario 1: both factors priced,
3. Integration Scenario 2: only factor 1 priced,
4. Segmentation Scenario 2: only factor 2 priced,

we define two different measures of dependence:

1. The dependence between the sum of the two demeaned factors X_t and the returns $R_{i,t}$, where:
$$X_t = \tilde{f}_{1,t} + \tilde{f}_{2,t}, \quad \text{for all } t$$
2. The dependence between the stochastic discount factor m_t and the returns $R_{i,t}$, where:
$$m_t = 1 - \lambda_1 \tilde{f}_{1,t} - \lambda_2 \tilde{f}_{2,t}, \quad \text{for all } t$$

Tables 3.4 and 3.5 report the results for the first portfolio, $i = 1$.

First, we compute the simple Pearson correlation (without copula) for both measures[7]:

$$\rho_{\text{sum}} = \rho^{\text{Pearson}}\big(\tilde{f}_{1,t} + \tilde{f}_{2,t}, \quad R_{1,t}\big)$$

[7] For theoretical moments of the random variables we suppress the time index, while for empirical correlations we keep it to emphasize the time-series dimension

and

$$\rho_{\text{SDF}} = \rho^{\text{Pearson}}\big(m_t, \quad R_{1,t}\big), \quad m_t = 1 - \lambda_1 \tilde{f}_{1,t} - \lambda_2 \tilde{f}_{2,t}.$$

Scenario	$\boldsymbol{\rho}_{\text{sum}}$	$\boldsymbol{\rho}_{\text{SDF}}$
Integration Scenario 1	$+0.935$	-0.900
Segmentation Scenario 1	$+0.921$	-0.923
Integration Scenario 2	$+0.900$	-0.643
Segmentation Scenario 2	$+0.905$	-0.627

Table 3.4: Pearson Correlations, Factors and Returns, Simulated Data

Then, we compute the Gaussian copula (C) correlations, see Section 3.2.3:

$$\rho_{C,\,\text{sum}} = \text{Corr}\Big(\Phi^{-1}\big(F_{\tilde{f}_1+\tilde{f}_2}\,(\,\tilde{f}_{1,t} + \tilde{f}_{2,t})\big), \quad \Phi^{-1}\big(F_R\,(R_{1,t})\big)\Big),$$

where $F_{\tilde{f}_1+\tilde{f}_2}$ and F_R denote the marginal cumulative distribution functions.

$$\rho_{C,\,\text{SDF}} = \text{Corr}\Big(\Phi^{-1}\big(F_m\,(m_t)\big), \quad \Phi^{-1}\big(F_R\,(R_{1,t})\big)\Big), \quad m_t = 1 \;-\; \lambda_1\,\tilde{f}_{1,t} \;-\; \lambda_2\,\tilde{f}_{2,t}.$$

Scenario	**Copula** $\boldsymbol{\rho}_{C,\text{sum}}$	**Copula** $\boldsymbol{\rho}_{C,\text{SDF}}$
Integration Scenario 1	$+0.881$	-0.813
Segmentation Scenario 1	$+0.877$	-0.814
Integration Scenario 2	$+0.882$	-0.670
Segmentation Scenario 2	$+0.879$	-0.676

Table 3.5: Gaussian Copula Correlations, Factors and Returns, Simulated Data

The strong positive dependence between $(\tilde{f}_1 + \tilde{f}_2)$ and returns arises because both factors contribute to return variability through the exposures $\beta_{1,l}$, regardless of whether they are priced. Even in scenarios where a factor is unpriced (i.e., $\lambda_l = 0$), it continues to affect return variation, which leads to strong positive dependence between the sum of factors and returns.

By contrast, the SDF dependence decreases in scenarios where only one-factor is priced, because it does not take into account unpriced factor variance. The negative dependence between the SDF and returns reflects how the stochastic discount factor assigns higher value to payoffs in bad times, when marginal utility is high and consumption is low.

Figures 3.2 and 3.3 display the scatter plots across the four scenarios of the transformed Gaussian scores (PIT) used in the copula estimation for the two different measures.

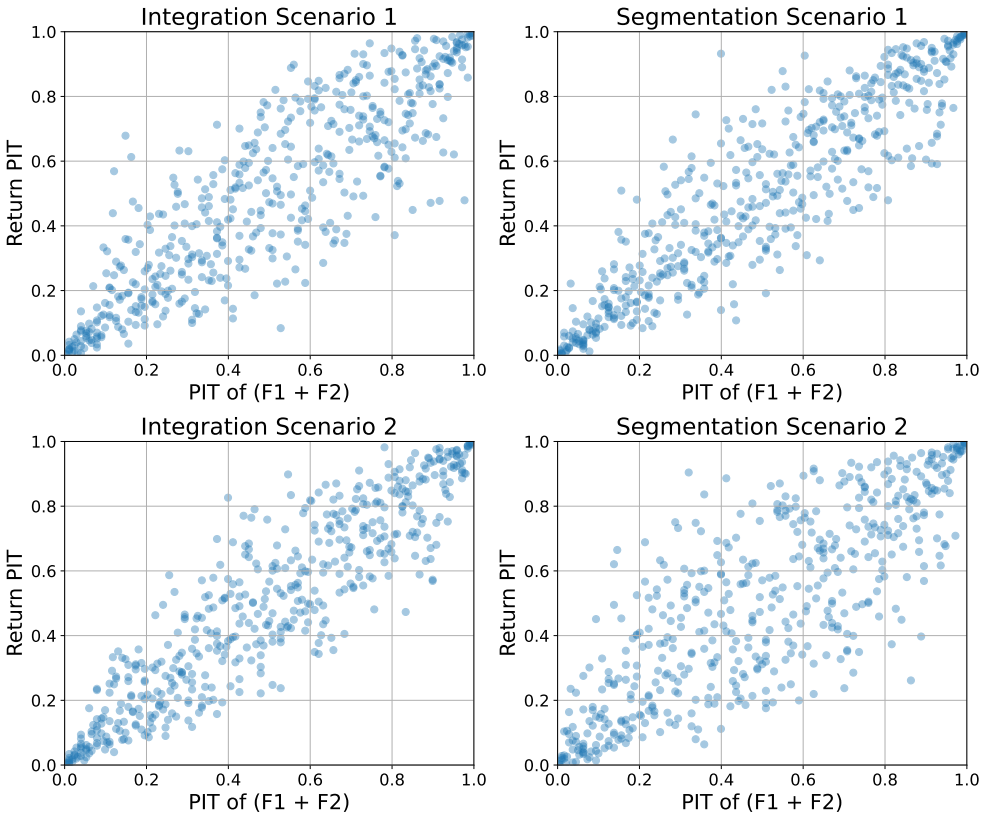

Figure 3.2: PIT Scatter Plots, Factors and Returns, Simulated Data

These plots confirm the numerical results: the dependence between the sum of factors and returns is strong and positive, while the dependence between the SDF and returns is negative and weaker in scenarios with only one priced factor.

Although the correlation analysis already explains the distinction between exposure and pricing – while exposures $\beta_{1,l}$ determine how much returns fluctuate with factors, only nonzero λ_l generate compensation for bearing factor risk and affect the expected returns in the pricing equation – we show the copula dependence measures because they characterize the entire joint distribution, including tail dependence and rank co-movement.

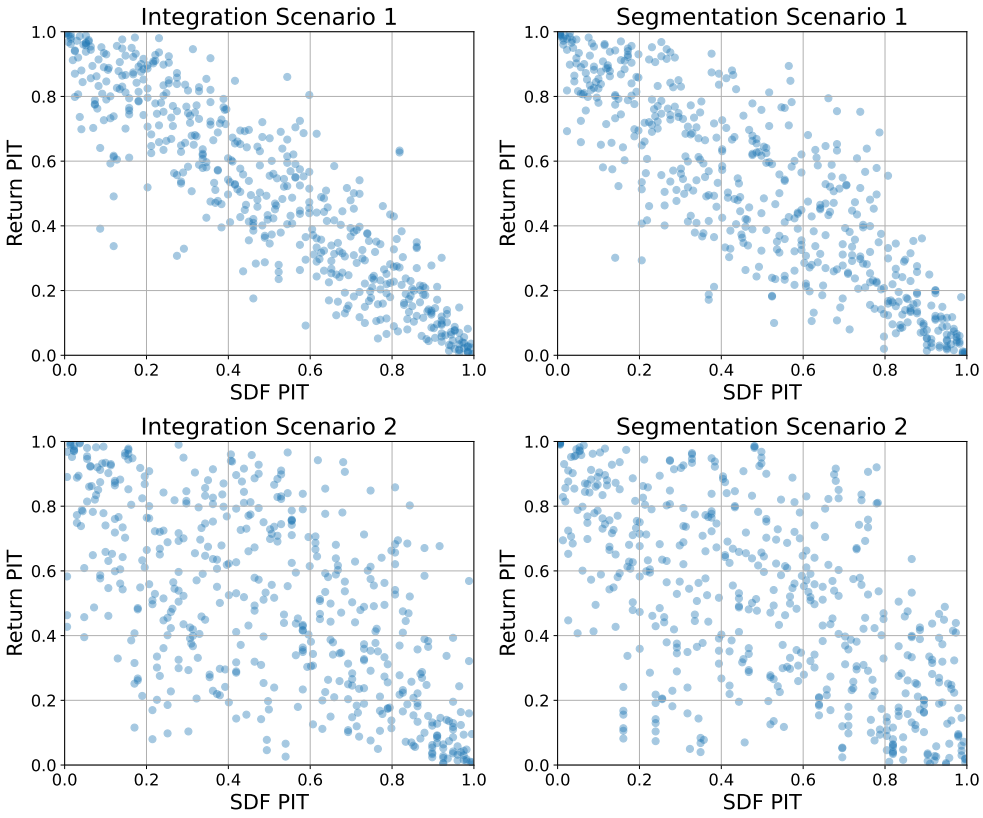

Figure 3.3: PIT Scatter Plots, SDF and Returns, Simulated Data

3.4.1.5 Integration and Segmentation

This section reports the dependence analysis between the factors and the returns using the historical data: the observed returns of six European portfolios and two factors, the US market, the EU market and their orthogonal components, over the period January 2003 to December 2022 (240 observations).

Integration

For the integration model, refer to equation 2.66. The estimated Pearson correlations between the sum of factors and returns, shown in Table 3.6, are extremely high, ranging from approximately $+0.93$ to $+0.99$, which indicates that realised factor innovations strongly co-move with portfolio returns in the time series.

The Gaussian copula correlations in Table 3.7 show a similar pattern.

The stochastic discount factor was constructed using a negative estimated risk price on the US market factor ($\hat{\lambda}_{\text{US}} = -1.96$, highly significant, p-value $= 0.001$), while the EU orthogonal factor

Portfolio	ρ_{sum}	ρ_{SDF}
SS_VH	$+0.950$	$+0.812$
SS_VM	$+0.957$	$+0.844$
SS_VL	$+0.935$	$+0.849$
SB_VH	$+0.956$	$+0.821$
SB_VM	$+0.993$	$+0.866$
SB_VL	$+0.958$	$+0.860$

Table 3.6: Integration Correlations, Factors and Returns, Equities

Portfolio	$\rho_{C,\text{sum}}$	$\rho_{C,\text{SDF}}$
SS_VH	$+0.945$	$+0.780$
SS_VM	$+0.952$	$+0.813$
SS_VL	$+0.932$	$+0.821$
SB_VH	$+0.951$	$+0.797$
SB_VM	$+0.990$	$+0.849$
SB_VL	$+0.956$	$+0.849$

Table 3.7: Integration Copula Correlations, Factors and Returns, Equities

was found to be statistically insignificant ($\hat{\lambda}_{\text{EU}} = 0.733$) and therefore set to zero in the SDF calculation: the SDF behaves as an inverse of the US market factor.

The resulting correlations between the SDF and returns are positive and large (approximately $+0.78$ to $+0.85$), in contrast to classic asset pricing intuition, where the SDF is typically negatively correlated with returns. In this case, the significant negative risk price implies that exposure to the US factor reduces expected returns, so the SDF increases in high-return states, generating positive dependence.

Figures 3.4 and 3.5 display the scatter plots of the transformed Gaussian scores (PIT) used in the copula estimation. In Figure 3.4, the PIT ranks of the sum of factors and returns are very tight clustered along the diagonal, reflecting the high dependence measured in the correlation tables. In contrast, Figure 3.5 shows visibly more dispersed ranks for the SDF, consistent with the lower copula correlations.

The greater dispersion of the SDF PIT reflects that, with the EU orthogonal factor excluded, the SDF depends primarily on a single factor and therefore provides less stable rankings of return realizations. Consistent with this, the Gaussian copula correlations between the SDF and returns are uniformly lower than the correlations between the sum of factors and returns, highlighting how the SDF, compared to the factor sum, captures only the priced component of this variation.

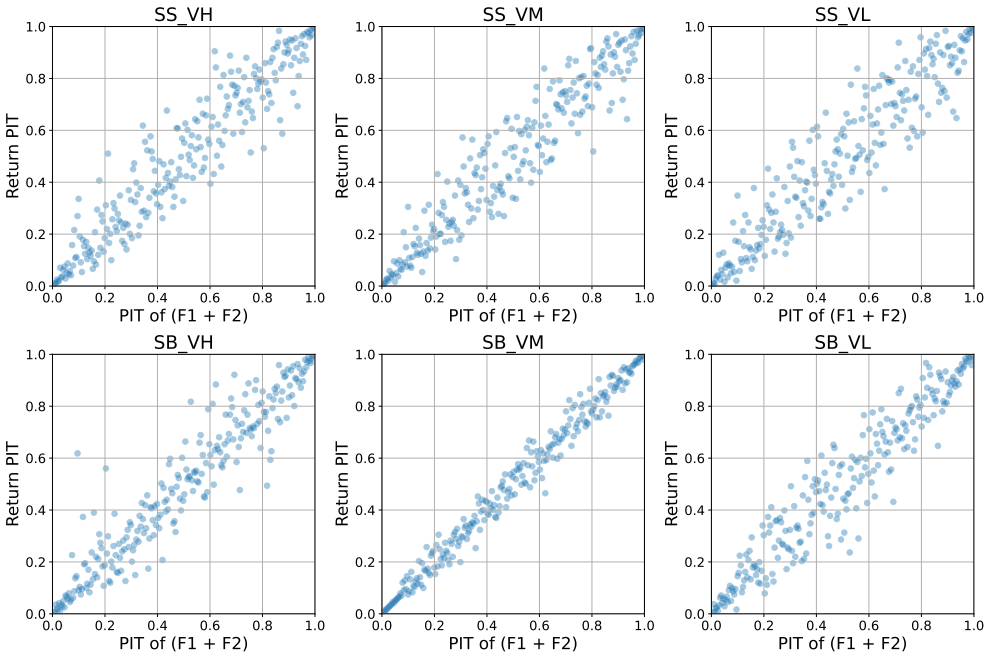

Figure 3.4: Integration PIT, Factors and Returns, Equities

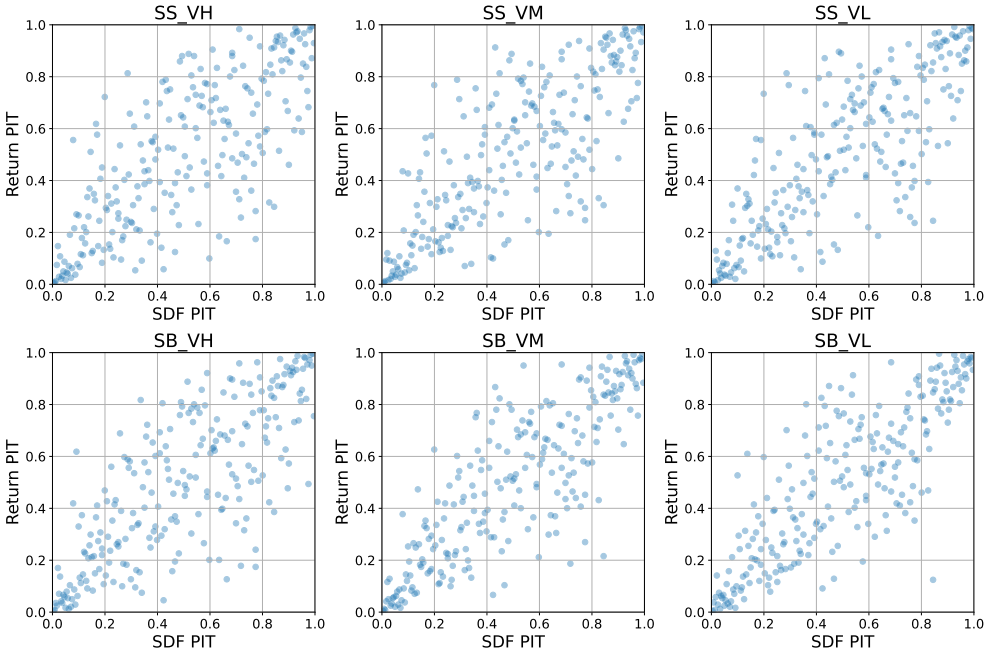

Figure 3.5: Integration PIT, SDF and Returns, Equities

Segmentation

For the segmentation model, refer to equation 2.67. The estimated Pearson correlations between the sum of factors and returns, shown in Table 3.8, are high (approximately $+0.86$ to $+0.92$).

Portfolio	ρ_{sum}	ρ_{SDF}
SS_VH	$+0.864$	$+0.894$
SS_VM	$+0.891$	$+0.917$
SS_VL	$+0.890$	$+0.911$
SB_VH	$+0.873$	$+0.902$
SB_VM	$+0.918$	$+0.946$
SB_VL	$+0.904$	$+0.928$

Table 3.8: Segmentation Correlations, Factors and Returns, Equities

The Gaussian copula correlations in Table 3.9 show a similar pattern.

Portfolio	$\rho_{C,\text{sum}}$	$\rho_{C,\text{SDF}}$
SS_VH	$+0.841$	$+0.880$
SS_VM	$+0.868$	$+0.902$
SS_VL	$+0.869$	$+0.900$
SB_VH	$+0.858$	$+0.894$
SB_VM	$+0.906$	$+0.941$
SB_VL	$+0.896$	$+0.927$

Table 3.9: Segmentation Copula Correlations, Factors and Returns, Equities

The segmentation risk price on the EU market factor ($\hat{\delta}_{\text{EU}} = -1.31$, p-value $= 0.014$) is significant. The US orthogonal factor was found to be statistically significant as well ($\hat{\delta}_{\text{US}} = -1.00$, p-value $= 0.031$).

In contrast to the integration scenario, the correlations between the SDF and returns are uniformly higher than those between the sum of factors and returns, suggesting that the SDF captures a larger portion of the return variation.

Figures 3.6 and 3.7 display the scatter plots of the Gaussian scores (PIT) used in the copula estimation. The PIT scatter plots confirm that the ranks of the SDF are more concentrated along the diagonal with the return ranks, showing less dispersion compared to the PITs of the sum of factors: the SDF ranks track return ranks more closely than the raw factor sum, reflecting the different role of pricing and exposure when the US orthogonal factor is not excluded.

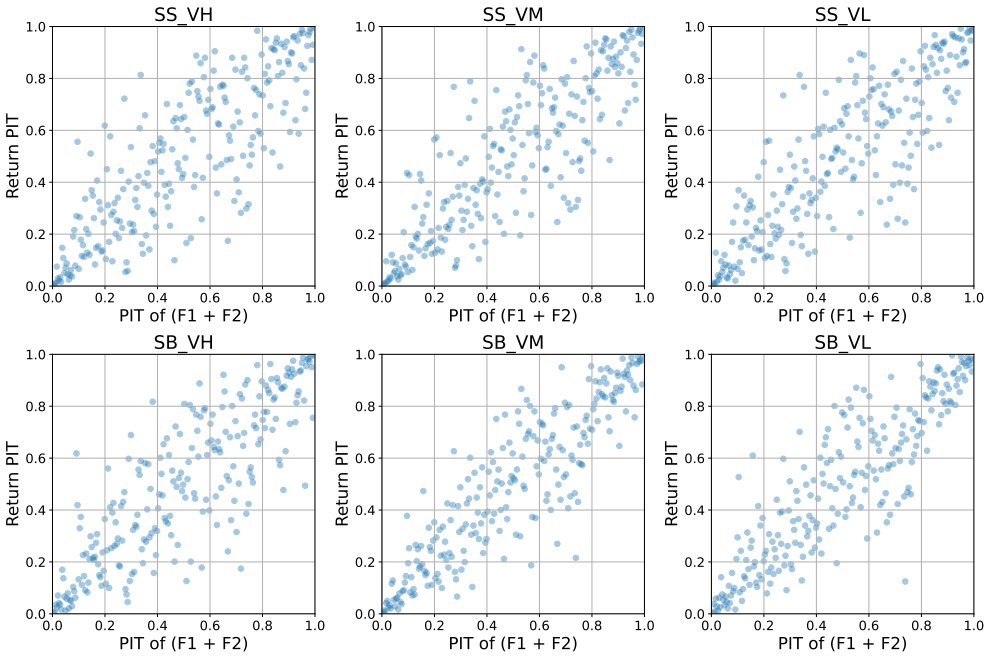

Figure 3.6: Segmentation PIT, Factors and Returns, Equities

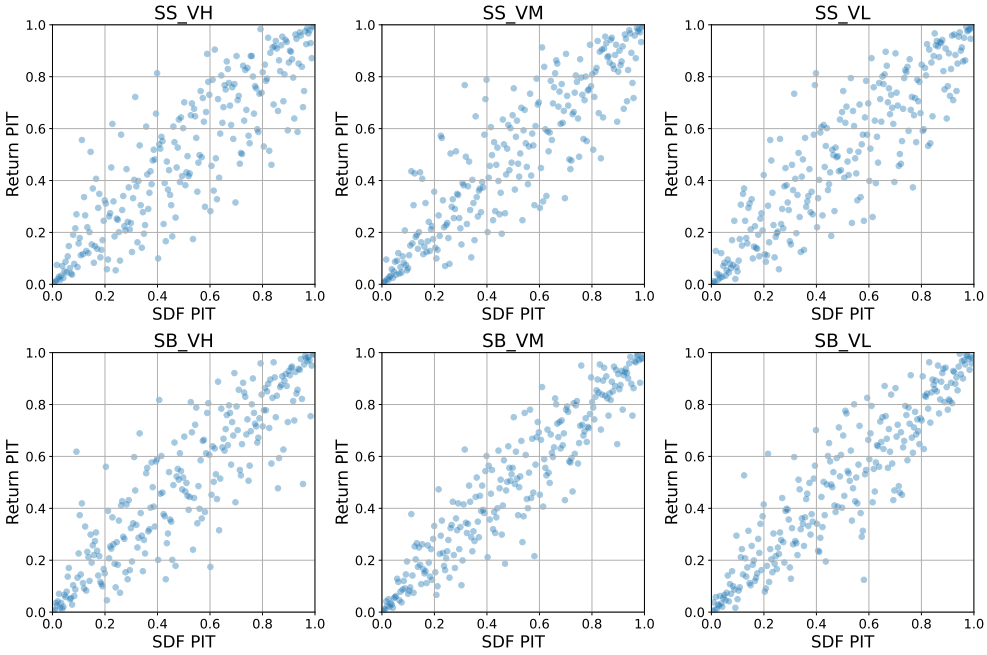

Figure 3.7: Segmentation PIT, SDF and Returns, Equities

3.4.1.6 Regularisation

In the regularisation of global economies, we use a panel of global size and value portfolios excess returns (Asia, AS, Europe, EU, Japan, JP, and North America, US) together with the corresponding global market factors excess returns. The dataset, consisting of 24 portfolios and 4 market factors covering the period from January 2003 to December 2022, was downloaded from the Fama–French website.

To analyse the effect of the different regularisation methods on factor loadings for global market economies, equations 3.8, 3.9, 3.10, we generate heatmaps of the estimated coefficients for each technique with the same penalty value ($\rho = 0.001$): Figures 3.8, 3.9, 3.10. The heatmaps help identify which factors remain relevant across portfolios as regularisation is applied. Darker

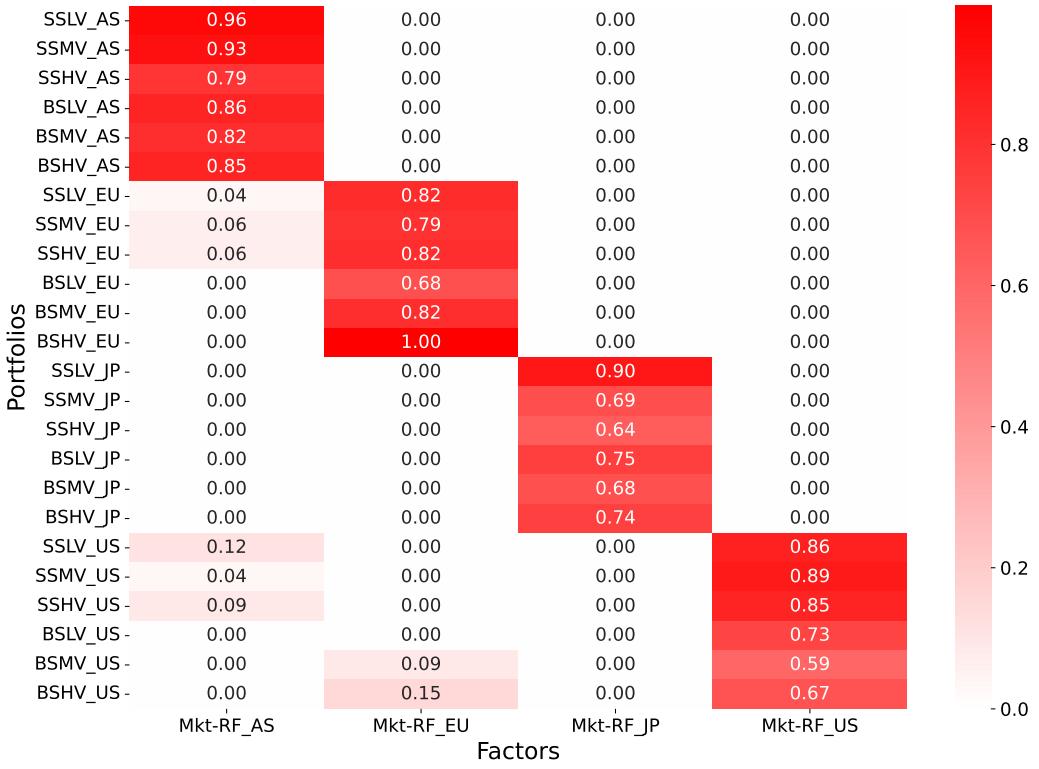

Figure 3.8: Lasso Loadings Heatmap

colors indicate stronger exposures; white areas correspond to zero loadings due to regularisation. The Lasso regularisation produces sparse solutions: it effectively filters out factors that are not locally relevant. For instance, only the European market factor retains significant loadings in the European portfolios, while most of the non-local factors (e.g., U.S. or Asia-Pacific) are driven to exact zeros: the only exceptions are Asia for the US portfolio and EU for Big in Size and High Book-to-Market Value US portfolio. This selective sparsity reflects Lasso's strength in performing

variable selection by eliminating irrelevant predictors entirely. In contrast, Ridge tends to preserve

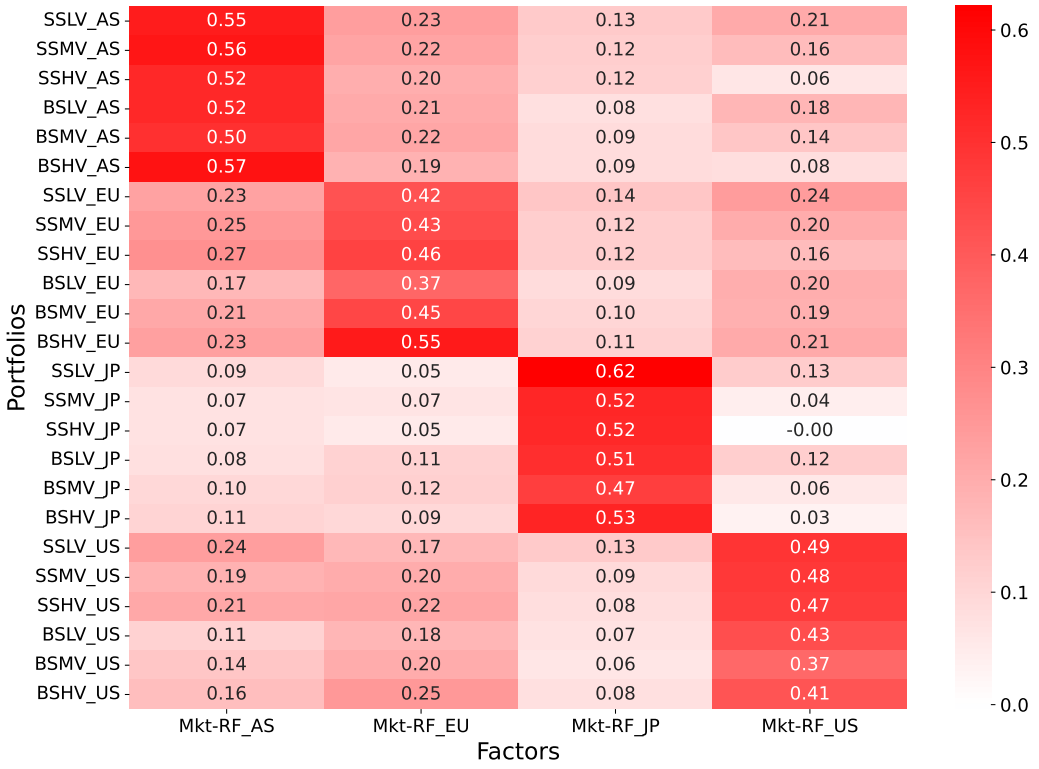

Figure 3.9: Ridge Loadings Heatmap

more diffuse structure across portfolios, under the same penalty, it retains non-local factors, with smaller coefficients. The result is a more diffuse dependence structure, more suited for models that assume some degree of global integration or where multicollinearity among predictors is present.

Elastic Net regularisation combines the properties of both Lasso and Ridge by combining L1 and L2 penalties. On one hand, it filters out irrelevant global market factors, helping reveal region-specific drivers as Lasso does. On the other hand, it avoids over-aggressive zeroing that may occur under pure Lasso, preserving some structure when factors exhibit collinearity—as is often the case across international markets. With a balanced penalty ($\rho = 0.001$, $\alpha = 0.5$), Elastic Net offers a flexible trade-off, particularly in cross-sectional financial data where local and global exposures may overlap.

Although heatmaps of exposures provide a clear visual summary of factor relevance, they do not take into account collinearity and assumes that risk price is not relevant to the model.

Then, we switch to the Lasso p-value penalised model, where we introduce orthogonal factors to address collinearity and account for marginal contributions, and regularise the Fama-MacBeth

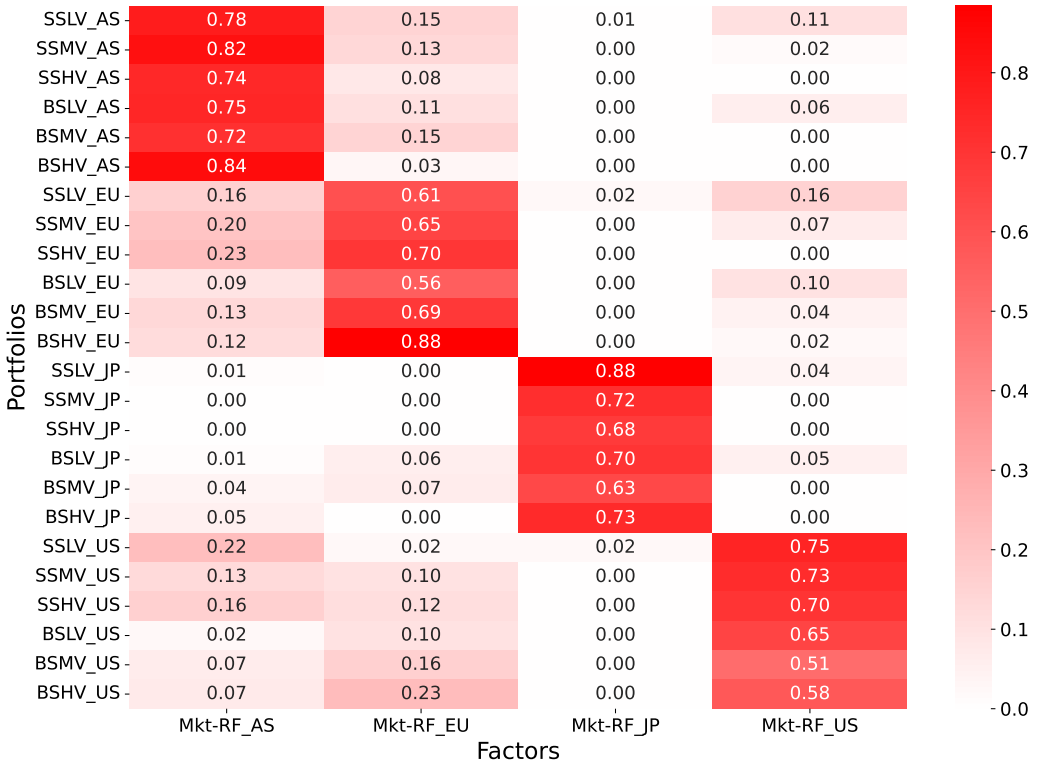

Figure 3.10: Elastic Net Loadings Heatmap.

second step using the Lasso penalty based on betas together with a soft constraint to shrink factors with high p-values (factors above a threshold p-value will be considered spurious), see equation 3.12. The soft p-value penalty is introduced directly into the GLS objective to shrink statistically insignificant risk prices. While this approach encourages sparsity, it is heuristic as it relies on approximate p-values within the optimisation.

The integrated model assumes the US as global market, with regional markets (EU, Asia, Japan) entering only after orthogonalization with respect to the U.S.

$$R_{i,t} = \lambda_0 + \beta_i^{US}(R_{US,t} + \lambda_{US}) + \beta_i^{EU \perp US}(R_{EU \perp US,t} + \lambda_{EU}) \tag{3.15}$$
$$+ \beta_i^{AS \perp US}(R_{AS \perp US,t} + \lambda_{AS}) + \beta_i^{JP \perp US}(R_{JP \perp US,t} + \lambda_{JP}) + \eta_{i,t}, \quad \forall i, t$$

The segmented model places the EU market at the centre and treats other regions as external and orthogonalised with respect to Europe.

$$R_{i,t} = \delta_0 + \zeta_i^{EU}(R_{EU,t} + \delta_{EU}) + \zeta_i^{US \perp EU}(R_{US \perp EU,t} + \delta_{US}) \tag{3.16}$$
$$+ \zeta_i^{AS \perp EU}(R_{AS \perp EU,t} + \delta_{AS}) + \zeta_i^{JP \perp EU}(R_{JP \perp EU,t} + \delta_{JP}) + \nu_{i,t}, \quad \forall i, t$$

In Figures 3.11 and 3.12, we report the heatmap results for the EU portfolios according to the regularisation objective function in equation 3.12.

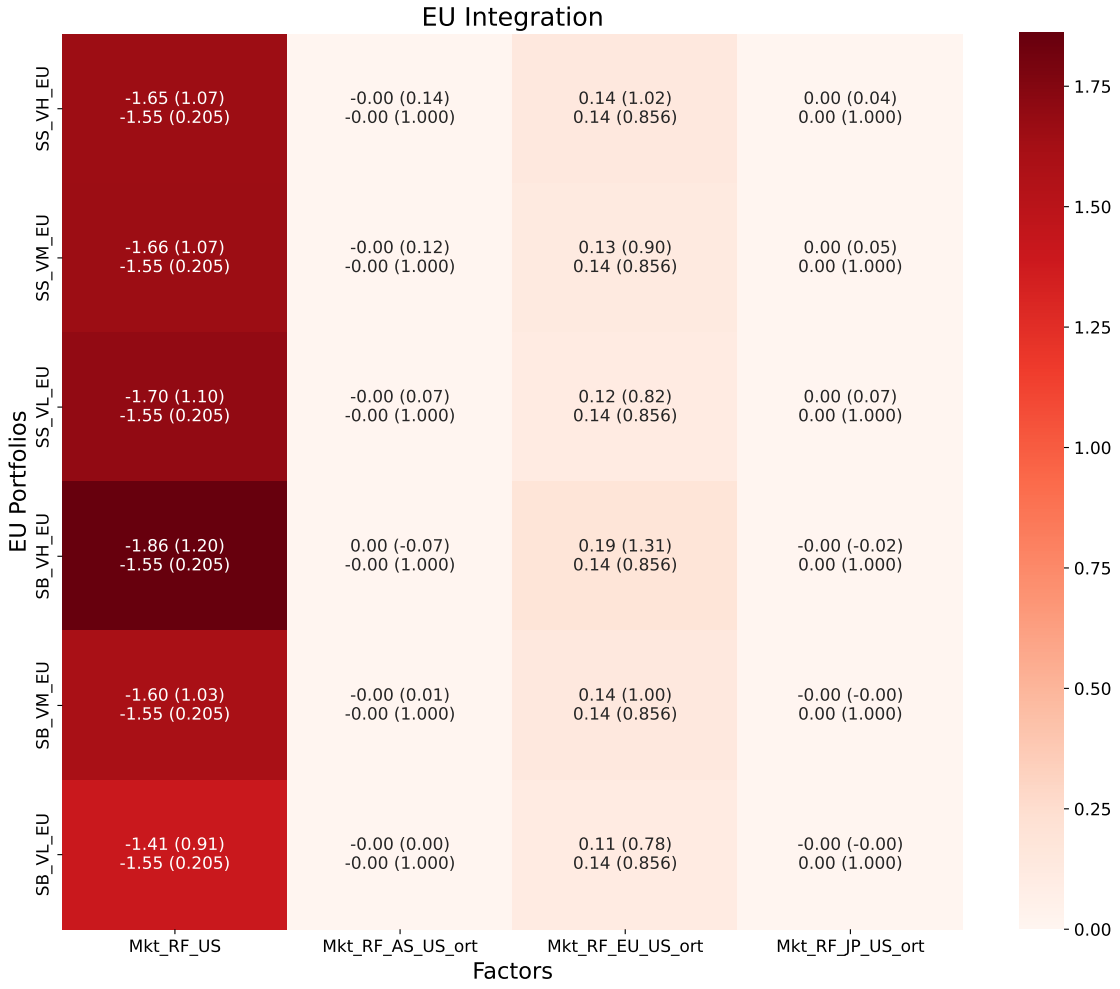

Figure 3.11: Penalised p-value Fama-MacBeth GLS, Equities Integration

The segmentation model uses European and orthogonalised foreign factors, while the integration model uses U.S. and orthogonalised non-U.S. factors. We apply adaptive Lasso with a p-value-driven penalty with $\rho_{\text{int}} = 0.0005$, $d_{\text{int}} = 5$, $\rho_{\text{seg}} = 0.0005$, $d_{\text{seg}} = 4$, p-value threshold $\tau = 0.05$, and p-value penalty weight $\rho_{\text{pval}} = 0.001$.

Each cell in the heatmaps reports four values:

- Top: the product $\beta_{i,l}\lambda_l$, i.e., the contribution of factor l to the expected return of portfolio i, with $\beta_{i,l}$ in parentheses.
- Bottom: the estimated risk price λ_l, with its GLS-based p-value in parentheses.

From the heatmaps, we observe that the only surviving (non-zero) orthogonal factors are the US

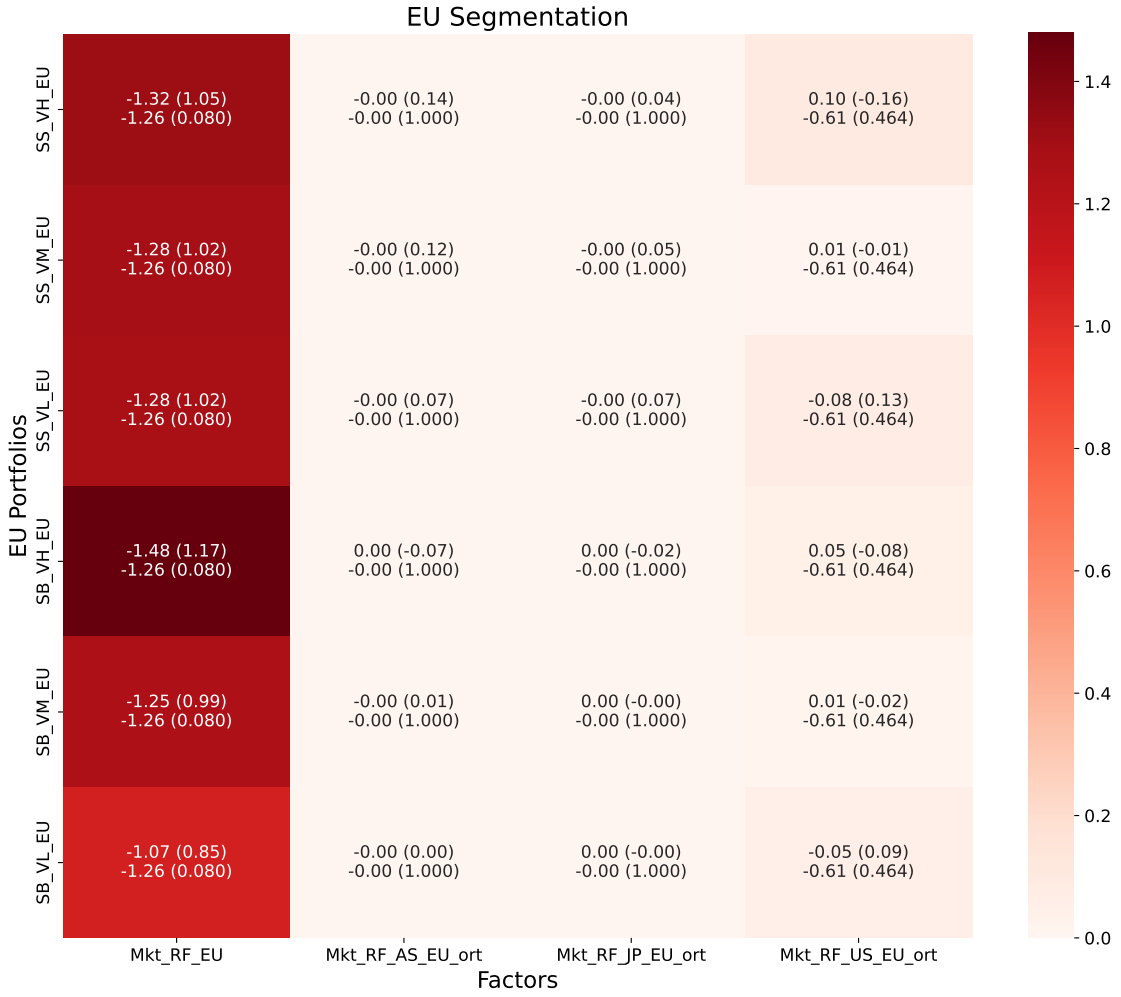

Figure 3.12: Penalised p-value Fama-MacBeth GLS, Equities Segmentation

market factor orthogonal to EU, $\text{Mkt\_RF}_{\text{US\_EU}}$, in the segmentation model, and the EU market factor orthogonal to US, $\text{Mkt\_RF}_{\text{EU\_US}}$, in the integration model. All other orthogonalised non-local factors are shrunk to (or near) zero due to the joint regularisation and p-value penalty.

Including spurious factors in the model can lead to overfitting and distorted inference: highly correlated factors may partially explain the same variation in portfolio returns, which leads to lower estimates of true λ_l values (due to diluted attribution) and inflated p-values for real factors, because the noise explained by spurious variables undermines statistical significance. This motivates a simplified model that includes only the two relevant factors, that is shown in Figure 3.13. Estimates are obtained with: $\rho_{\text{int}} = 0.00005$, $d_{\text{int}} = 4$, $\rho_{\text{seg}} = 0.00005$, $d_{\text{seg}} = 4$, $\tau = 0.05$, $\rho_{\text{pval}} = 0.0001$. We have now relaxed the penalty as we are interested in allowing non-primary components to enter the model. In the integration model there is no contribution of the EU orthogonal component, while in the segmentation model, however the EU factor is

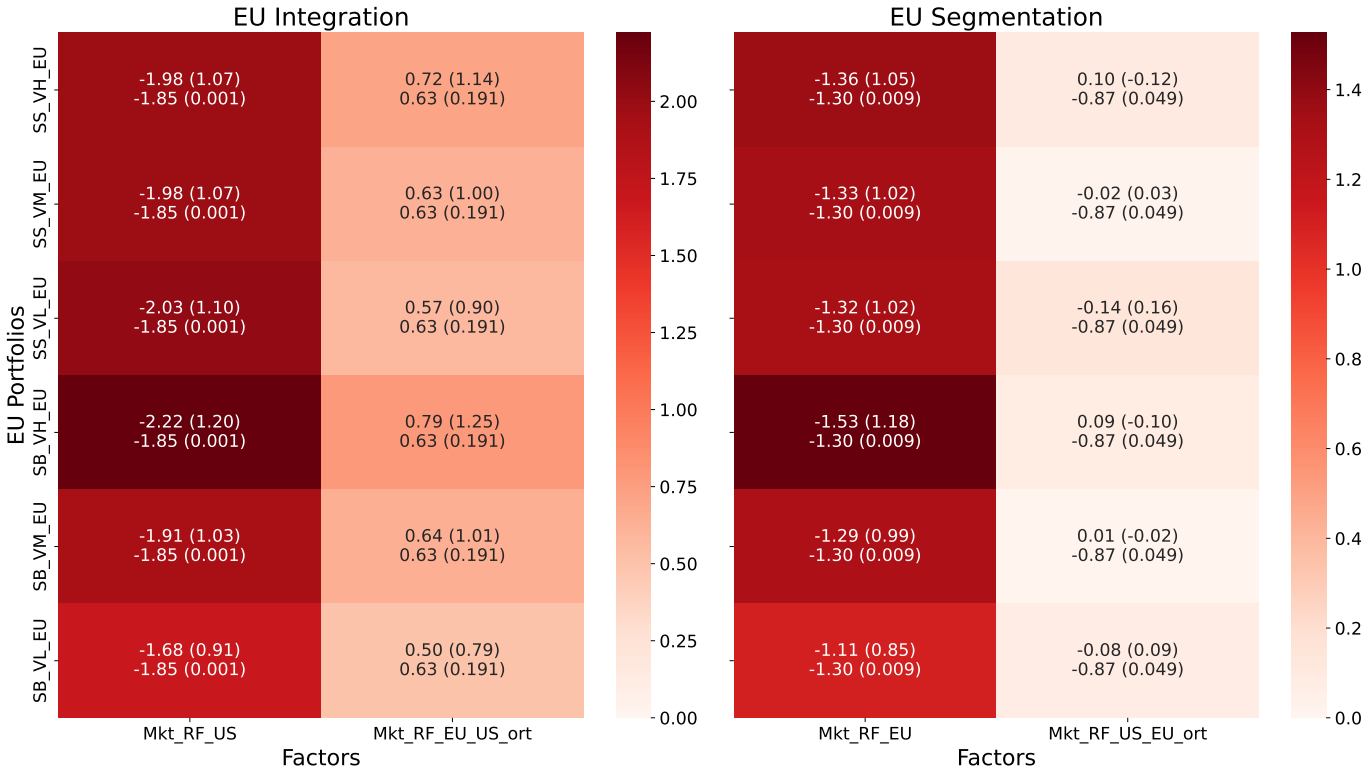

Figure 3.13: Penalised p-value Fama-MacBeth GLS, Equities Two-Factor

the main component, the US orthogonal factor is also significant, suggesting partial segmentation.

The displayed table reports the product $\lambda_l \beta_{i,l}$, while the SDF is defined as the negative of this product: a negative value in the heatmap implies a positive contribution to the SDF, and vice versa. This sign inversion is consistent with the economic role of discount factors: assets that covary positively with the SDF are priced to offer lower expected returns (i.e., they hedge in bad times), while those with negative SDF exposure require higher expected returns.

Finally, we note that the risk price estimates and p-values in the heatmaps are consistent with those reported in Table 2.16 and 2.17 for the FM GLS model.

3.4.2 Commodities

Similar to the global stock market, this section extends the copula density application to Oil-related exchange-traded funds (ETFs) and global commodity markets, namely Natural Gas (GAS), Aluminium (AL), and Soybean (SOY). We use the OIL commodity as local market. All returns are computed as excess percentage returns, using the US Coupon bond risk-free rate as benchmark.

The ETF tickers are listed below, the full description can be found in Section 2.6.2.1: 'USO', 'DBO', 'IEO', 'IXC', 'IYE', 'XLE'.

3.4.2.1 Data Processing

The analysis covers $T = 180$ monthly observations from January 2007 to December 2022. The ETFs price data are downloaded from Yahoo Finance (2025) while the commodity prices are extracted from the International Monetary Fund, IMF (2025), historical prices database. The reader should refer to Section 2.6.2.1 for more details.

3.4.2.2 One-Factor Model

We estimate a one-factor pricing model for the given ETFs using the OIL factor as the systematic risk driver. Estimation is performed via maximum likelihood copula densities and compared to standard methods such as Fama-MacBeth (FM), GLS, and GMM.

The results are shown in Table 3.10 for the centred factor. We report estimates from both univariate and multivariate copula likelihood models using Gaussian and Student-t distributions. The univariate Gaussian with a risk price of -1.64 (p-value 0.1) closely matches FM and OLS estimates, however with lower p-value. As expected, the Gauss multivariate values, -0.69 (p-value 0.41), are close to those obtained by FIML, GLS, SUR classic methods. Under multivariate copula specifications and heavy-tailed Student-t models, standard errors and estimates are comparable with the classic methods as well.

For the copula PIT we use the notation Empirical (E) and Parametric (P). Multivariate copula takes into account cross-sectional dependence, and produces parameter values comparable to FM GLS and GMM. Among standard estimators, FM GLS matches closely the Gauss multivariate risk price estimate. In the univariate case, the copula reduces to the identity and carries no extra information, so empirical and parametric margins yield identical risk price estimates.

Both Gauss univariate and Gauss multivariate start from separate (independent) marginals, but Gauss multivariate replaces the diagonal covariance with a full $\boldsymbol{\Sigma}$ from the copula, which introduces cross-sectional dependence.

These results confirm that copula-based estimation can replicate and extend standard inferences, offering robustness through flexible tail modelling and cross-sectional correlation structures. The Student-t multivariate copula, with optimized degrees of freedom, provides a more conservative estimate, consistent with the behaviour of robust estimators.

Method	Parameter	Estimate	Std. Error	t-Statistic	*p*-value
Gauss Uni	Intercept	0.9319	0.7709	1.2088	0.2267
Gauss Uni	Risk price	-1.6432	1.0076	-1.6308	0.1029
Gauss Copula Uni E, P	Intercept	0.9315	0.7707	1.2086	0.2268
Gauss Copula Uni E, P	Risk price	-1.6425	1.0073	-1.6306	0.1030
Gauss Multi	Intercept	-0.1087	0.5422	-0.2005	0.8411
Gauss Multi	Risk price	-0.6893	0.8380	-0.8226	0.4107
Gauss Copula Multi P	Intercept	-0.1087	0.5630	-0.1931	0.8469
Gauss Copula Multi P	Risk price	-0.6894	0.9505	-0.7253	0.4683
Gauss Copula Multi E	Intercept	0.0716	0.5339	0.1341	0.8933
Gauss Copula Multi E	Risk price	-1.0345	0.8924	-1.1593	0.2463
t Uni nu = 2.01	Intercept	1.0454	0.5090	2.0539	0.0400
t Uni nu = 2.01	Risk price	-1.5307	1.2469	-1.2276	0.2196
t Copula Uni E[1] P [2]	Intercept	1.0454	0.5090	2.0539	0.0400
t Copula Uni E[1] P [2]	Risk price	-1.5307	1.2469	-1.2276	0.2196
t Multi nu = 4.65	Intercept	-0.5640	0.6380	-0.8839	0.3768
t Multi nu = 4.65	Risk price	-0.1604	1.1008	-0.1457	0.8841
t Copula Multi P [3]	Intercept	-0.5338	0.5971	-0.8939	0.3714
t Copula Multi P [3]	Risk price	-0.1034	1.0772	-0.0960	0.9235
t Copula Multi E [4]	Intercept	-0.5175	0.6623	-0.7814	0.4346
t Copula Multi E [4]	Risk price	-0.1994	1.2703	-0.1570	0.8753
t Uni nu = 2.5	Intercept	1.0516	0.5283	1.9906	0.0465
t Uni nu = 2.5	Risk price	-1.5417	1.2639	-1.2198	0.2225
t Uni nu = 25	Intercept	0.8314	0.6427	1.2936	0.1958
t Uni nu = 25	Risk price	-1.3139	1.1371	-1.1555	0.2479
t Uni nu = 100	Intercept	0.6665	0.6816	0.9779	0.3281
t Uni nu = 100	Risk price	-1.2819	1.0417	-1.2305	0.2185
t Uni nu = 10000	Intercept	0.8766	0.7615	1.1512	0.2497
t Uni nu = 10000	Risk price	-1.5725	0.9972	-1.5770	0.1148
t Multi nu = 2.5	Intercept	-0.6881	0.6736	-1.0215	0.3070
t Multi nu = 2.5	Risk price	0.0230	1.1548	0.0199	0.9841
t Multi nu = 4.65	Intercept	-0.5639	0.6297	-0.8955	0.3705
t Multi nu = 4.65	Risk price	-0.1606	1.0857	-0.1479	0.8824
t Multi nu = 100	Intercept	-0.2189	0.5439	-0.4026	0.6873
t Multi nu = 100	Risk price	-0.5084	0.8924	-0.5697	0.5689
t Multi nu = 100000	Intercept	-0.1100	0.5306	-0.2073	0.8357
t Multi nu = 100000	Risk price	-0.6876	0.8284	-0.8301	0.4065
Nonlinear L1	Intercept	0.7887	0.5407	1.46	0.1447
Nonlinear L1	Risk price	-1.3216	1.0394	-1.27	0.2036
FM	Intercept	0.9242	0.8388	1.102	0.2706
FM	Risk price	-1.6276	1.3058	-1.246	0.2126
OLS	Intercept	0.9242	0.5947	1.550	0.1220
OLS	Risk price	-1.6276	1.0927	-1.490	0.1381
FIML	Intercept	-0.1068	0.5342	-0.200	0.8418
FIML	Risk price	-0.6932	0.8705	-0.796	0.4269
GLS	Intercept	-0.1203	0.5216	-0.231	0.8179
GLS	Risk price	-0.6631	0.8416	-0.788	0.4318
SUR	Intercept	-0.0803	0.5291	-0.150	0.8795
SUR	Risk price	-0.7169	0.8479	-0.850	0.3990
GMM	Intercept	-0.1945	0.5436	-0.360	0.7210
GMM	Risk price	-0.4453	0.8890	-0.500	0.6170

[1] The estimated marginal d.o.f. are 2.01, and the copula d.o.f. are 7.01

[2] The estimated marginal d.o.f. are 2.01, and the copula d.o.f. are 3.01

[3] The estimated marginal d.o.f. are 2.01, and the copula d.o.f. are 8.01

[4] The estimated marginal d.o.f. are 2.01, and the copula d.o.f. are 6.01

Table 3.10: One-Factor Risk Price Estimates, ETFs Copula, Centred Factor

In the empirical Gaussian copula case, the results tend to lie between the univariate (OLS) benchmark and the multivariate GLS benchmark, due to the following reasons:

- Ranks Correlation matrix vs. residuals. In the parametric Gaussian case, the correlation matrix $\hat{\rho}$ is computed from the standardised residuals, matching the GLS benchmark exactly. In the empirical case, residuals are first transformed to ranks and then Gaussian inverse CDF is applied $Z = \Phi^{-1}(u)$, where u is the empirical PIT. This produces a Spearman-type correlation, which is generally smaller in magnitude than the Pearson correlation (see Section 3.2.3, Table 3.5), pulling GLS estimates closer to the univariate values.

- Weaker GLS weighting. GLS uses the weight matrix $\boldsymbol{\Sigma}^{-1}$. When $\boldsymbol{\Sigma}$ is based on rank correlations, it is typically closer to the identity matrix than the Pearson-based $\boldsymbol{\Sigma}$, due to the inverse scaling factor depending on the square of partial correlations, resulting in weaker GLS adjustments and estimates that lie between OLS and GLS.

- Difference with the t-copula case. In the t-copula, even when empirical ranks are used for PIT, the tail-dependence structure inflates the estimated correlations toward Pearson-like values, especially for small degrees of freedom ν. This explains why the empirical t-copula estimates remain close to the multivariate benchmark, whereas the Gaussian empirical estimates do not.

Notably, the Univariate Student-t with fixed $\nu = 25$ delivers risk price estimates close to those obtained via L1 minimization, supporting the interpretation of heavy-tailed likelihoods as robust alternatives to Gaussian loss functions.

3.4.2.3 Integration and Segmentation

In this section, we use the ETFs together with the mimicking portfolio factors that were defined in Section 2.6.2.5 and defined in equation 2.88.

We report the results for the integration model of Oil ETFs and Gas as global factor and Crude Oil orthogonal component as local factor, and the corresponding segmentation with Oil as local factor and Gas orthogonal component as global factor.

Integration

For the integration model, refer to equation 2.82, the dependence between mimicked market factor and ETFs excess return is visibly decreased in the constructed stochastic discount factor compared with the sum of the factors. Being only the Gas factor significant, we set to zero the risk price of the Oil orthogonal component. While the estimated Pearson correlations between

the raw sum of factors and ETFs remain moderate to high (ranging from approximately $+0.5$ to $+0.88$), the correlations between the SDF and ETFs are weaker, ranging from about 0.4 to 0.67 (Table 3.11).

The Gaussian copula correlations in Table 3.12 confirm this pattern, with $\rho_{\text{C,SDF}}$ values uniformly lower than $\rho_{\text{C,sum}}$. Despite the strong individual factor exposures, the estimated SDF effectively balances them via the cross-sectional GLS risk prices, such that their combined pricing kernel exhibits lower residual dependence with the ETFs excess returns.

On average, the R^2 of the first step regression is 0.57 (USO, $R^2 = 0.97$, and DBO $R^2 = 0.89$), the ETFs exhibit weaker average exposure ($\bar{\beta}^{GAS} = -0.07$) to the Gas factor and, as expected, stronger exposure to Oil Gas orthogonal factor ($\bar{\beta}^{OIL \perp GAS} = 0.84$), which is however balanced by the prices of risk: $\lambda_{\text{GAS}} = -2.14$ (p = 0.0014) and $\lambda_{OIL} = -1.22$ (p = 0.07).

ETF	ρ_{sum}	ρ_{SDF}
USO	$+0.843$	$+0.575$
DBO	$+0.883$	$+0.679$
IEO	$+0.514$	$+0.386$
IXC	$+0.670$	$+0.542$
IYE	$+0.600$	$+0.470$
XLE	$+0.574$	$+0.428$

Table 3.11: Integration Correlations, Mimicked Factors and Excess returns

ETF	$\rho_{C,\text{sum}}$	$\rho_{C,\text{SDF}}$
USO	$+0.822$	$+0.571$
DBO	$+0.856$	$+0.644$
IEO	$+0.554$	$+0.423$
IXC	$+0.684$	$+0.565$
IYE	$+0.645$	$+0.522$
XLE	$+0.624$	$+0.487$

Table 3.12: Integration Copula Correlations, Mimicked Factors and Excess returns

Figures 3.14 and 3.15 display the scatter plots of the Gaussian scores (PIT) used in the copula estimation.

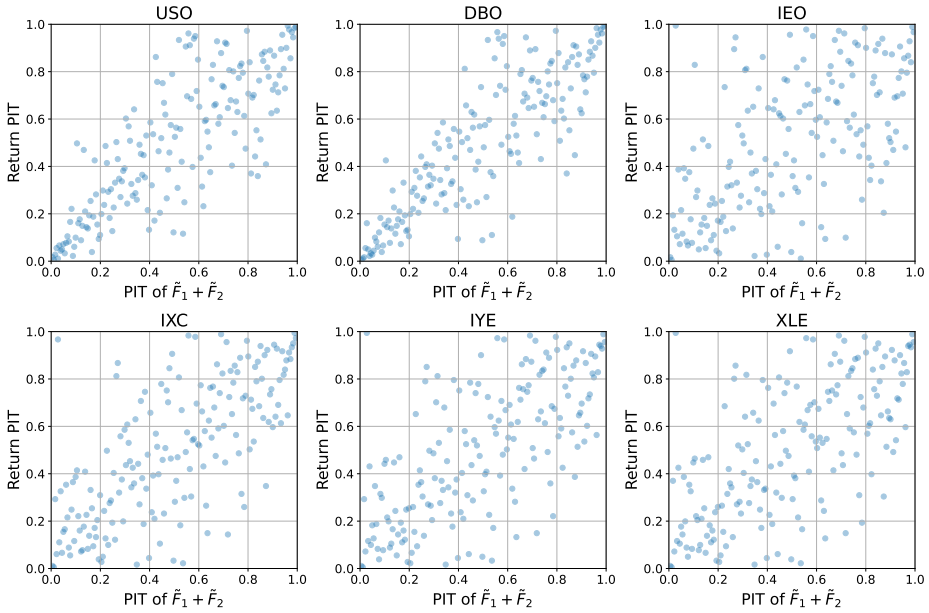

Figure 3.14: Integration PIT, Factors and Returns, ETFs

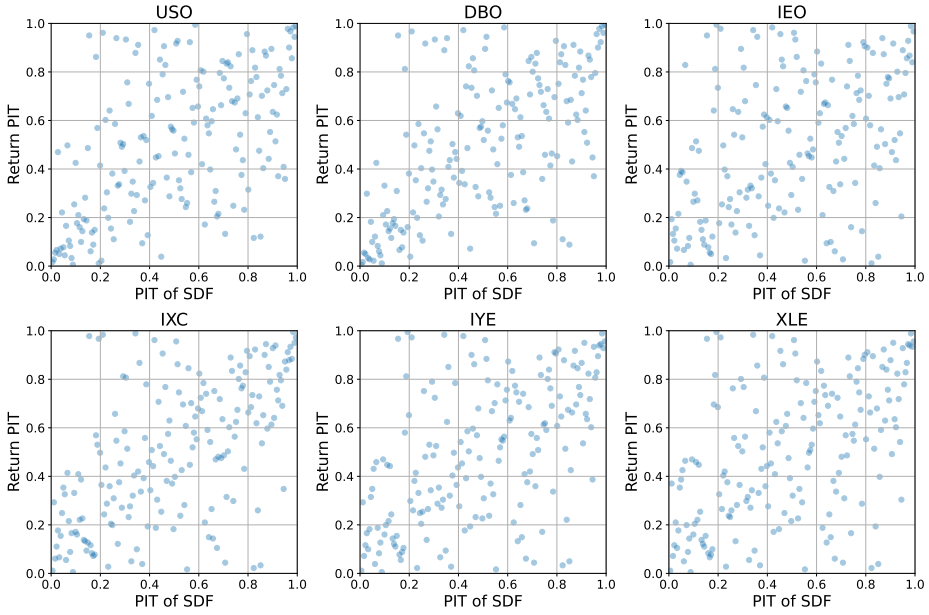

Figure 3.15: Integration PIT, SDF and Returns, ETFs

The PIT scatter plots confirm that the ranks of the SDF are more dispersed and less diagonally aligned with the mimicked factor ranks than the raw factor sum, revealing how the SDF pricing kernel reflects real statistical dependence through estimated λ reweighting.

Segmentation

The segmentation results, refer to equation 2.83, are shown below. On average, the R^2 of the first step regression is 0.56, the estimated SDF reflects a strong positive exposure ($\bar{\beta}^{OIL} = 0.79$) to the OIL factor and a weaker exposure to the Gas Oil orthogonal component ($\bar{\beta}^{GAS \perp OIL} = -0.034$) ; however, this is compensated by the cross-sectional GLS estimates of the risk prices, with $\delta_{\text{OIL}} = -1.42$ (p = 0.052) and $\delta_{GAS} = -2.1$ (p = 0.004), such that the combined pricing kernel neutralizes much of the return co-movement. However as the p-value is not significant, we set the risk price of the Oil component to zero.

In Tables 3.13 and 3.14, we report the segmentation correlation values. After we set the risk price of the Oil factor to zero, the dependence structure breaks down: the sum of the two-factor components shows correlations with ETFs excess returns (ranging from 0.56 to $+0.87$), while the correlation between the constructed SDF and the ETFs returns is weaker (approximately 0.37 to 0.62 across both Pearson and copula measures).

ETF	ρ_{sum}	ρ_{SDF}
USO	$+0.815$	$+0.488$
DBO	$+0.866$	$+0.614$
IEO	$+0.508$	$+0.371$
IXC	$+0.669$	$+0.535$
IYE	$+0.597$	$+0.462$
XLE	$+0.569$	$+0.420$

Table 3.13: Segmentation Correlations, Mimicked Factors and Excess Returns

ETF	$\rho_{C,\text{sum}}$	$\rho_{C,\text{SDF}}$
USO	$+0.796$	$+0.495$
DBO	$+0.835$	$+0.583$
IEO	$+0.543$	$+0.409$
IXC	$+0.679$	$+0.562$
IYE	$+0.638$	$+0.519$
XLE	$+0.615$	$+0.482$

Table 3.14: Segmentation Copula Correlations, Mimicked Factors and Excess Returns

Figures 3.16 and 3.17 show the scatter plots of the Gaussian PIT scores used in the copula estimation. The PITs for the raw sum of factors do show a stronger dependence along the diagonal with the return PITs, due to spurious Oil factor interference, while the PITs for the SDF show a weaker dependence as we manually removed the effect of the Oil factor, which is not

priced. This visual pattern reinforces the numerical findings: the SDF ranks co-move lightly with Gas Excess return ranks, indicating integrated risk exposures in the pricing kernel.

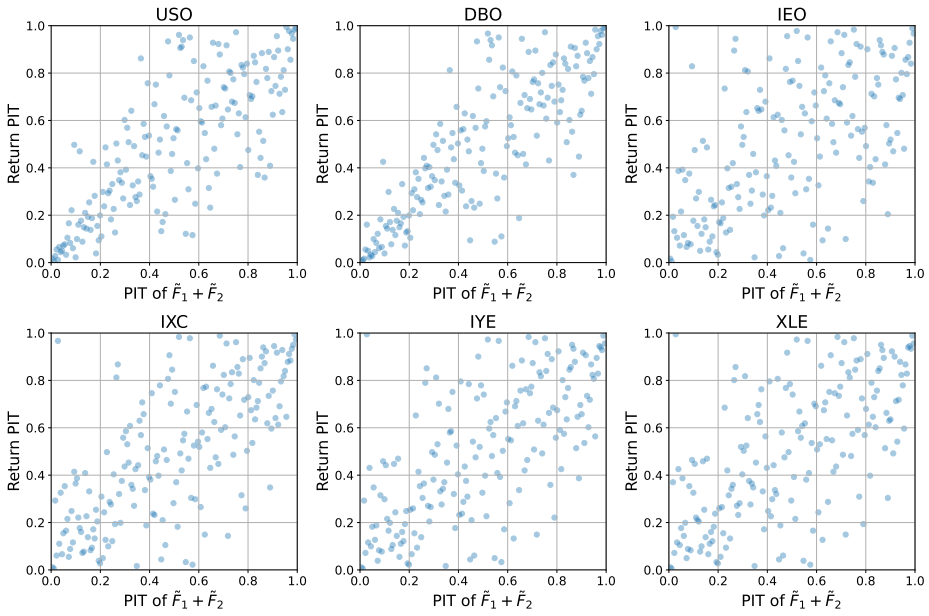

Figure 3.16: Segmentation PIT, Factors and Returns, ETFs

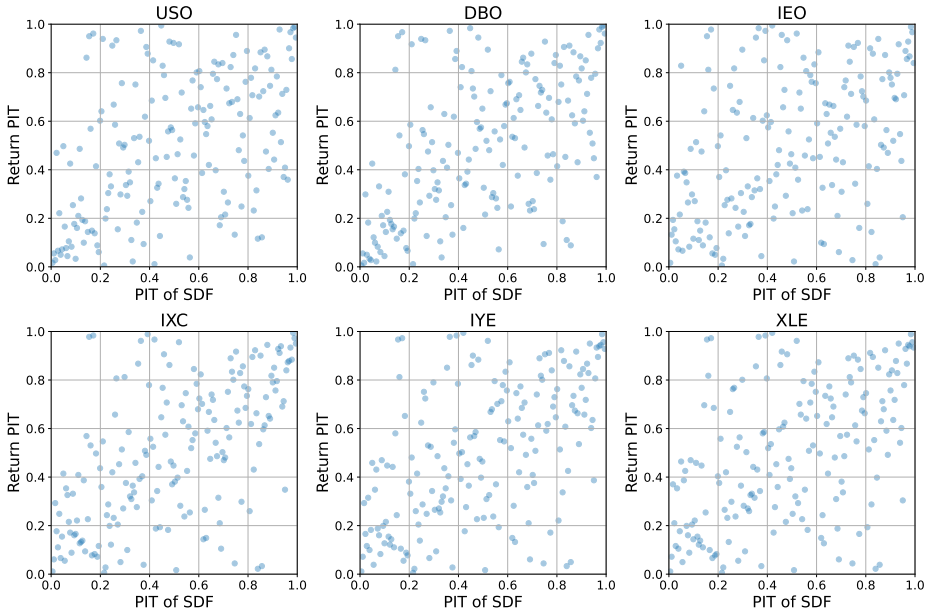

Figure 3.17: Segmentation PIT, SDF and Returns, ETFs

3.4.2.4 Regularisation

We apply penalised regression methods—Lasso, Ridge, and Elastic Net—as described in equations 3.8, 3.9, 3.10 to commodity indices and OIL ETF excess returns: Figures 3.18, 3.19, 3.20. The analysis is limited to the OIL ETFs, using mimicked factors for the commodity indices: Aluminium, Gas, Oil, Soybean. We start by running time series regularisation using the same penalty level ($\rho = 0.0001$). The resulting factor loading coefficients (exposures) are visualized through heatmaps.

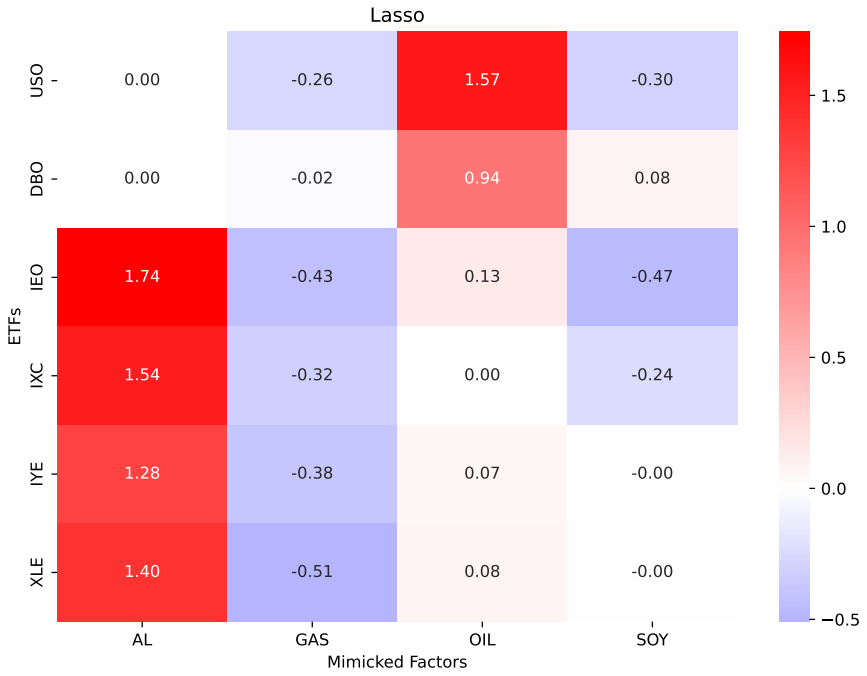

Figure 3.18: Lasso Loadings Heatmap, ETFs

Darker colors represent stronger factor exposures; white areas indicate coefficients shrunk to zero due to regularisation. Lasso produces sparse solutions by performing factor selection.

In Figure 3.18, we can see that Lasso mostly captures Oil-related exposures and filters out weaker signals: DBO and USO with AL, IYE and XLE with SOY. In contrast, Ridge regularisation, Figure 3.19, retains most of the factor loadings but shrinks them proportionally. This results in a more diffuse exposure profile where all factors contribute marginally. Balanced Elastic Net regularisation, $\alpha = 0.5$, combines Lasso's sparsity and Ridge's stability, filtering out irrelevant factors while preserving structure among correlated predictors. In Figure 3.20, we can see that Oil factor is dominant but still appreciate partial influence from Aluminium and Gas.

The exposure heatmaps, as already noticed for global economies, have two important limitations:

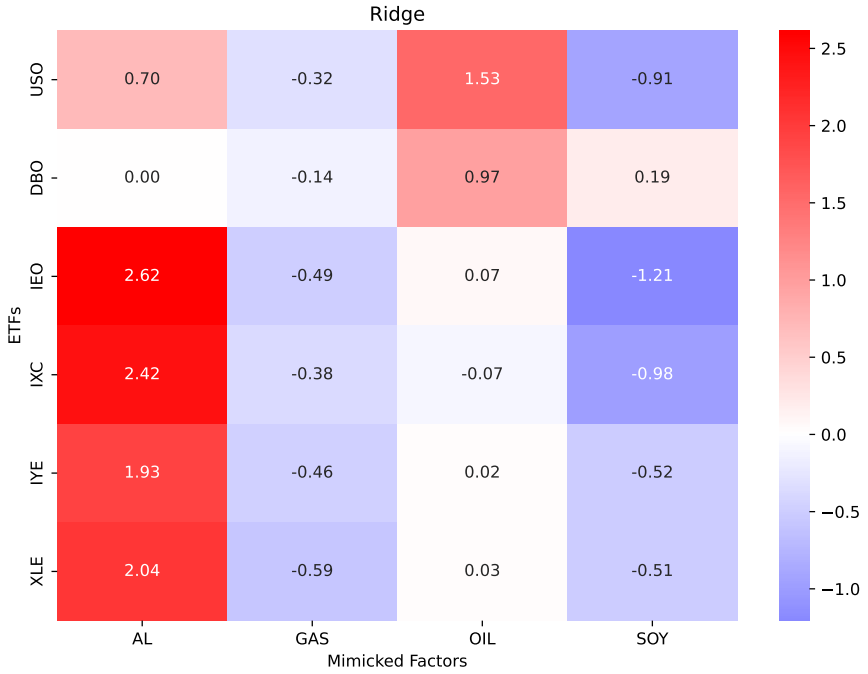

Figure 3.19: Ridge Loadings Heatmap, ETFs

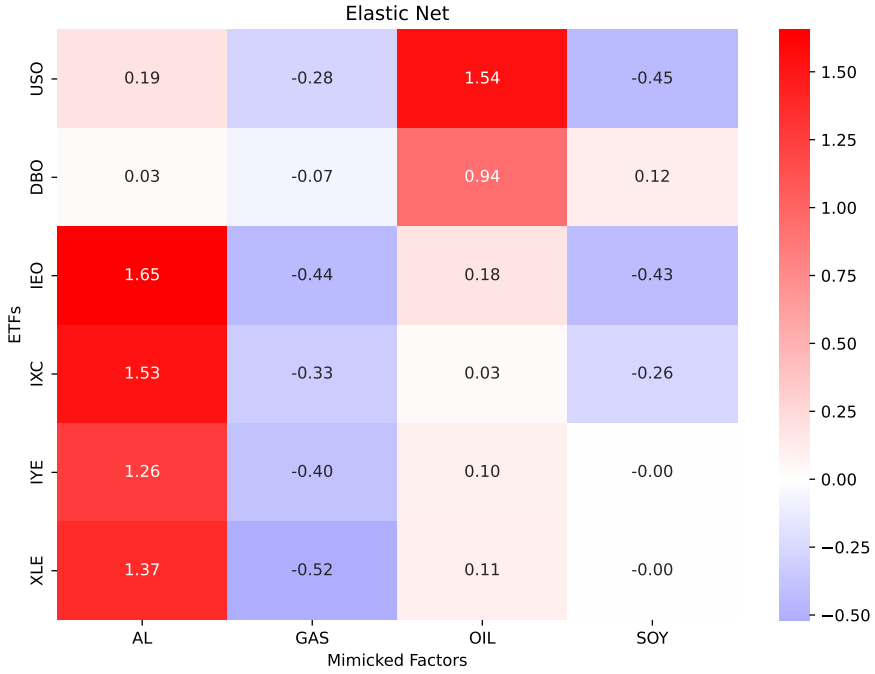

Figure 3.20: Elastic Net Loadings Heatmap, ETFs

they ignore collinearity among factors and do not consider that risk price is not the same for all the factors.

To incorporate cross-sectional risk prices (λ_l) and link exposures more explicitly to asset returns, we deploy the Lasso p-value penalised model. We also introduce orthogonal factors to address collinearity, and to account for marginal contributions. We regularise the Fama–MacBeth second step using a Lasso penalty on the risk prices, with adaptive weights derived from the betas. In addition, we impose a soft constraint that shrinks factors with a high p-value (factors above a threshold p-value will be considered spurious), see equation 3.12.

The integration model assumes the GAS factor as the global source of risk, while OIL, AL, and SOY are introduced only for their marginal contribution, using their GAS orthogonal component:

$$\begin{aligned} R_{i,t} = \lambda_0 + \beta_i^{GAS}(f_{GAS,t} + \lambda_{GAS}) + \beta_i^{OIL\perp GAS}(f_{OIL\perp GAS,t} + \lambda_{OIL}) \\ + \beta_i^{AL\perp GAS}(f_{AL\perp GAS,t} + \lambda_{AL}) + \beta_i^{SOY\perp GAS}(f_{SOY\perp GAS,t} + \lambda_{SOY}) + \eta_{i,t}, \quad \forall i,t \end{aligned} \tag{3.17}$$

where $f_{l,t}$ is the mimicked factor l at time t.

The segmentation model treats OIL as the local reference market, and other commodity factors are orthogonalised with respect to it:

$$\begin{aligned} R_{i,t} = \delta_0 + \zeta_i^{OIL}(f_{OIL,t} + \delta_{OIL}) + \zeta_i^{AL\perp OIL}(f_{AL\perp OIL,t} + \delta_{AL}) \\ + \zeta_i^{GAS\perp OIL}(f_{GAS\perp OIL,t} + \delta_{GAS}) + \zeta_i^{SOY\perp OIL}(f_{SOY\perp OIL,t} + \delta_{SOY}) + \nu_{i,t}, \quad \forall i,t \end{aligned} \tag{3.18}$$

Figures 3.21 and 3.22 show the multifactor heatmap results, with the following regularisation setting: $\rho_{\text{int}} = 0.00005$, $d_{\text{int}} = 5$, $\rho_{\text{seg}} = 0.00005$, $d_{\text{seg}} = 4$, p-value threshold $\tau = 0.05$, and p-value penalty weight $\rho_{\text{pval}} = 0.0005$.

Each heatmap cell contains:

- Top: the product $\beta_{i,l}\lambda_l$ of Excess Return factor l vs. ETFs i's expected return, with the beta $\beta_{i,l}$ in parentheses.
- Bottom: the estimated risk price λ_l, with the corresponding GLS p-value in parentheses.

The choice of the regularisation penalty in Lasso estimation is not straightforward, especially in settings with orthogonal but weak factors where Lasso can yield unstable results.

We opt to exclude Soybean, which is not priced, and Aluminium, which has higher betas but

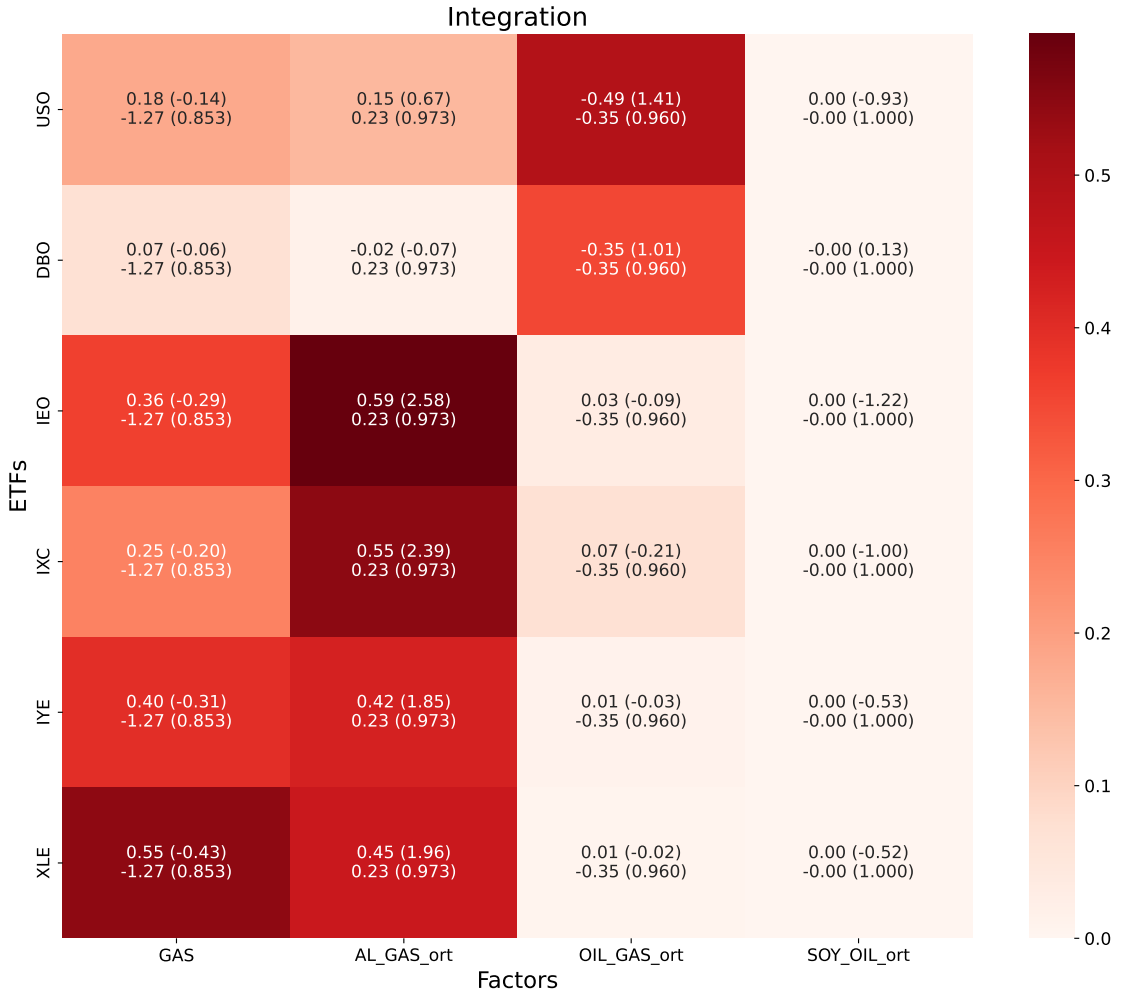

Figure 3.21: Penalised p-value Fama-MacBeth GLS, ETFs Integration

shows a lower p-value in the dual factor model (refer to tables 2.28, and 2.29). However both integration and segmentation models with four factors are unstable due to the high beta collinearity in the second step of the Fama-Macbeth regression. In this setting, we also tried to use a backward selection, but factor shrinkage was not successful. By contrast, a forward selection model, starting from GAS for the integration and from OIL for the segmentation, consistently selected OIL_GAS_ort and GAS_OIL_ort, respectively. The numerical results of these experiments are available upon request.

For this reason, we retain GAS and OIL_GAS_ort in the integration model, and OIL and GAS_OIL_ort in the segmentation model. The simplified two-factor specification is shown in Figure 3.23. The reduced model was run relaxing the penalty as we are interested to allow non-primary components to enter the model: $\rho_{\text{int}} = 0.00005$, $d_{\text{int}} = 5$, $\rho_{\text{seg}} = 0.00005$, $d_{\text{seg}} = 4$, p-value threshold $\tau = 0.0005$, and p-value penalty weight $\rho_{\text{pval}} = 0.0005$.

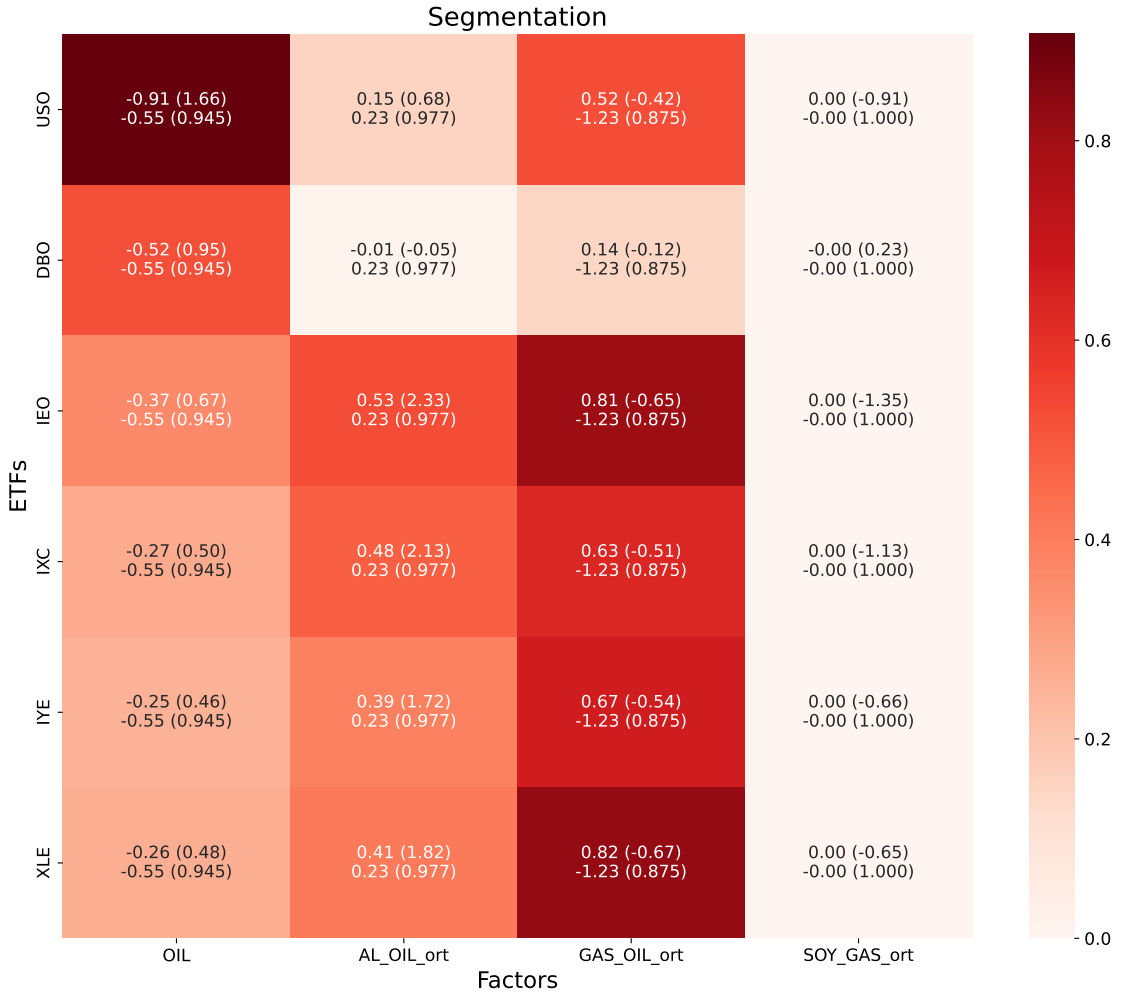

Figure 3.22: Penalised p-value Fama-MacBeth GLS, ETFs Segmentation

The estimated risk prices and p-values shown in the heatmaps are consistent with the values reported in Tables 2.28 and 2.29. The regularisation integration heatmap shows that the GAS factor, although it has lower betas, is priced. In the segmentation model, the orthogonal GAS factor is also priced while the OIL factor is weakly priced. This evidence supports total integration between the Gas and Oil commodities.

The sign of the estimated products $\lambda_f \beta_{i,f}$ implies opposite signs for the stochastic discount factor contributions, as we have already noticed for global economies.

We can notice that in the presence of weak factors and beta collinearity the Lasso regularisation becomes unstable, and highly sensitive to the choice of regularisation parameters.

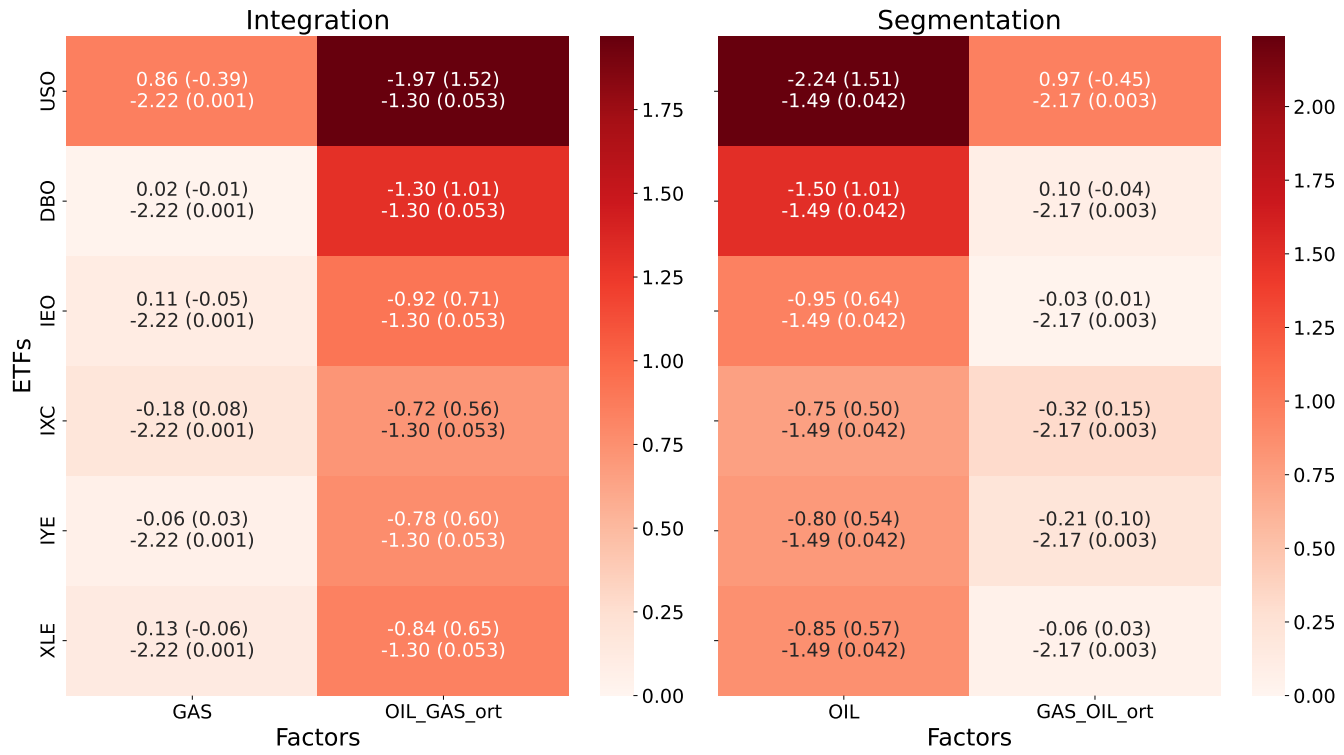

Figure 3.23: Penalised p-value Fama-MacBeth GLS, ETFs Two-Factor

3.5 Practical Implications for Investors

This section addresses the practical implications for portfolio managers, institutional investors and commodity traders, particularly regarding the understanding of market integration, tail dependence, and risk transmission mechanisms.

3.5.1 Risk Transmission

In classical portfolio analysis, the focus is on the first stage time-series regression introduced in equation 2.7. Beta captures the exposure to the market, and alpha captures the idiosyncratic component associated with active management.

Multifactor models, such as the Fama–French framework, introduce additional factors beyond the market—such as size (SMB,Small Minus Big), value (HML, High Minus Low), profitability (RMW,Robust Minus Weak), and investment (CMA,Conservative Minus Aggressive)—challenging the interpretation of portfolio managers' α as genuine skill.

According to our study, in multifactor models the relevance of a factor should be assessed not

only through time-series fit (adjusted R^2) but also by the significance of its estimated risk price λ in the cross-sectional regression, equation 2.8. A factor l with $\lambda_l \approx 0$ is considered spurious, even if portfolios show high or statistically significant loadings ($\beta_{i,l}$) on it in the time-series regressions.

In addition, classical linear estimation of λ and β assumes normally distributed returns. These assumptions fail under heavy tails and nonlinear dependence, precisely when risk transmission becomes most important. During the 2007–2008 financial crisis, for example, collateralized debt obligations (CDOs) were priced using the Gaussian copula, which implied independence in the tails of joint default probabilities, and led to a systematic underestimation of the probability of joint extreme losses.

Estimating risk premia with Student-t copula-based maximum likelihood methods, as proposed in this study, provides a more realistic measure of priced risk, which enables the identification of non-spurious factors even under heavy-tailed distributions, capturing the higher probability of simultaneous losses under stress conditions. This ensures that, in periods of market stress, the inferred risk prices λ remain valid and informative for portfolio construction, and systemic risk assessment.

The idea is to identify the factors that drive performance under different volatility regimes, such as crisis and non-crisis periods: the investors can use λ as a signal for factor-based stock picking, allocating more weight to securities with high risk premium (high $\beta_{il}\lambda_l$).

Furthermore, λ can be integrated directly into portfolio construction if the expected returns are modelled as $\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\lambda}$, where $\mathbf{B} \in \mathbb{R}^{N\times\ell}$ collects the factor loadings for N portfolios and ℓ factors, and $\boldsymbol{\lambda} \in \mathbb{R}^{\ell\times 1}$ is the vector of corresponding risk prices. While the classical mean–variance optimization relies on historical sample means $\hat{\boldsymbol{\mu}}$, this approach derives expected returns from economically priced risk factors. When $\boldsymbol{\lambda}$ and $\mathbf{B}$ are estimated via the copula-based MLE method proposed in this study, tail dependence and nonlinear co-movements are taken into account, making the optimization robust to fat tails and high volatility dynamics.

3.5.2 Integration

In commodity trading, the approach extends naturally to the classical pricing relation between the Forward (F_t) and the Spot Price (S_t), $F_t = S_t\, e^{(r_t + c_t - y_t)\tau}$, where r_t is the interest rate, c_t the cost of carry, y_t the convenience yield, and $\tau = T - t$ is the time to maturity, with t denoting the current time (when the forward is priced) and T the maturity or delivery date of the forward contract. The structural pricing model can be expressed in a equivalent multi-factor

framework. The interest rate, cost of carry, and convenience yield play roles analogous to the market, size, and value factors in equity models, acting as systematic drivers of expected yields and term structure dynamics through their associated risk premia, see Ballestra and Tezza (2025) for a continuous-time commodity model where the spot price, convenience yield, and interest rate follow a system of stochastic differential equations (SDE).

On one hand, integration, as defined by Brooks and Iorio (2009) and applied in this study, represents the economic linkage among the relevant factors through a shared pricing kernel that explains why prices move together. On the other hand, cointegration provides statistical evidence of a shared long-run trend, confirming price comovements. Taken together, integrated markets should exhibit cointegration, allowing practitioners to identify benchmarks, redundant prices references, and mean-reverting spreads that can be exploited through pairs trading strategies.

Integration is typically stronger at the micro level—for example, between two futures that reference the same product specification, delivery point, and incoterm—where arbitrage and price discovery operate efficiently. For example, the ICE and CME ULSD 10 ppm CIF Mediterranean Platts futures contracts both settle to the same price assessment, so if they are economically integrated, their cointegration should be stable. The stationary spread between them would then represent only short-term noise or liquidity differences. In such a case, either future series can serve as the representative benchmark for hedging Mediterranean diesel, as both embed the same market information. However, a breakdown in cointegration would signal partial segmentation between contracts.

While cointegration analysis is purely statistical and detects the presence of a stable long-run relation between price series, it does not explain why the relation exists or what economic factors drive it. Market integration, by contrast, provides the underlying mechanism: it identifies the common factors that generate the co-movement. Here comes the importance of the SDF copula analysis and the p-value penalised Fama-MacBeth GLS Lasso regularisation proposed in this study: they potentially enable the application of trading strategies (for example, the classic pair-trading mean-reversion approach based on cointegration) through a guided selection of factors that effectively drive risk pricing. The trader will use the integration signal to decide whether and how much to rely on a cointegration strategy, and at the same time, the cointegration signal will be used to time the trades and manage the spread.

3.6 Contribution

In this Chapter we presented two applications of copula: one for risk price estimation and another, based on the dependence between SDF and returns, that we apply to market integration analysis.

The risk price estimation is a novel application of copula-based maximum likelihood optimisation of the one-factor asset pricing model. While traditional methods such as Fama-MacBeth, GMM, and SUR rely on linear covariance structures, the copula framework allows explicit modelling of nonlinear and tail dependencies in residuals. In particular, our results show that the multivariate copula density likelihood recovers cross-sectional GLS estimates, while the univariate copula likelihood recovers OLS estimates under the assumptions of homoscedasticity and no cross-sectional correlation. Although computationally more intensive, the copula method can be extended to more flexible specifications such as the Student-t copula, which provides robustness to heavy tails, and to dynamic copulas capturing time-varying dependence.

The second application, the copula correlation between the SDF and returns, helps to differentiate the contribution of joint dependence from the contribution of risk prices. We show that in a multifactor model, the sum of the factor realizations depends only on exposures and explains return variability regardless of whether factors are priced. By contrast, the SDF-return copula correlation depends directly on which factors are priced in the model: setting a factor's risk price to zero, removes its impact on the SDF while it does still contribute to return variation. This distinction illustrates the essential difference between exposure and pricing in factor models:

- Exposure drives variation in realised returns through the factor loadings β_l. Even if the factor's risk price is zero, $\lambda_l = 0$, fluctuations in f_l induce fluctuations in returns.
- Risk prices determine the compensation for risk in expected returns. Only nonzero λ_l affect the expected returns and the SDF.

The SDF copula analysis helps to visualize whether strong dependence arises primarily from factor exposure or from pricing. If a factor is unpriced, the SDF and returns will show weaker dependence in the copula analysis, even though the factor still drives return variation. In contrast, when risk prices are significant, the SDF systematically discounts returns, creating stronger dependence. SDF copula helps to identify scenarios where dependence is driven by exposure alone rather than priced risk, highlighting potential spurious factors that would otherwise appear significant in variance-based analyses.

On the other hand, this approach is not intended to replace classic risk price estimation. The

sign and magnitude of λ, which are key inputs for the SDF copula, determines whether a factor is perceived as a good or a bad risk and how much compensation investors require per unit exposure. Rather, SDF copula complements risk-price estimation by providing a rank-based view of dependence that is separate from the marginal distributions and invariant to scaling and other monotone transformations. For example, a Student-t copulas can accommodate heavy tails and tail dependence, which are common in financial data. This extension is particularly useful if the data are non-Gaussian or exhibit extreme co-movements during periods of high volatility.

Finally, in the regularisation experiment, we test a Penalised p-value Fama-MacBeth GLS Lasso model, which provides several advantages over other regularisation methods, including the FM Lasso penalised beta estimation of risk prices introduced by Bryzgalova (2015):

- ensuring that retained factors contribute not only to fit, but also to risk pricing.
- accounting for collinearity both via full GLS weighting (cross-sectional covariance of residuals) and via prior orthogonalization of factors, which isolates the marginal contribution of each factor-specific component.
- penalising factors with high p-values, avoiding false positives and overfitting.

Unlike other approaches, this method regularises the pricing kernel directly. Factors that lack significance or explanatory power are penalised and removed, while priced and relevant sources of risk are preserved. However, in the presence of weak factors and beta collinearity, as it is the case of commodities indices, the regularisation method becomes very unstable.

Chapter 4

Conclusions and Future Research

In this chapter, we present the conclusions and summarize the main contributions of this work. For a detailed list of contributions we refer to the respective sections:
- Risk price estimation and linear approximation, Section 2.7.
- Dependence, Copula, and Regularisation, Section 3.6.

4.1 Conclusions

The risk premium is the product of two elements: the risk price and the beta exposure. A factor is priced only when both components are non-zero. In this framework, the integration model can be seen as a binary application: integration requires that only the global factor is priced, meaning it exhibits both high exposure and a significant risk price. Segmentation is defined as the symmetric case, in which the local factor is priced while the global factor is not. However, the segmentation model is not intended to verify the same economic mechanism as integration, but rather to provide a complementary definition by construction. The SDF Copula analysis developed in the final chapter is particularly useful for visualizing the product factor contribution, as it highlights the risk premium as a whole rather than separating exposure and risk price components, which would otherwise appear as a simple sum of factors exposure. A major outcome of this thesis is the clarification of the concept of integration as originally applied in Brooks and Iorio (2009). As explained above, this idea extends beyond the integration of global economies and commodity indices, and addresses the estimation of risk premia and the representation of their bilinear structure through the SDF Copula analysis. More generally, the concept of integration serves as a starting point for clarifying how true factors should be distinguished from spurious ones. This is effectively achieved through risk premium estimation, which combines exposures and risk prices, and through the analysis of their bilinear product via the SDF Copula. In the final chapter, we also propose a heuristic Fama–MacBeth GLS Lasso regularisation with penalties on both betas and p-values, representing a first attempt to implement this idea within a multifactor model.

Another contribution is the systematic approach to the linear approximation of the product factor,

namely the risk premium. We show how the estimates and inference of the linear approximation methods are consistent with the results obtained with more classic estimation techniques. Comparing solver performance, we see that the sum of squared residuals (SSR) naturally favours L2-norm methods and the sum of absolute errors (SAE) favours L1-norm methods. However, we find that in the presence of heavy tails, the L1-norm approaches show lower bias and variance for the estimation of the unknown parameters than the L2-norm approaches.

Finally, the risk price estimation presented in Chapter 3 introduces a novel application of copula-based maximum likelihood optimisation to the one-factor asset pricing model. Our results demonstrate that, when errors are homoskedastic and cross-sectional correlation is absent, the univariate copula likelihood yields the same estimates as OLS, whereas the multivariate copula density likelihood matches the GLS estimates. The copula formulation therefore provides a natural bridge between linear correlation-based methods and dependence modelling based on transformed ranks.

4.2 Future Research

The findings of this thesis open up several directions for future research that may prove valuable for financial modelling. In particular, the main applications are expected to lie in the areas of factor selection and portfolio optimisation via the factor mimicking portfolio method.

- Integration and segmentation. The methodology described in Section 2.6.1.4 can be applied to study the integration and segmentation of European economies before and after the introduction of the Euro. Another natural application would be to analyse the degree of integration of the UK economy with the EU, the US, and the Asia-Pacific region in the periods before and after Brexit. The main challenge for this line of research lies in the construction of appropriate local portfolios and market factors, as well as their standardisation across countries, which is essential to ensure a consistent assessment of changes in financial integration over time. For both studies, we recommend the use of the SDF copula analysis as developed in Section 3.2.3, to help better visualise the integration results.

- Linear approximation. The linear approximation methods proposed in Section 2.3 can be easily extended to two-factor and more complex multifactor models, in order to compare bias, variance, and performance under both the L1- and L2-norm dimensions.

- Copula density likelihood estimation. The method proposed in Section 3.2.2 can be extended to multiple factor model and to the use of non-elliptical copulas with empirical PIT. A promising direction for future research is to simulate the robustness of copula-based likelihood estimation

relative to standard correlation-based methods in environments with time-varying volatility and tail dependence. This would help assess the potential gains of copula likelihood methods in realistic crisis scenarios, where standard linear factor models may fail to provide an adequate description of joint risks.

- Regularisation. An interesting direction for future research is to extend the penalised p-value Fama-MacBeth Generalized Least Squares (GLS) Lasso regularisation introduced in this thesis. Our approach improves upon standard methods by ensuring that retained factors contribute not only to statistical fit but also to risk pricing, through direct regularisation of the stochastic discount factor. The current implementation is heuristic: p-values are computed ex post and incorporated as a soft penalty in the objective function.

A promising extension of this work is the use of the regularised GMM framework developed by Belloni et al. (2018). In this context, the estimation involves a low-dimensional structural parameter of interest (such as the vector of risk prices λ) together with high-dimensional parameters (such as the portfolio exposures β_i). When the β_i are estimated with shrinkage methods (e.g. Lasso), the resulting estimators are biased and direct plug-in approaches yield invalid inference for λ. The idea is to construct a modified score function that is locally insensitive to errors in the high-dimensional estimates. Applied to factor pricing, this approach would allow regularisation to be used in the time-series stage for exposure estimation, while maintaining valid inference on the cross-sectional risk prices. This offers a rigorous way of unifying shrinkage and inference, and is a natural direction for future research building on the penalised FM-GLS framework.

- Commodities indices. In the Rolling Yield Appendix A, we have shown that the two-month WTI futures rolling yield provides a good proxy for the twelve-month rolling yield, and that the Kalman-filtered ETFs and factor-based synthetic rolling yields display the highest correlation with the one-period lagged WTI futures term. However, further evidence is required to establish whether the Kalman synthetic yield, as defined in the appendix, is the most reliable practical proxy in the absence of full futures curve data. As already noted, the synthetic rolling yield construction mixes genuine rolling effects with other sources of variation arising from log returns. In addition, the very high correlations between the Kalman-filtered OIL factor rolling yields and their ETF counterparts is partially due to the fact that the Kalman filter is removing noise. Future research should therefore investigate whether the Kalman synthetic yield can be validated against alternative measures of term premia (i.e., the risk compensation investors demand for holding longer maturities) and whether its apparent efficiency is robust across different commodities, maturities, and market regimes. This would help to clarify whether it can be considered a general-purpose proxy for rolling yields when futures curve data are unavailable.

Appendices

A Rolling Yield

In this appendix, we aim to mimic futures contracts with historical prices. For this, we construct a synthetic rolling yield (RYS) that infers the curve term behaviour indirectly.

From the ETFs description, Section 2.6.2.1, it can be seen that, while most of the Oil ETFs include large- and mid-cap US energy sector firms, USO and DBO are futures related ETFs. This is the motivation to replace the natural log returns (hereafter log returns) and adopt a carry dynamics factor such as the rolling yield, which is a more natural risk driver for commodities.

The WTI crude oil futures contract prices from December 2007 to December 2022 were downloaded from the LSEG (London Stock Exchange Group) Refinitiv platform. We use the tickers NYMWTI1 (Crude Oil WTI NYMEX Close M+1, USD/BBL, $F_{t,1}$) and NYMWTI2 (M+2, $F_{t,2}$), where NYMWTI1 represents the front-month contract—that is, the contract with the nearest delivery date. For the twelve monthly contracts, we used the tickers NCLSM02–NCLSM13 (NYMEX Light Crude Oil Strip M02–M13 Settlement Prices) over the same time period, where NCLSM02 corresponds to the front-month contract.

Rolling yield is the curve slope: the (log) difference between futures maturities (or spot vs front), measured at time t. Because the spot price is not available, it is derived from the slope of the futures curve.

In this work, the two-month WTI rolling yield (RY) is defined as:

$$\mathsf{RY}_{\mathsf{WTI}_{2M},t} = \ln\left(\frac{F_{t,2}}{F_{t,1}}\right), \quad \forall t \tag{4.1}$$

and the twelve-month rolling yield is defined as:

$$\mathsf{RY}_{\mathsf{WTI}_{12M},t} = \sum_{n=3}^{13} \frac{\ln\left(\frac{F_{t,n}}{F_{t,2}}\right)}{n-2}, \quad \forall t \tag{4.2}$$

where:
$F_{t,1}$= price of the 1-month WTI futures contract at time t
$F_{t,2}$= price of the 2-month WTI futures contract at time t [1]

[1] Observed in December, the +1 month contract, that matures in late December, is the January contract (delivers in January). The +2 month contract is the February delivery contract that matures in late January. In detail, according to the CME Group, trading for a given contract month terminates three business days before the

$F_{t,n}$ = price of the n-month WTI futures at time t
n = Delivery month and t =monthly time index that all the future contract refers to.

In Figure 4.1, we compare the rolling yield built using the full twelve-month future term curve with the one built using the first two-month future term.

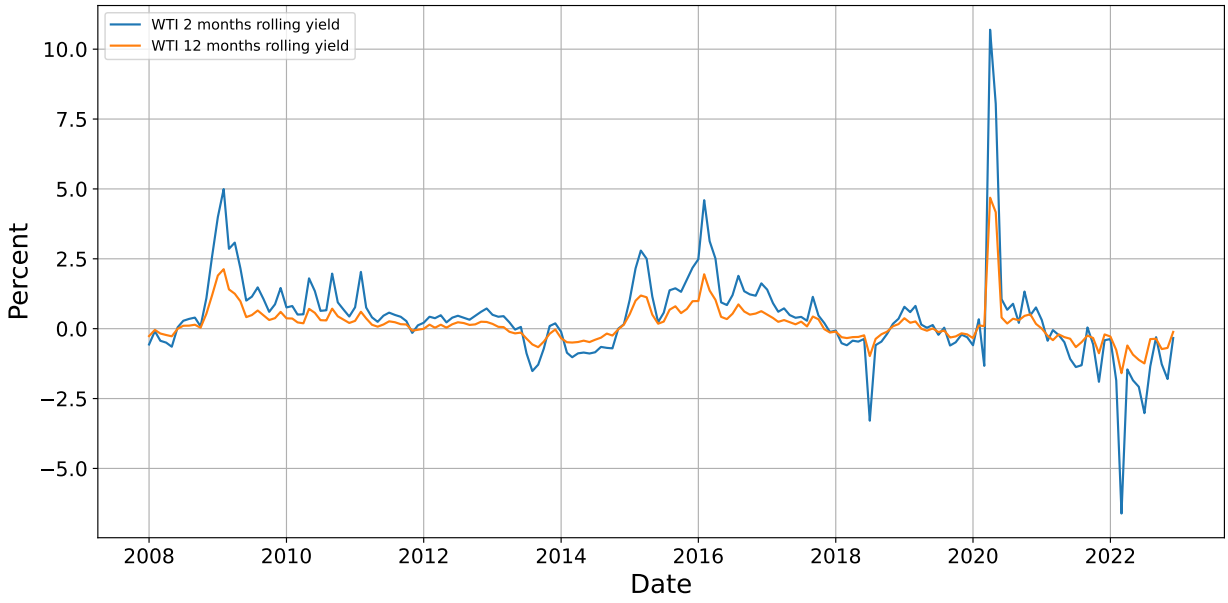

Figure 4.1: WTI 2-month and 12-month Rolling Yields

Sakkas and Tessaromatis (2020) show that momentum is among the most important pricing factors, delivering statistically and economically significant risk prices, outperforming specifications that rely only on untransformed returns. Our risk driver, the rolling yield, captures a term-structure (carry) component that is conceptually distinct from momentum (past returns) yet often empirically related in commodities because trends and curve slopes can co-move. We therefore test whether rolling yield explains part of the momentum premium or provides incremental pricing power relative to untransformed returns.

The rolling yield is positive in backwardation and negative in contango, reflecting the return component from the futures curve shape. The two month and twelve-month WTI rolling yields have a very high correlation (0.975), they show the same trend, the only difference being the higher volatility of the two-month curve. This behaviour is confirmed in the literature: Miffre and Rallis (2007) have shown that rolling yields across different maturities are highly correlated being closely linked to the slope of the term structure. Shorter-dated contracts exhibit more

25th calendar day of the month preceding the delivery month. For example, the February delivery 2008 WTI +2 month futures contract expired on January 22, 2008.

volatile rolling yields, but they largely follow the same trend as longer-dated ones. According to this result, from now on, we operate with the 2 month RY and we refer to it as WTI rolling yield.

Now, we define a synthetic rolling yield (RYS) as the deviation of the current excess return from its long-run average:

$$\mathsf{RYS}_{i,t} = R_{i,t} - \frac{1}{12}\sum_{s=1}^{12} R_{i,t-s} = \mathsf{ln}\left(\frac{P_{i,t}}{P_{i,t-1}}\right) - \frac{1}{12}\sum_{s=1}^{12}\mathsf{ln}\left(\frac{P_{i,t-s}}{P_{i,t-s-1}}\right), \quad \forall i,t \tag{4.3}$$

The formula captures short-term momentum or deviation from trend, which reflect the realised effects of rolling but is also influenced by "spot" price movements, the ETF-specific structure and tracking noise. While the WTI rolling yield directly measures the economic gain or cost of rolling contracts on the futures curve, the synthetic rolling yield infers curve behaviour indirectly by comparing short-term ETFs returns to long-term averages.

In addition, we estimate the rolling yield as a latent factor, using a state-space model based on the discrete time Ornstein-Uhlenbeck process. We postulate that the unobserved rolling yield $y_{i,t}$ follows a mean-reverting process, while observed excess log returns $R_{i,t}$ are modelled as noisy observations of the true yield:

$$y_{i,t} = \phi_i y_{i,t-1} + (1-\phi_i)\mu_i + \eta_{i,t} \quad \eta_{i,t} \sim \mathcal{N}(0, \sigma^2_{\eta_i}), \quad \forall i,t \tag{4.4}$$

$$R_{i,t} = y_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_i}), \quad \forall i,t \tag{4.5}$$

where for each ETF and factor:

- $\phi_i \in (0,1)$: parameter controlling the mean reversion speed,
- μ_i: equilibrium rate or long-run mean level of the latent yield process,
- $\sigma^2_{\eta_i}$: variance of the state innovation (yield shock),
- $\sigma^2_{\varepsilon_i}$: variance of the observation noise (measurement error),
- $y_{i,t}$: latent rolling yield at time t,
- $R_{i,t}$: observed excess log return at time t.

The model is estimated by maximizing the likelihood of the observed returns with respect to the parameters $(\phi_i, \mu_i, \sigma^2_{\eta_i}, \sigma^2_{\varepsilon_i})$. The Kalman filtered estimate of the rolling yield is obtained by using the python package MLEModel. We have $N = 6$ ETFs and $\bar{\ell} = 4$ factors: in total 10 Kalman

filtered rolling yields.

The filtering is performed separately for each time series: the optimised parameter are used for the USO and DBO ETFs and the commodities factors; while for the rest of the ETFs as the parameter estimated by filtering where not realistic, we applied a fix vector parameter that ensures uniform smoothing behaviour.

We expect that USO and DBO, commodities ETFs that are built as futures-based ETFs derive a large portion of their return from the rolling yield, i.e. the gain or loss incurred when rolling futures contracts forward. Standard log returns mix the rolling yield component with the spot price movements, masking the role of the future term structure (i.e., contango and backwardation). Rolling yield provides a cleaner measure of the systematic risk priced in the cross-section of commodities.

In equation 4.1, we defined the two-month WTI rolling yield (RY). However, futures prices converge to the spot price at contract maturity, which typically occurs a few days before the start of the stated delivery month; therefore, the term structure observed at time $t-1$ (Lag 1) better captures the information relevant for the return realised by the ETFs over the interval $[t-1, t]$. For example, observed in December, the WTI February contract physical delivery begins in February but reflects expectations about spot prices in late January. This will coincide in between the one-month lag and the two-month lag WTI rolling yield, however we decide to use the one-period lag WTI rolling yield, as it exhibits the highest correlation with the ETFs over the interval $[t-1, t]$. :

$$\mathrm{RY}^{(1)}_{\mathrm{WTI}_{2M},t} = \ln\left(\frac{F_{t-1,2}}{F_{t-1,1}}\right), \tag{4.6}$$

In Figures 4.2, 4.3, and 4.4, we compare the WTI 2 months RY lag 1 defined in equation 4.6 with the synthetic rolling yield and the Kalman smoothed RYS for USO, DBO and XLE:

We also show the correlation matrices in Tables 4.1, 4.2, 4.3.

	USO Return	USO RYS	USO Kalman RYS	OIL Kalman RYS	OIL Return	WTI Return	WTI RY	WTI RY (Lag 1)
USO Return	1.000	0.937	0.863	0.671	0.793	0.340	-0.100	0.145
USO RYS	0.937	1.000	0.953	0.799	0.778	0.322	0.101	0.336
USO Kalman RYS	0.863	0.953	1.000	0.864	0.770	0.439	0.065	0.339
OIL Kalman RYS	0.671	0.799	0.864	1.000	0.865	0.553	0.226	0.515
OIL Return	0.793	0.778	0.770	0.865	1.000	0.605	-0.015	0.315
WTI Return	0.340	0.322	0.439	0.553	0.605	1.000	-0.359	0.269
WTI RY	-0.100	0.101	0.065	0.226	-0.015	-0.359	1.000	0.630
WTI RY (Lag 1)	0.145	0.336	0.339	0.515	0.315	0.269	0.630	1.000

Table 4.1: Correlation Matrix: USO

We can see that the Kalman synthetic yield is the best practical proxy in the absence of the

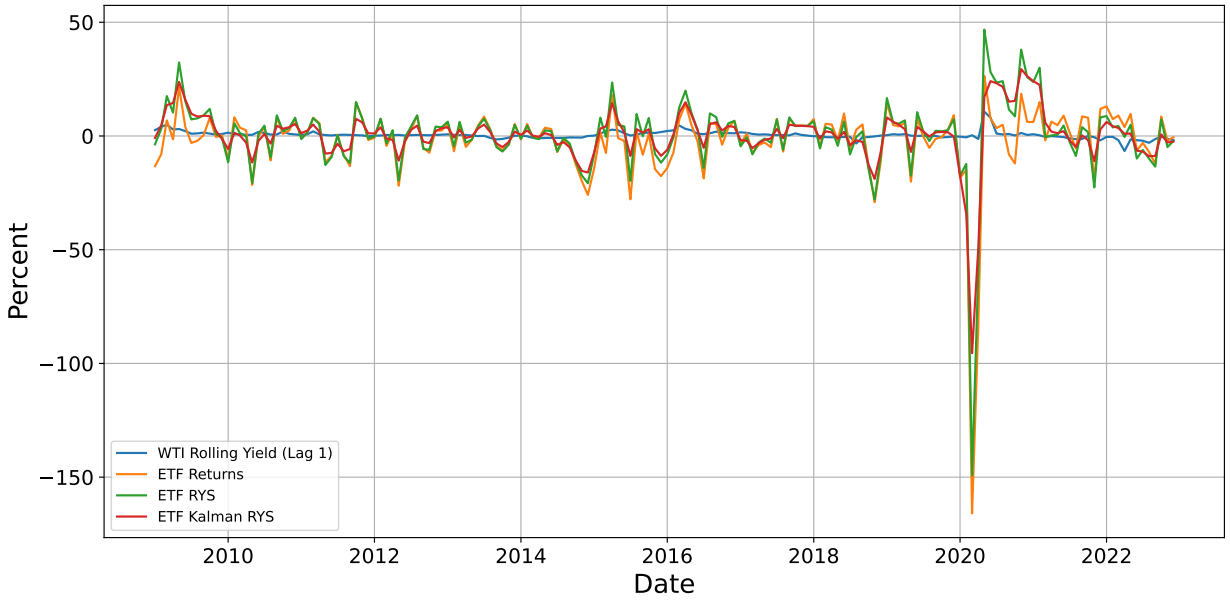

Figure 4.2: USO Kalman Filtered Synthetic Rolling Yields

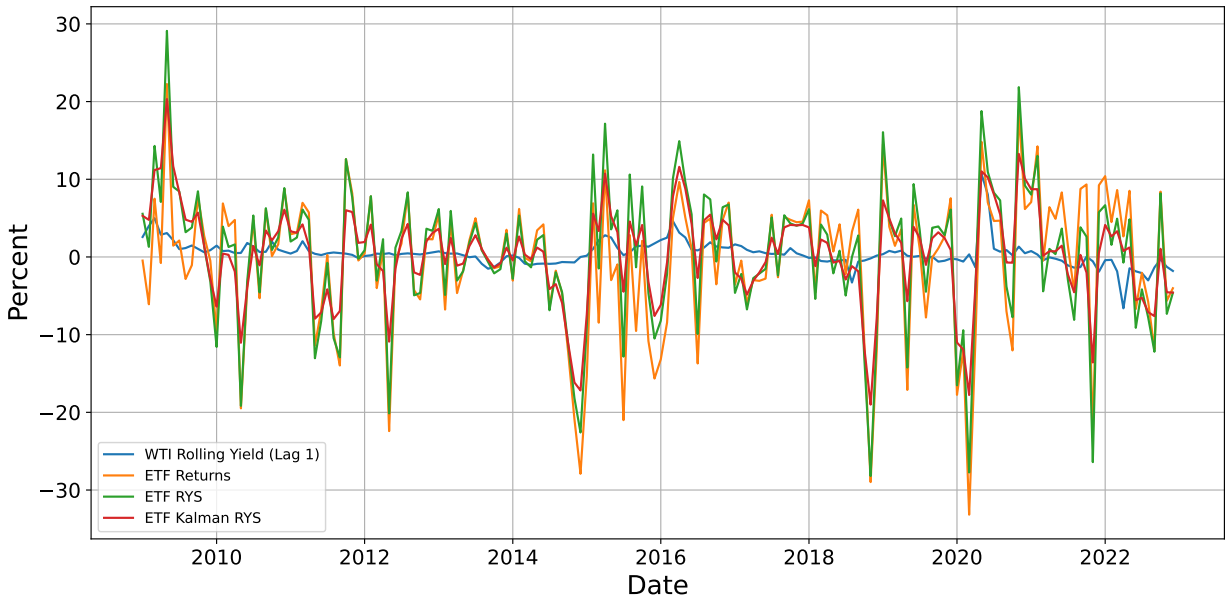

Figure 4.3: DBO Kalman Filtered Synthetic Rolling Yields

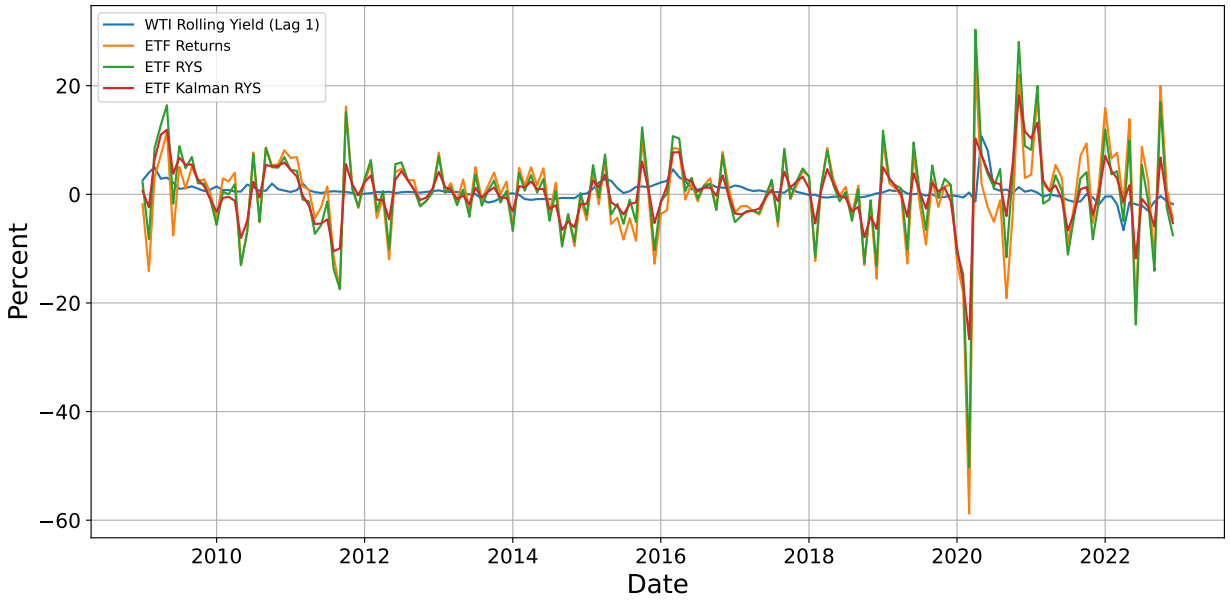

Figure 4.4: XLE Kalman Filtered Synthetic Rolling Yields

	DBO Return	DBO RYS	DBO Kalman RYS	OIL Kalman RYS	OIL Return	WTI Return	WTI RY	WTI RY (Lag 1)
DBO Return	1.000	0.933	0.837	0.634	0.679	0.259	0.002	0.159
DBO RYS	0.933	1.000	0.941	0.723	0.621	0.204	0.219	0.356
DBO Kalman RYS	0.837	0.941	1.000	0.841	0.653	0.339	0.262	0.428
OIL Kalman RYS	0.634	0.723	0.841	1.000	0.865	0.553	0.226	0.515
OIL Return	0.679	0.621	0.653	0.865	1.000	0.605	-0.015	0.315
WTI Return	0.259	0.204	0.339	0.553	0.605	1.000	-0.359	0.269
WTI RY	0.002	0.219	0.262	0.226	-0.015	-0.359	1.000	0.630
WTI RY (Lag 1)	0.159	0.356	0.428	0.515	0.315	0.269	0.630	1.000

Table 4.2: Correlation Matrix: DBO

	XLE Return	XLE RYS	XLE Kalman RYS	OIL Kalman RYS	OIL Return	WTI Return	WTI RY	WTI RY (Lag 1)
XLE Return	1.000	0.952	0.849	0.438	0.479	-0.024	0.108	0.023
XLE RYS	0.952	1.000	0.929	0.528	0.445	-0.048	0.270	0.178
XLE Kalman RYS	0.849	0.929	1.000	0.694	0.536	0.130	0.277	0.281
OIL Kalman RYS	0.438	0.528	0.694	1.000	0.865	0.553	0.226	0.515
OIL Return	0.479	0.445	0.536	0.865	1.000	0.605	-0.015	0.315
WTI Return	-0.024	-0.048	0.130	0.553	0.605	1.000	-0.359	0.269
WTI RY	0.108	0.270	0.277	0.226	-0.015	-0.359	1.000	0.630
WTI RY (Lag 1)	0.023	0.178	0.281	0.515	0.315	0.269	0.630	1.000

Table 4.3: Correlation Matrix: XLE

futures curve data, although it should be interpreted with caution as we have already pointed out that it mixes rolling effects with other sources of variation from the log returns. We also notice that the correlation of the Kalman smoothed RYS for USO, DBO and XLE (the same is observed for the other ETFs) with the WTI 2 months RY is higher than the log return one. Finally we observe very high correlation between the Kalman filtered OIL factor RYS and the ETFs one (USO 0.864, DBO 0.841, XLE 0.694). Of course, this is partially due to the fact that

the Kalman filter is removing the noise.

In summary, we have shown that the WTI future 2 months rolling yield is a good proxy for the WTI future 12 months RY. We also show that Kalman-filtered ETF and factor RYS have the highest correlations with the one-lag two-month WTI RY.

According to these results, the Kalman filtered time series appears to be more suitable risk drivers than log returns to be used in the commodities Fama–MacBeth integrated regression, for both the dependent variables (de-noised portfolio returns) and the independent variables (factor exposures).

The Fama-MacBeth regression for the integrated and segmented model are shown in Table 4.4 and 4.5, the excess log return percentage have been replaced by the rolling yields percentage. Our results show that rolling yield regressions compared with excess log return regressions, exhibit lower residual variance, and stronger pricing of key factors under both GLS and OLS.

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
AL_OIL	FM OLS	λ_{AL}	-0.0308	0.958		Estimate not statistically significant; no evidence of integration	
AL_OIL	FM OLS	λ_{OIL}	0.3225	0.4201		Estimate is not statistically significant; no evidence of segmentation	
GAS_OIL	FM OLS	λ_{GAS}	0.3712	0.4962		Estimate not statistically significant; no evidence of integration	
GAS_OIL	FM OLS	λ_{OIL}	0.254	0.0753		Estimate is not statistically significant; no evidence of segmentation	
SOY_OIL	FM OLS	λ_{SOY}	0.3176	0.4105		Estimate not statistically significant; no evidence of integration	
SOY_OIL	FM OLS	λ_{OIL}	0.2774	0.0141	**	Statistically significant and economically large; integration rejected	IR
AL_OIL	FM GLS	λ_{AL}	0.4799	0.1854		Estimate not statistically significant; no evidence of integration	
AL_OIL	FM GLS	λ_{OIL}	0.0374	0.8757		Estimate is not statistically significant; no evidence of segmentation	
GAS_OIL	FM GLS	λ_{GAS}	-2.4614	0.0065	***	Statistically significant and economically large; evidence of integration	PI
GAS_OIL	FM GLS	λ_{OIL}	0.9161	0.0019	***	Statistically significant and economically large; partial integration	PI
SOY_OIL	FM GLS	λ_{SOY}	2.0474	0.0974		Estimate not statistically significant; no evidence of integration	
SOY_OIL	FM GLS	λ_{OIL}	0.1657	0.3261		Estimate is not statistically significant; no evidence of segmentation	

Table 4.4: ETFs Synthetic Rolling Yield Integration Test Results

For the segmented model, the estimated price of risk of the rolling yield OIL factor in the pair with the GAS orthogonal component is 0.76% with $p < 0.01$. This implies that an ETF with unit exposure to the rolling yield factor earns, on average, 0.76% higher monthly excess return. Annualized, this corresponds to a risk price of approximately: $0.76 \times 12 = 9.1\%$ per year

Area	Method	Parameter	Estimate	Pr $> \|t\|$	Signif.	Comments	Results
OIL_AL	FM OLS	δ_{OIL}	0.1706	0.1104		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	FM OLS	δ_{AL}	-0.7189	0.1222		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	FM OLS	δ_{OIL}	0.2756	0.0306	**	Significant and different from zero, which shows segmentation	TS
OIL_GAS	FM OLS	δ_{GAS}	0.3601	0.5401		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_SOY	FM OLS	δ_{OIL}	0.3657	0.0154	**	Significant and different from zero, which shows segmentation	TS
OIL_SOY	FM OLS	δ_{SOY}	0.2735	0.5692		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_AL	FM GLS	δ_{OIL}	0.4596	0.0701		Estimate not statistically significant; no evidence of segmentation	
OIL_AL	FM GLS	δ_{AL}	0.332	0.3146		Estimate not statistically significant; no evidence of segmentation rejection	
OIL_GAS	FM GLS	δ_{OIL}	0.7555	0.0015	***	Significant and different from zero, which shows segmentation	PS
OIL_GAS	FM GLS	δ_{GAS}	-2.5566	0.0053	***	Significant and different from zero, which shows partial segmentation	PS
OIL_SOY	FM GLS	δ_{OIL}	0.7678	0.0342	**	Significant and different from zero, which shows segmentation	TS
OIL_SOY	FM GLS	δ_{SOY}	2.0093	0.1187		Estimate not statistically significant; no evidence of segmentation rejection	

Table 4.5: ETFs Synthetic Rolling Yield Segmentation Test Results

The result is statistically significant and confirms that rolling yield, as a futures curve-derived return component, is priced in the cross-section of commodity-linked ETFs.

Finally, in Table 4.6 we report the Harvey–Liu Incremental factor significance test using the synthetic rolling yield.

Factor Pair	Mean	p (mean)	Signif.	Median	p (median)	Signif.
AL-OIL	-0.0013	0.7430		-0.0040	0.7920	
AL-OIL_AL_ort	-0.0011	0.6410		-0.0056	0.7420	
GAS-OIL	0.0062	0.0570		0.0079	0.0600	
GAS-OIL_GAS_ort	0.0064	0.0300	*	0.0083	0.0250	*
SOY-OIL	0.0042	0.1080		0.0041	0.4170	
SOY-OIL_SOY_ort	0.0042	0.0910		0.0041	0.4090	
OIL-AL	-0.0002	0.7400		-0.0007	0.7950	
OIL-AL_OIL_ort	0.0024	0.2850		0.0008	0.4720	
OIL-GAS	0.0022	0.0520		0.0013	0.0400	*
OIL-GAS_OIL_ort	0.0027	0.0240	*	0.0024	0.0180	*
OIL-SOY	0.0023	0.3180		0.0012	0.3050	
OIL-SOY_OIL_ort	0.0005	0.3560		-0.0037	0.5550	

Table 4.6: Synthetic Rolling Yield Harvey–Liu Bootstrap Test for Factor Significance

For the pairs Gas Oil and Oil Gas, the mean reduction in pricing errors is statistically significant, that suggests the Gas / Oil factors adds explanatory power to the OIL ETFs. The median reduction is also significant, indicating that the effect is spanned across all ETFs. This is consistent with the previous analysis where we detected partial segmentation of the OIL and Gas commodities factor.

B Error Estimation

For the stack vector of parameters $\hat{\boldsymbol{\theta}}$, we compute the sum of squared residuals (SSR):

$$\text{SSR} = \hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2 \tag{4.7}$$

the sample variance:

$$\hat{\sigma}^2 = \frac{\text{SSR}}{NT - J} \tag{4.8}$$

where $T = 240$ months, corresponds to the number of time periods; $N = 6$, is the number of portfolios; and j is the parameter index from 1 to J, and J denotes the number of parameters, which changes depending on the method.

The standard errors are derived using the design matrix $\mathbf{X} \in \mathbb{R}^{NT \times J}$ evaluated at $\hat{\boldsymbol{\theta}}$[2]. Then residuals are computed as:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}} \tag{4.9}$$

The variance-covariance matrix of the estimators is given by:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) = \hat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \tag{4.10}$$

The standard error for each parameter $\hat{\theta}_j$ is:

$$\text{SE}(\hat{\theta}_j) = \sqrt{\left[\widehat{\text{Var}}(\hat{\boldsymbol{\theta}})\right]_{jj}}, \quad \forall j \tag{4.11}$$

The corresponding test statistic and p-value are computed as:

$$t_j = \frac{\hat{\theta}_j}{\text{SE}(\hat{\theta}_j)}, \quad \forall j \tag{4.12}$$

[2] In the Fama-MacBeth method, we use the two-pass pseudo design matrix, Shanken (1992), which addresses errors-in-variables in cross-sectional regressions with estimated betas

$$p_j = 2\left(1 - \mathcal{T}_\nu\big(|t_j|\big)\right), \quad \forall j \tag{4.13}$$

where $\mathcal{T}_\nu(\cdot)$ denotes the cumulative distribution function of the Student's t-distribution with $\nu = NT - J$ degrees of freedom.

For the Taylor Product method, $J = N + 2$ corresponds to the number of estimated parameters, and the parameter vector is:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\lambda}_0 & \hat{\lambda}_f & \hat{\beta}_1 & \dots & \hat{\beta}_N \end{bmatrix}^\top \in \mathbb{R}^{N+2},$$

For the Product Factor method, $J = 2N + 2$ is the total number of estimated parameters in $\hat{\boldsymbol{\theta}}$ and the parameter vector is:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\lambda}_0 & \hat{\lambda}_f & \hat{\beta}_1 & \dots & \hat{\beta}_N & \hat{\gamma}_1 & \dots & \hat{\gamma}_N \end{bmatrix}^\top \in \mathbb{R}^{2N+2}.$$

For the Integer Programming method, $J = 2N+2$ is the number of parameters and the parameter vector is:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\lambda}_0 & \hat{\lambda}_f & \hat{y}_{1,1} & \dots & \hat{y}_{N,1} & \hat{y}_{1,2} & \dots & \hat{y}_{N,2} \end{bmatrix}^\top \in \mathbb{R}^{2N+2},$$

C Kronecker Product

If $\mathbf{M}$ is an $N \times S$ matrix and $\mathbf{Z}$ is a $T \times \bar{p}$ matrix, then the matrix direct product or Kronecker product $\mathbf{M} \otimes \mathbf{Z}$ is an $(NT) \times (S\bar{p})$ block matrix defined by:

$$\mathbf{M} \otimes \mathbf{Z} = \begin{bmatrix} m_{11}\mathbf{Z} & m_{12}\mathbf{Z} & \cdots & m_{1S}\mathbf{Z} \\ m_{21}\mathbf{Z} & m_{22}\mathbf{Z} & \cdots & m_{2S}\mathbf{Z} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1}\mathbf{Z} & m_{N2}\mathbf{Z} & \cdots & m_{NS}\mathbf{Z} \end{bmatrix} \in \mathbb{R}^{(NT)\times(S\bar{p})} \tag{4.14}$$

Examples:
From equation 2.50:

If $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the identity matrix and $\mathbf{x}_t + \boldsymbol{\lambda}_f = \begin{bmatrix} x_{1,t} + \lambda_{f1} \\ x_{2,t} + \lambda_{f2} \end{bmatrix} \in \mathbb{R}^2$, then the Kronecker

product $\left(\mathbf{I}_N \otimes (\mathbf{x}_t + \boldsymbol{\lambda}_f)^\top\right) \in \mathbb{R}^{N \times 2N}$ can be written block-wise as:

$$\begin{bmatrix} (x_{1,t}+\lambda_{f1}) & (x_{2,t}+\lambda_{f2}) & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & (x_{1,t}+\lambda_{f1}) & (x_{2,t}+\lambda_{f2}) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (x_{1,t}+\lambda_{f1}) & (x_{2,t}+\lambda_{f2}) \end{bmatrix}$$

From equation 2.54:
If $\mathbf{M}_t(\boldsymbol{\theta}) \in \mathbb{R}^N$ is a residual column vector and $\mathbf{z}_t \in \mathbb{R}^{\bar{p}}$ is a column vector of instruments, then their Kronecker product is:

$$\mathbf{g}_t(\boldsymbol{\theta}) = \mathbf{M}_t(\boldsymbol{\theta}) \otimes \mathbf{z}_t = \begin{bmatrix} m_{1,t}\mathbf{z}_t \\ m_{2,t}\mathbf{z}_t \\ \vdots \\ m_{N,t}\mathbf{z}_t \end{bmatrix} \in \mathbb{R}^{N\bar{p}} \tag{4.15}$$

That is:

$$\mathbf{g}_t(\boldsymbol{\theta}) = \begin{bmatrix} m_{1,t}z_{1,t} \\ m_{1,t}z_{2,t} \\ m_{1,t}z_{3,t} \\ m_{2,t}z_{1,t} \\ m_{2,t}z_{2,t} \\ \vdots \\ m_{N,t}z_{\bar{p},t} \end{bmatrix}$$

D Vectorization

The operator vec($\boldsymbol{B}$) denotes the vectorization of the matrix $\boldsymbol{B} \in \mathbb{R}^{\bar{\ell} \times N}$:

$$\boldsymbol{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1N} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{\bar{\ell}1} & \beta_{\bar{\ell}2} & \cdots & \beta_{\bar{\ell}N} \end{bmatrix} \in \mathbb{R}^{\bar{\ell} \times N} \tag{4.16}$$

It is obtained by stacking the columns of $\boldsymbol{B}$ into a single column vector:

$$\text{vec}(\boldsymbol{B}) = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{\bar{\ell}1} \\ \beta_{12} \\ \vdots \\ \beta_{\bar{\ell}N} \end{bmatrix} \in \mathbb{R}^{\bar{\ell}N}. \tag{4.17}$$

E Ordinary Least Squares for a System of Equations

We consider a system consisting of N regression models, each observed over T time periods, yielding N×T scalar equations.

$$y_{i,t} = \boldsymbol{\beta}_i \mathbf{x}_t + \varepsilon_{i,t}, \quad \forall i, t \tag{4.18}$$

Here, i indexes the cross-sectional unit (e.g., portfolio or equation), and t is the index of the observation at time t. We use the row vector $\boldsymbol{\beta}_i \in \mathbb{R}^{1\times\bar{\ell}}$ for equation i to match the dimension of column vector $\mathbf{x}_t \in \mathbb{R}^{\bar{\ell}}$. Each equation i has a single response variable $y_{i,t}$, while the vector of regressors $\mathbf{x}_t$ is common across all equations.

In econometrics, it is customary to denote observations as $y_{i,t}$, where i refers to the cross-sectional unit (e.g., asset or equation) and t refers to the time period. Thus, $y_{i,t}$ represents the observation for asset i at time t. For matrix operations, however, it is convenient to arrange the data by time in rows and portfolios in columns. Therefore, we define the observation matrix:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & y_{ti} & \vdots \\ y_{T1} & y_{T2} & \cdots & y_{TN} \end{bmatrix}.$$

Similarly, the coefficient vectors $\boldsymbol{\beta}_i$ are rearranged to form the matrix

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\beta}_1^\top & \boldsymbol{\beta}_2^\top & \cdots & \boldsymbol{\beta}_N^\top \end{bmatrix} \in \mathbb{R}^{\bar{\ell}\times N},$$

where each column corresponds to the coefficients of one equation, and each row to a regressor.

With these definitions, the scalar model

$$y_{i,t} = \boldsymbol{\beta}_i \mathbf{x}_t + \varepsilon_{i,t}$$

can be expressed in the standard matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \mathbf{E},$$

where:

- $\mathbf{X} \in \mathbb{R}^{T \times \bar{\ell}}$ is the regressor matrix with rows $\mathbf{x}_t^\top$,
- $\boldsymbol{B} \in \mathbb{R}^{\bar{\ell} \times N}$ is the coefficient matrix,
- $\mathbf{E} \in \mathbb{R}^{T \times N}$ is the error matrix.

Alternatively, stacking observations corresponding to the i-th equation over time into T-dimensional vectors and matrices, the model can be written in vector form as:

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad \forall i \tag{4.19}$$

where:

- $\mathbf{y}_i \in \mathbb{R}^T$: response vector for equation i,
- $\mathbf{X} \in \mathbb{R}^{T \times \bar{\ell}}$: regressor matrix (same for all equations),
- $\boldsymbol{\beta}_i \in \mathbb{R}^{\bar{\ell}}$: coefficient vector specific to equation i,
- $\boldsymbol{\varepsilon}_i \in \mathbb{R}^T$: error vector for equation i.

Stacking all N equations, we obtain:

$$\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}) + \boldsymbol{\varepsilon}, \tag{4.20}$$

where:

- $\mathbf{y} = \text{vec}(\mathbf{Y}) = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{bmatrix}^\top = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{T1} & y_{12} & \cdots & y_{TN} \end{bmatrix}^\top \in \mathbb{R}^{TN}$: stacked response vector obtained by stacking columns of $\mathbf{Y}$ (ordered by time periods and equations, all T observations for $i = 1$, then for $i = 2$, and so on),
- $\boldsymbol{B} \in \mathbb{R}^{\bar{\ell} \times N}$: coefficient matrix (columns correspond to equations, rows to regressors),

- $\text{vec}(\boldsymbol{B}) = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_N \end{bmatrix}^\top \in \mathbb{R}^{\bar{\ell}N}$: vectorized coefficient matrix obtained by stacking columns of $\boldsymbol{B}$ (ordered by regressors and equations),
- $\boldsymbol{\varepsilon} = \text{vec}(\mathbf{E}) = \begin{bmatrix} \boldsymbol{\varepsilon}_1 & \boldsymbol{\varepsilon}_2 & \cdots & \boldsymbol{\varepsilon}_N \end{bmatrix}^\top \in \mathbb{R}^{TN}$: stacked error vector obtained by stacking columns of $\mathbf{E}$ (ordered by time periods and equations).

Note: The notation with transpose indicates that stacking is column-wise, consistent with the standard vectorisation, vec().

The OLS estimator minimizes the sum of squares residual (SSR) to estimate the parameter vector $\text{vec}(\boldsymbol{B})$ from the stacked model:

$$\widehat{\text{vec}(\boldsymbol{B})} = \arg\min_{\text{vec}(\boldsymbol{B})} (\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}))^\top (\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B})) \tag{4.21}$$

Taking the derivative with respect to $\text{vec}(\boldsymbol{B})$ and setting it equal to zero:

$$\frac{\partial}{\partial \, \text{vec}(\boldsymbol{B})} \left[(\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}))^\top (\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B})) \right] = 0 \tag{4.22}$$

We obtain:

$$-2(\mathbf{I}_N \otimes \mathbf{X})^\top \mathbf{y} + 2(\mathbf{I}_N \otimes \mathbf{X})^\top (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}) = 0 \tag{4.23}$$

Solving for $\text{vec}(\boldsymbol{B})$ gives the OLS estimator:

$$\widehat{\text{vec}(\boldsymbol{B})} = \left[(\mathbf{I}_N \otimes \mathbf{X})^\top (\mathbf{I}_N \otimes \mathbf{X}) \right]^{-1} (\mathbf{I}_N \otimes \mathbf{X})^\top \mathbf{y} \tag{4.24}$$

The OLS single equation linear objective function is :

$$\hat{\boldsymbol{\beta}}_{i,OLS} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y_i} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y_i} - \mathbf{X}\boldsymbol{\beta}) \tag{4.25}$$

In SAS-style notation, the OLS objective function is [3]:

$$\text{Objective}_{\text{OLS}} = \frac{1}{T} \mathbf{r}^\top \mathbf{r} \tag{4.26}$$

where:

- $\mathbf{r} = \mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}) \in \mathbb{R}^{TN}$ is the stacked vector of residuals $\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_N \end{bmatrix}^\top$

[3] The scaling by $\frac{1}{T}$ in the SAS objective function, Section 2.4, is used to express it as a sample average rather than a total sum, see also Appendix K.

and each $\mathbf{r}_i \in \mathbb{R}^T$ contains residuals from a (non) linear equation i.

We can observe that the multi-equation stacked model:

$$\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}) + \boldsymbol{\varepsilon} \tag{4.27}$$

can be interpreted as a generalization of the standard single-equation regression model. In fact, for each individual equation $i \in \{1, \ldots, N\}$, we have:

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \tag{4.28}$$

where $\boldsymbol{\beta}_i \in \mathbb{R}^{\bar{\ell}}$ is the i-th column of $\boldsymbol{B}$ transposed, and $\mathbf{X} \in \mathbb{R}^{T \times \bar{\ell}}$ is the regressor matrix.

We can switch from the multi-equation model to the single equation formulation substituting:

$$\boldsymbol{\beta}_i \longleftrightarrow \text{vec}(\boldsymbol{B}) \tag{4.29}$$

$$\mathbf{X} \longleftrightarrow \mathbf{I}_N \otimes \mathbf{X} \tag{4.30}$$

In this way, we obtain the classic single equation OLS estimator from equation 4.24[4]:

$$\hat{\boldsymbol{\beta}}_i = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}_i \tag{4.31}$$

From now on, we will use the single equation model. The multi-equation equivalent results can be obtained by applying equivalence 4.29 and 4.30.

F Gauss-Markov Assumptions

The OLS estimator is subject to the Gauss-Markov assumptions:

- Identification condition or no multicollinearity of the independent variables: $\mathbf{X} \in \mathbb{R}^{T \times \bar{\ell}}$ is a full-rank matrix ($\text{rank}(\mathbf{X}) = \bar{\ell}$). We denote its columns by $\mathbf{x}_j \in \mathbb{R}^T$, $j = 1, \ldots, \bar{\ell}$. This condition can be written in terms of correlation as: $\rho(\mathbf{x}_i, \mathbf{x}_j) \not\approx 1$ for $i \neq j$.

- The error term $\boldsymbol{\varepsilon}$ has conditional expectation zero given the stacked regressor matrix[5]:

$$\mathbb{E}[\boldsymbol{\varepsilon} \mid (\mathbf{I}_N \otimes \mathbf{X})] = \mathbf{0}.$$

  - This implies that for any measurable function $f(\mathbf{X}) = \mathbf{I}_N \otimes \mathbf{X}$, we have the orthogonal-

[4] Using the same method, we can derive equation 4.25 from equation 4.21.

[5] Since all N equations share the same regressor matrix $\mathbf{X}$, conditioning on $\mathbf{X}$ or on $(\mathbf{I}_N \otimes \mathbf{X})$ is equivalent.

ity condition $\mathbb{E}[(\mathbf{I}_N \otimes \mathbf{X})^\top \boldsymbol{\varepsilon}] = \mathbf{0}$, which follows from the law of iterated expectations:

$$\mathbb{E}[f(\mathbf{X})^\top \boldsymbol{\varepsilon}] = \mathbb{E}\big[\mathbb{E}[f(\mathbf{X})^\top \boldsymbol{\varepsilon} \mid \mathbf{X}]\big] = \mathbb{E}\big[f(\mathbf{X})^\top \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}]\big] = 0.$$

By choosing $f(\mathbf{X}) = \mathbf{x}$ [6], we obtain the population orthogonality condition: $\mathbb{E}[\mathbf{x} \otimes \boldsymbol{\varepsilon}] = \mathbf{0}$.

- Hence (since $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$), we have no endogeneity or omitted variable bias, the regressors are uncorrelated with the error term:

$$\mathbb{E}\big[(\mathbf{I}_N \otimes \mathbf{X})^\top \boldsymbol{\varepsilon}\big] = \mathbf{0}.$$

- Homoscedasticity and no autocorrelation of the errors: for the stacked system, the disturbance vector $\boldsymbol{\varepsilon} \in \mathbb{R}^{TN}$ is assumed to satisfy $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \mid \mathbf{X}] = \sigma^2 \mathbf{I}_{TN}$. This implies homoscedasticity (constant variance across time and portfolios) and no autocorrelation (errors are uncorrelated across both time and equations):

$$\mathsf{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_{TN} = \sigma^2 (\mathbf{I}_T \otimes \mathbf{I}_N)$$

$$\mathsf{Cov}(\varepsilon_{i,t},\, \varepsilon_{j,s} \mid \mathbf{X}) = 0 \quad \text{for } (i,t) \neq (j,s)$$

If, in addition, (conditional) normality is assumed, then $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim \mathcal{N}(\mathbf{0},\, \sigma^2(\mathbf{I}_T \otimes \mathbf{I}_N))$. The Kronecker structure $\mathbf{I}_T \otimes \mathbf{I}_N$ indicates zero covariance across both time and cross-sectional portfolios (independence follows under normality).

These two conditions together define the case of spherical disturbances. More generally, the variance-covariance matrix of the error vector is given by, see Section 2.4:

$$\mathsf{Cov}(\mathbf{r} \mid \mathbf{X}) = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \mid \mathbf{X}] = \mathsf{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \boldsymbol{\Omega} = \sigma^2 (\boldsymbol{\Omega}_T \otimes \boldsymbol{\Omega}_N) \neq \sigma^2 \mathbf{I}_{TN} \tag{4.32}$$

where $\boldsymbol{\Omega}_T$ and $\boldsymbol{\Omega}_N$ are correlation matrices (unit diagonal):

- $\boldsymbol{\Omega}_T \in \mathbb{R}^{T \times T}$ captures time-series autocorrelation within each equation.

[6] In more general models, the functional form can be written as a normalized nonlinear representation:

$$\mathbf{y}_t = \mathbf{f}(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{f}(\cdot)$ may be nonlinear and incorporate both endogenous and exogenous variables. In the GMM model, the orthogonality condition generalizes to the instrumental variables: $\mathbb{E}[\mathbf{z} \otimes \boldsymbol{\varepsilon}] = \mathbf{0}$, for any valid instrument $\mathbf{z}_t$ that is a measurable function of $\mathbf{x}_t$.

- $\mathbf{\Omega}_N \in \mathbb{R}^{N \times N}$ captures cross-sectional correlation across portfolios,

Spherical disturbances correspond to $\mathbf{\Omega}_T = \mathbf{I}_T$ and $\mathbf{\Omega}_N = \mathbf{I}_N$.

In general, $\mathbf{\Omega}$ is a matrix, whose main diagonal elements are the scaled variances of errors and all other elements are the scaled covariances of errors. We can have heteroscedasticity and autocorrelation within $\mathbf{\Omega}_N$ and\or $\mathbf{\Omega}_T$. For the cross-sectional correlation, for example, the elements of the cross correlation matrix are: $\omega_{ij} = \frac{\mathsf{Cov}(\varepsilon_i, \varepsilon_j)}{\sigma^2}, \quad \text{for } i, j \in \{1, \ldots, N\}$. In the case of homoscedasticity and no autocorrelation, $\mathbf{\Omega_N} = \mathbf{I}_N$. If the disturbances are only homoscedastic, then $\omega_{ii} = 1$ for all i, and if they are uncorrelated, all off-diagonal elements $\omega_{ij} = 0$ for $i \neq j$. Similarly, for the time-series component, under homoscedasticity and no autocorrelation, the same structure applies, $\mathbf{\Omega_T} = \mathbf{I}_T$.

  - OLS: assumes $\mathbf{\Omega}_N = \mathbf{I}_N$, $\mathbf{\Omega}_T = \mathbf{I}_T$.
  - SUR / ITSUR: assumes $\mathbf{\Omega}_T = \mathbf{I}_T$, but allows $\mathbf{\Omega}_N \neq \mathbf{I}_N$.
  - HAC (robust SEs only): allows $\mathbf{\Omega}_T \neq \mathbf{I}_T$; typically assumes $\mathbf{\Omega}_N = \mathbf{I}_N$. HAC modifies only the covariance estimator (e.g., Newey–West); point estimates are unchanged.
  - Time-series GLS: specifies a structure for $\mathbf{\Omega}_T \neq \mathbf{I}_T$ (e.g., AR(1), ARMA), usually with $\mathbf{\Omega}_N = \mathbf{I}_N$. GLS changes both the estimator and its covariance.
  - Full GLS: allows both $\mathbf{\Omega}_N \neq \mathbf{I}_N$ and $\mathbf{\Omega}_T \neq \mathbf{I}_T$; fully general error structure.

- Each observation $(\mathbf{x}_t, \mathbf{y}_t), \forall t$ is independently and identically distributed (i.i.d.) and independently drawn from their joint distribution.
- Finally, for inference and hypothesis testing, a common assumption is normality of the error terms: $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim \mathcal{N}(0, \sigma^2(\mathbf{I}_T \otimes \mathbf{I}_N))$, based on the Central Limit Theorem.

Conditional on the assumptions mentioned above, the Gauss-Markov Theorem states that there is no other linear and unbiased estimator of the vec($\boldsymbol{B}$) parameters with smaller sampling variance. That is, the OLS estimator $\widehat{\text{vec}(\boldsymbol{B})}$ is the Best Linear Unbiased Estimator (BLUE).

G Generalized Least Squares (GLS)

The full Generalized Least Squares objective function is defined below:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathbf{r}(\boldsymbol{\theta})^{\top} \left(\text{Cov}(\mathbf{r})\right)^{-1} \mathbf{r}(\boldsymbol{\theta}), \qquad \text{where } \mathbf{r}(\boldsymbol{\theta}) = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}) \tag{4.33}$$

The function $f(\boldsymbol{\theta}) \in \mathbb{R}^{TN}$ is a nonlinear function of the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{\bar{k}}$, which replaces $(\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B})$ for the nonlinear model (see Section I), $\text{Cov}(\mathbf{r}) \in \mathbb{R}^{TN \times TN}$.

The matrix $\text{Cov}(\mathbf{r})^{-1}$ is a non-identity, positive-definite weighting matrix $\mathbf{W}$ that:

- in the full GLS, accounts for heteroscedasticity, autocorrelation and correlation across equations in the residuals;
- reduces to Seemingly Unrelated Regression if $\text{Cov}(\mathbf{r})$ has the form $\mathbf{S} \otimes \mathbf{I_T}$ where $\mathbf{S}$ accounts for both heteroscedasticity and correlation across equations in the residuals:

$$\hat{\boldsymbol{\theta}}_{SUR} = \arg\min_{\boldsymbol{\theta}} \mathbf{r}(\boldsymbol{\theta})^{\top} \left(\mathbf{S}^{-1} \otimes \mathbf{I}_T\right) \mathbf{r}(\boldsymbol{\theta}), \qquad \text{where } \mathbf{r}(\boldsymbol{\theta}) = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}) \tag{4.34}$$

- reduces to nonlinear Weighted Least Squares (WLS) if $\text{Cov}(\mathbf{r}) \in \mathbb{R}^{TN \times TN}$ has the form $\text{Cov}(\mathbf{r}) = \mathbf{W} \otimes \mathbf{I}_T$, where $\mathbf{W} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2) \in \mathbb{R}^{N \times N}$, which corresponds to heteroscedasticity across equations, In this case, the nonlinear WLS objective function is:

$$\hat{\boldsymbol{\theta}}_{WLS} = \arg\min_{\boldsymbol{\theta}} \ \mathbf{r}(\boldsymbol{\theta})^{\top} \left(\mathbf{W}^{-1} \otimes \mathbf{I}_T\right) \mathbf{r}(\boldsymbol{\theta}) \tag{4.35}$$

  The name Weighted Least Squares is used because each squared residual is weighted by the inverse of its variance.

- reduces to Ordinary Least Squares if $\text{Cov}(\mathbf{r}) = \sigma^2 \mathbf{I}_{TN}$, that is, under the assumption of homoscedasticity and no autocorrelation;

This class of estimators (GLS) has better properties than OLS with non-spherical errors defined in equation 4.32. We derive the single equation linear estimator for the GLS method[7]:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^{\top} \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^{\top} \boldsymbol{\Omega}^{-1} \mathbf{y} \tag{4.36}$$

In fact, if $\boldsymbol{\Omega}$ is symmetric and positive definite, there is an invertible matrix $\boldsymbol{\Gamma}^{1/2}$ such that

[7] The derivation below applies to a linear system, where the relationship between the regressors and the parameters is linear: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and the design matrix $\mathbf{X}$ does not depend on $\boldsymbol{\beta}$. In nonlinear systems, because the residuals $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta})$ depend on the parameter vector $\boldsymbol{\theta}$ in a nonlinear way, the role of the design matrix $\mathbf{X}$ is played by the Jacobian matrix: $\mathbf{J} = \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}$. Nonlinear GLS is derived in the next section.

$\mathbf{\Omega} = \mathbf{\Gamma}^{1/2}(\mathbf{\Gamma}^{1/2})^\top$. Then, by multiplying the regression equation 4.28 by $\mathbf{\Gamma}^{-1/2}$, we can obtain the transformed regression equation:

$$\acute{\mathbf{y}} = \acute{\mathbf{X}}\boldsymbol{\beta} + \acute{\boldsymbol{\varepsilon}} \tag{4.37}$$

$$\acute{\mathbf{y}} = \mathbf{\Gamma}^{-1/2}\mathbf{y}; \quad \acute{\mathbf{X}} = \mathbf{\Gamma}^{-1/2}\mathbf{X}; \quad \acute{\boldsymbol{\varepsilon}} = \mathbf{\Gamma}^{-1/2}\boldsymbol{\varepsilon} \tag{4.38}$$

We assume $N = 1$, we omit the index i. This linear transformation is known as *whitening* (i.e., imposing white noise properties), as it transforms the error vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{\Gamma})$ into a new error term $\acute{\boldsymbol{\varepsilon}} \sim \mathcal{N}(0, \mathbf{I})$ with identity covariance. The transformation removes heteroscedasticity and autocorrelation from the disturbances.

The OLS estimator of the transformed regression equation is the GLS estimator, that also solves the so-called generalized least squares problem (see Pericoli and Taboga (2012) for details of the proof).

$$\hat{\boldsymbol{\beta}}_{GLS} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4.39}$$

$$\hat{\boldsymbol{\beta}}_{GLS} = (\acute{\mathbf{X}}^\top \acute{\mathbf{X}})^{-1} \acute{\mathbf{X}}^\top \acute{\mathbf{y}} \tag{4.40}$$

We substitute into the GLS estimator:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left((\mathbf{\Gamma}^{-1/2}\mathbf{X})^\top (\mathbf{\Gamma}^{-1/2}\mathbf{X}) \right)^{-1} (\mathbf{\Gamma}^{-1/2}\mathbf{X})^\top (\mathbf{\Gamma}^{-1/2}\mathbf{y})$$

Using the transpose identity $(AB)^\top = B^\top A^\top$:

$$= \left(\mathbf{X}^\top (\mathbf{\Gamma}^{-1/2})^\top \mathbf{\Gamma}^{-1/2} \mathbf{X} \right)^{-1} \mathbf{X}^\top (\mathbf{\Gamma}^{-1/2})^\top \mathbf{\Gamma}^{-1/2} \mathbf{y}$$

$\mathbf{\Gamma}$ is symmetric being a covariance matrix, then: $\mathbf{\Gamma}^{-1/2} = (\mathbf{\Gamma}^{-1/2})^\top$, so:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{y} \tag{4.41}$$

In practice we rarely know the error covariance matrix $\mathbf{\Omega}$ so we have to replace it with an estimator $\widehat{\mathbf{\Omega}} = \widehat{\text{Cov}}(\mathbf{r})$, to obtain the so-called Feasible Generalized Least Squares estimator (FGLS). In the case of Seemingly Unrelated Regressions (SUR), this estimator takes the form $\widehat{\mathbf{\Omega}} = \mathbf{S} \otimes \mathbf{I}_T$, where $\mathbf{S}$ is the $N \times N$ matrix of estimated contemporaneous covariances across equations.

It is normal practice to estimate $\mathbf{S}$ by means of the residuals of the first-step OLS regression.

Nonlinear SUR (NSUR) is based on the same two-step logic: the first step uses OLS residuals to estimate the variance–covariance of disturbances, and the second step applies FGLS.

The estimator for a system of linear equations is obtained via equation 4.29 and 4.30:

$$\widehat{\text{vec}(\boldsymbol{B})}_{SUR} = \left((\mathbf{I}_N \otimes \mathbf{X})^\top \left(\mathbf{S}^{-1} \otimes \mathbf{I_T}\right)(\mathbf{I}_N \otimes \mathbf{X})\right)^{-1} (\mathbf{I}_N \otimes \mathbf{X})^\top \left(\mathbf{S}^{-1} \otimes \mathbf{I_T}\right) \mathbf{y} \tag{4.42}$$

where:

- $\mathbf{I}_N \otimes \mathbf{X} \in \mathbb{R}^{NT \times N\bar{\ell}}$ is the system design matrix that extends the regressor $\mathbf{X}$ for each equation;
- $\mathbf{S}^{-1} \otimes \mathbf{I}_T \in \mathbb{R}^{NT \times NT}$ is the SUR/GLS weighting matrix that assumes heteroscedasticity, cross-equation residual correlation and no autocorrelation over time;
- $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{R}^{NT \times 1}$ is the stacked portfolios vector.

For a system of nonlinear equation, SAS provides 4.34, which we derive in the next sections and we use to estimate the parameter vector in equation 2.50 for the NSUR results.

H Nonlinear Least Squares Gauss-Newton

The Gauss-Newton nonlinear least square method is based on a first-order Taylor expansion around the starting value of the function:

$$f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{4.43}$$

where $\mathbf{J}(\boldsymbol{\theta}_0) = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}_0} \in \mathbb{R}^{TN \times \bar{k}}$ is the Jacobian matrix of f evaluated at $\boldsymbol{\theta}_0$.

Then, the residual function becomes:

$$\begin{aligned} \mathbf{r}(\boldsymbol{\theta}) &= \mathbf{y} - f(\boldsymbol{\theta}) \\ &\approx \mathbf{y} - [f(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] \\ &= [\mathbf{y} - f(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0] - \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta} \end{aligned} \tag{4.44}$$

We define a pseudo-response vector constructed to centre the nonlinear model around $\boldsymbol{\theta}_0$, which enables linear estimation techniques (GLS):

$$\mathbf{y}^* := \mathbf{y} - f(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 \tag{4.45}$$

So that the linearized residual becomes:

$$\mathbf{r}(\boldsymbol{\theta}) \approx \mathbf{y}^* - \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta} \tag{4.46}$$

Substituting into the SUR original objective:

$$\hat{\boldsymbol{\theta}}_{\text{SUR}} \approx \arg\min_{\boldsymbol{\theta}} (\mathbf{y}^* - \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta})^\top (\mathbf{S}^{-1} \otimes \mathbf{I}_T)(\mathbf{y}^* - \mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\theta}). \tag{4.47}$$

We can also replace $\mathbf{X}$ with $\mathbf{J}$ in the linear estimator 4.36 as equation 4.47 is analogue to equation 4.33.

The nonlinear single equation GLS estimator is:

$$\hat{\boldsymbol{\theta}}_{\text{GLS}} = \left(\mathbf{J}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Omega}^{-1} \mathbf{J}(\boldsymbol{\theta}_0)\right)^{-1} \mathbf{J}(\boldsymbol{\theta}_0)^\top \boldsymbol{\Omega}^{-1} \mathbf{y}^* \tag{4.48}$$

If $\boldsymbol{\theta}_0$ is updated iteratively using $\hat{\boldsymbol{\theta}}$, the procedure corresponds to iteratively reweighted nonlinear GLS. Here below, we write the formula for a system of nonlinear equations N:

$$\hat{\boldsymbol{\theta}}_{\text{NSUR}} = \left(\mathbf{J}^\top (\mathbf{S}^{-1} \otimes \mathbf{I}_T)\mathbf{J}\right)^{-1} \mathbf{J}^\top (\mathbf{S}^{-1} \otimes \mathbf{I}_T)\mathbf{y}^*) \tag{4.49}$$

I Nonlinear Least Squares SAS ITSUR

SAS Feasible Generalized Least Squares (FGLS) estimator (`ITSUR` method in `PROC MODEL`) iteratively updates both the parameter vector and the cross-equation covariance matrix without linearizing the model.

The parameter vector $\boldsymbol{\theta}^{(0)}$, is initialised using OLS.

The following steps are run for each iteration $m = 0, 1, 2, \ldots$, until convergence:

- Residuals are calculated: $\mathbf{r}^{(m)} = \mathbf{y} - f(\boldsymbol{\theta}^{(m)})$.
- The cross-equation error covariance matrix $\mathbf{S}^{(m)} \in \mathbb{R}^{N \times N}$ is estimated from the residuals:

$$\mathbf{S}^{(m)} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t^{(m)} \boldsymbol{\varepsilon}_t^{(m)\top} \tag{4.50}$$

  where $\boldsymbol{\varepsilon}_t^{(m)} \in \mathbb{R}^N$ stacks residuals across equations at time t.
- The Jacobian matrix of the nonlinear function is evaluated numerically using finite

difference methods, unless analytic derivatives are explicitly provided:

$$\mathbf{J}^{(m)} = \left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}^{(m)}} \in \mathbb{R}^{TN \times \bar{k}}$$

- The parameter vector is updated using the FGLS estimator and the pseudo–response $\mathbf{y}^{*(m)} = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^{(m)}) + \mathbf{J}^{(m)}\boldsymbol{\theta}^{(m)}$:

$$\boldsymbol{\theta}_{NSUR}^{(m+1)} = \left(\mathbf{J}^{(m)\top}(\mathbf{S}^{(m)^{-1}} \otimes \mathbf{I}_T)\mathbf{J}^{(m)}\right)^{-1} \mathbf{J}^{(m)\top}(\mathbf{S}^{(m)^{-1}} \otimes \mathbf{I}_T)\mathbf{y}^{*(m)}$$

If $\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}\|$ and the change in $\mathbf{S}$ are below a given tolerance, the algorithm terminates.

J Inference

The covariance matrix is a fundamental element in statistical inference. It is used to determine whether the regression coefficients are statistically significant and to construct confidence intervals.

First we define $\mathbf{X}^+$, the Moore–Penrose pseudoinverse of $\mathbf{X}$. If $\mathbf{X} \in \mathbb{R}^{T \times \bar{\ell}}$ has full column rank ℓ, then:

$$\mathbf{X}^+ = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top.$$

We assume that $\mathbf{X}$ is full rank and we write the OLS estimator as a linear function of the error:

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^+(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + \mathbf{X}^+\boldsymbol{\varepsilon} \tag{4.51}$$

and derive the variance-covariance (Cov) matrix of the OLS estimator:

$$\text{Cov}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top] = \mathbb{E}[(\mathbf{X}^+\boldsymbol{\varepsilon})(\mathbf{X}^+\boldsymbol{\varepsilon})^\top] \tag{4.52}$$

and:

$$\mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top] = \mathbf{X}^+ \, \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top] \, (\mathbf{X}^+)^\top = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1} \tag{4.53}$$

The main OLS assumptions are the homoscedasticity of the errors and their lack of autocorrelation. In practice, we cannot observe the error terms $\boldsymbol{\varepsilon}$, nor their variance σ^2. Instead, we can rely on the residual errors of the sample to compute the variance s^2. Then:

$$\text{Cov}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = s^2(\mathbf{X}^\top \mathbf{X})^{-1} \tag{4.54}$$

The general form is based on nonlinear least squares theory and use the Taylor approximation of the residual around the starting value:

$$\mathbf{r}(\boldsymbol{\theta}) \approx \mathbf{r}(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

The corresponding quadratic approximation of the objective becomes:

$$Q(\boldsymbol{\theta}) = \left(\mathbf{r}(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right)^\top \left(\mathbf{r}(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right)$$

Taking the derivative:

$$\frac{\partial Q}{\partial \boldsymbol{\theta}} = 2\mathbf{J}^\top \left(\mathbf{r}(\boldsymbol{\theta}_0) + \mathbf{J}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) = 0$$

and solving for $\boldsymbol{\theta}$:

$$\mathbf{J}^\top \mathbf{r}(\boldsymbol{\theta}_0) + \mathbf{J}^\top \mathbf{J}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = 0$$

We obtain the Gauss–Newton update step

$$\boldsymbol{\theta} - \boldsymbol{\theta}_0 = -(\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \mathbf{r}(\boldsymbol{\theta}_0)$$

Now, take the covariance on both sides:

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) = \mathrm{Cov}\left(-(\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \mathbf{r}(\boldsymbol{\theta}_0)\right)$$

Using the identity $\mathrm{Cov}(\mathbf{Az}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{z})\,\mathbf{A}^\top$, we get the general form:

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) = (\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top\, \mathrm{Cov}(\mathbf{r})\, \mathbf{J} (\mathbf{J}^\top \mathbf{J})^{-1}$$

which simplifies to equation 4.54 when:

$$\mathrm{Cov}(\mathbf{r}) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$$

However, while no error autocorrelation is expected in cross-sectional data, homoscedasticity is rarely satisfied: the SAS OLS covariance estimator relies on Weighted Least Squares (WLS) where $\mathbf{S} \in \mathbb{R}^{N \times N}$ has the form $\boldsymbol{\Omega}^{-1} = \mathrm{diag}(\mathbf{S})^{-1} \otimes \mathbf{I}_T$, and $\mathbf{W} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$, which corresponds to heteroscedasticity across equations, constant variance over time, and no autocorrelation. SAS replaces $\mathrm{Cov}(\mathbf{r}) = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top]$ with the diagonal approximation, $\mathrm{diag}(\mathbf{S}) \otimes \mathbf{I}_T$.

Instead of using the standard OLS "sandwich" variance estimator (equation 4.53 and 4.54), the Generalized Least Squares estimator solves the weighted least squares problem. It seeks a more efficient estimator by pre-whitening the residuals using as weighting matrix, the inverse of their covariance structure, $\mathbf{W} = \mathbf{\Omega}^{-1}$, where: SUR corresponds to the special case with no serial correlation, $\mathbf{\Omega} = \mathbf{S} \otimes \mathbf{I}_T$, and SAS "OLS" to diagonal WLS weighting, $\mathbf{W} = \left(\text{diag}(\mathbf{S})\right)^{-1} \otimes \mathbf{I}_T$..

In general, the GLS objective replaces the unweighted least squares objective with a weighted version:

$$\hat{\boldsymbol{\theta}}_{GLS} = \arg\min_{\boldsymbol{\theta}} \ \mathbf{r}(\boldsymbol{\theta})^\top \mathbf{W}\, \mathbf{r}(\boldsymbol{\theta}) \tag{4.55}$$

Then, we take the derivative of the objective with respect to $\boldsymbol{\theta}$ and set it to zero:

$$\mathbf{J}^\top \mathbf{W}\, \mathbf{r}(\boldsymbol{\theta}) = 0$$

Using Taylor approximation as we have done in the general form, we obtain:

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left(\mathbf{J}^\top \mathbf{W} \mathbf{J}\right)^{-1} \mathbf{J}^\top \mathbf{W} \mathbf{r}(\boldsymbol{\theta}_0)$$

Now, we take the covariance on both sides:

$$\text{Cov}(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) = \left(\mathbf{J}^\top \mathbf{W} \mathbf{J}\right)^{-1} \mathbf{J}^\top \mathbf{W}\, \text{Cov}(\mathbf{r})\, \mathbf{W} \mathbf{J} \left(\mathbf{J}^\top \mathbf{W} \mathbf{J}\right)^{-1}$$

Since $\mathbf{W} = \text{Cov}(\mathbf{r})^{-1}$, this simplifies to [8] :

$$\text{Cov}(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) = \left(\mathbf{J}^\top \mathbf{W} \mathbf{J}\right)^{-1} \tag{4.56}$$

SAS OLS form is based on WLS for multiple equations under heteroscedasticity assumptions without autocorrelation:

$$\text{Cov}_{WLS}(\hat{\boldsymbol{\theta}} \mid \mathbf{X}) = \left(\mathbf{J}^\top \left(\text{diag}(\mathbf{S})^{-1} \otimes \mathbf{I}_T\right) \mathbf{J}\right)^{-1} \tag{4.57}$$

Heteroscedasticity does not cause problems for estimating $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$, which does not rely on any assumptions about the disturbances. The estimator is not efficient, but it is unbiased. However, it causes problems when computing the *correct* standard errors for hypothesis testing, i.e., when using: $\hat{\boldsymbol{\beta}} \approx \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$, as we need to make assumptions about the disturbance process. In the presence of heteroscedasticity and error correlation, it is not true that: $\mathbf{X}^+ \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top] \mathbf{X}^{+\top} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$ and we should use equation

[8] We use the matrix identity $\left(\mathbf{A}^{-1} \mathbf{A} \mathbf{A}^{-1}\right) = \mathbf{A}^{-1}$, which holds when $\mathbf{A}$ is square and invertible.

4.32 for the error covariance.

The SUR covariance matrix estimator is given by:

$$\text{Cov}_{SUR}(\hat{\boldsymbol{\theta}}) = \left(\mathbf{J}^{\top}(\mathbf{S}^{-1} \otimes \mathbf{I}_T)\mathbf{J}\right)^{-1} \tag{4.58}$$

where $\mathbf{S}$ is the estimated covariance matrix of the residuals evaluated at $\hat{\boldsymbol{\theta}}$.

K Generalized Method of Moments

In the Generalized Method of Moments (GMM), a number of moment conditions are specified for the model as functions of the parameters and the data [9]:

$$\mathbf{g}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{f}(\mathbf{w}, \mathbf{z}, \boldsymbol{\theta})] \tag{4.59}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{\bar{k}}$ is a vector of parameters with true value $\boldsymbol{\theta}_0$ and $\mathbf{f}(\cdot) \in \mathbb{R}^{\bar{g}}$ is a (generally nonlinear) function vector. At time t, the vector $\mathbf{w}_t$ contains the endogenous and exogenous model variables ($\mathbf{y}_t \in \mathbb{R}^N, \mathbf{x}_t \in \mathbb{R}^{\bar{\ell}}$) and $\mathbf{z}_t \in \mathbb{R}^{\bar{p}}$ are the instrumental variables.

The true parameter vector value $\boldsymbol{\theta}_0$ in Equation 4.59 is found via the moment condition expectation. The system is identified (identification can be derived by model construction or from the data) if there is a unique solution $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for $\mathbf{g}(\boldsymbol{\theta}) = \mathbb{E}(\mathbf{f}(\mathbf{w}_t, \mathbf{z}_t, \boldsymbol{\theta}))= 0$. Since the population expectation is in general unknown, it is replaced by its sample average:

$$\hat{\mathbf{g}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}(\mathbf{w}_t, \mathbf{z}_t, \boldsymbol{\theta}) \tag{4.60}$$

In the next pages, we will define the GMM estimator $\hat{\boldsymbol{\theta}}_{GMM}$ as the parameter value that minimises the norm of $\hat{\mathbf{g}}_T(\boldsymbol{\theta})$. Under regularity conditions, $\hat{\boldsymbol{\theta}}_{GMM}$ converges in probability to $\boldsymbol{\theta}_0$ as $T \to \infty$.

In many applications the population moment condition has the form: $\mathbf{f}(\mathbf{w}_t, \mathbf{z}_t, \boldsymbol{\theta}) = \boldsymbol{\varepsilon}(\mathbf{w}_t, \boldsymbol{\theta}) \otimes \mathbf{z}_t$ where the $\bar{g}$ moment functions ($\bar{g} = N \times \bar{p}$) are derived multiplying the model residuals $\boldsymbol{\varepsilon}(\mathbf{w}_t, \boldsymbol{\theta})$ by the instruments $\mathbf{z}_t$:

$$\mathbf{g}(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\varepsilon}(\mathbf{w}, \boldsymbol{\theta}) \otimes \mathbf{z}] = 0 \tag{4.61}$$

In this class of estimator called instrumental variable (IV), the instruments are uncorrelated

[9] The reader should refer to Section 2.4 for the notation explanation in vector and matrix form.

with the error term of the model.

A unique estimator is found when there are at least as many equations $\bar{g} = N \times \bar{p}$ as parameters $\bar{k}$, in detail:

- When $\bar{g} = \bar{k}$ exact identification, it is named the Method of Moments estimator, $\hat{\boldsymbol{\theta}}_{MM}$
- When $\bar{g} > \bar{k}$, over identification, it is named the Generalized Method of Moments estimator, $\hat{\boldsymbol{\theta}}_{GMM}$.
- When $\bar{g} < \bar{k}$, under identification, the parameters are not identified and cannot be estimated consistently.

The solution can be seen as a special case of the minimum-distance estimation. We derive a quadratic form $Q_T(\boldsymbol{\theta})$, where the GMM estimator minimizes a certain distance of the sample averages of the moment conditions. The properties of the subsequent estimator depend on the choice of the norm function, the GMM theory considers a family of L^2-type norms, defined as:

$$Q_T(\boldsymbol{\theta}) = \|\hat{\mathbf{g}}_T(\boldsymbol{\theta})\|^2_{\mathbf{W}} = \hat{\mathbf{g}}_T(\boldsymbol{\theta})^\top \mathbf{W} \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \tag{4.62}$$

where $\mathbf{W}$ is a positive definite weighting matrix.

We define the GMM estimator via the standard objective function:

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}} \hat{\mathbf{g}}_T^\top(\boldsymbol{\theta})\, \mathbf{W}\, \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \tag{4.63}$$

In Section 2.4, we also present the SAS GMM objective in an equivalent scaled form:

$$\frac{1}{T} \left[T \hat{\mathbf{g}}_T(\boldsymbol{\theta})\right]^\top \hat{\mathbf{V}}^{-1} \left[T \hat{\mathbf{g}}_T(\boldsymbol{\theta})\right] \tag{4.64}$$

where the weighting matrix is defined as:

$$\mathbf{V} = T\boldsymbol{\Lambda}, \qquad \text{with } \boldsymbol{\Lambda} = \mathbb{E}[\mathbf{g}_t \mathbf{g}_t^\top] \tag{4.65}$$

Since $\boldsymbol{\Lambda}$ is unknown, we estimate it with the sample covariance:

$$\hat{\boldsymbol{\Lambda}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t^\top \tag{4.66}$$

However, SAS defines:

$$\hat{\mathbf{V}} = T\hat{\mathbf{\Lambda}} = \sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t^\top \tag{4.67}$$

The T scaling of the objective in Equation (4.63) ensures consistency with large-sample asymptotics. Both expressions are equivalent

$$\hat{\mathbf{g}}_T(\boldsymbol{\theta})^\top \, \hat{\mathbf{\Lambda}}^{-1} \, \hat{\mathbf{g}}_T(\boldsymbol{\theta}) = T \, \hat{\mathbf{g}}_T(\boldsymbol{\theta})^\top \, \hat{\mathbf{V}}^{-1} \, \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \tag{4.68}$$

From 4.63, to find the first-order condition, we take the derivative of the quadratic form $Q(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$:

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \left(\hat{\mathbf{g}}_T^\top(\boldsymbol{\theta}) \mathbf{W} \hat{\mathbf{g}}_T(\boldsymbol{\theta}) \right) \tag{4.69}$$

Using the rule for differentiating quadratic forms:

$$\frac{\partial}{\partial \boldsymbol{\theta}} (\mathbf{a}^\top \mathbf{M} \ \mathbf{a}) = 2\mathbf{a}^\top \mathbf{M} \frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} \tag{4.70}$$

where $\mathbf{a}$ is a function of $\boldsymbol{\theta}$ and $\mathbf{M}$ is a symmetric matrix, we obtain:

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\hat{\mathbf{g}}(\boldsymbol{\theta})^\top \mathbf{W} \frac{\partial \hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{4.71}$$

We define the Jacobian matrix of the moment conditions and apply the interchange rule:

$$\mathbf{G}(\boldsymbol{\theta}) = \frac{\partial \mathbb{E}[\mathbf{g}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^\top}, \quad \hat{\mathbf{G}}(\boldsymbol{\theta}) = \frac{\partial \hat{\mathbf{g}}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \in \mathbb{R}^{\bar{g} \times \bar{k}} \tag{4.72}$$

then, we set the first-order condition 4.71 to zero:

$$\hat{\mathbf{g}}(\boldsymbol{\theta})^\top \mathbf{W} \hat{\mathbf{G}}(\boldsymbol{\theta}) = 0 \tag{4.73}$$

Rearranging, we obtain:

$$\hat{\mathbf{G}}^\top \mathbf{W} \hat{\mathbf{g}} = 0. \tag{4.74}$$

The GMM estimator for linear moments[10] is found by solving for $\boldsymbol{\theta}$, multiplying both sides by the inverse of $\hat{\mathbf{G}}^\top \mathbf{W} \hat{\mathbf{G}}$:

$$\hat{\boldsymbol{\theta}}_{GMM} = \left(\hat{\mathbf{G}}^\top \mathbf{W} \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^\top \mathbf{W} \hat{\mathbf{g}} \tag{4.75}$$

[10] For non linear moments:
$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \Delta\boldsymbol{\theta}^{(m)}$, and $\Delta\boldsymbol{\theta}^{(m)} = \left(\hat{\mathbf{G}}(\boldsymbol{\theta}^{(m)})^\top \mathbf{W}^{(m)} \hat{\mathbf{G}}(\boldsymbol{\theta}^{(m)})\right)^{-1} \hat{\mathbf{G}}(\boldsymbol{\theta}^{(m)})^\top \mathbf{W}^{(m)} \hat{\mathbf{g}}_T(\boldsymbol{\theta}^{(m)})$.

From equation 4.27, the residual vector is defined as:

$$\varepsilon(\boldsymbol{B}) := \mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X}) \cdot \text{vec}(\boldsymbol{B}),$$

the sample moment function becomes:

$$\hat{\mathbf{g}} = \frac{1}{T}(\mathbf{I}_N \otimes \mathbf{Z}^\top)\varepsilon(\boldsymbol{B}) \quad \in \mathbb{R}^{N\bar{p}} \tag{4.76}$$

where: $\mathbf{I}_N \otimes \mathbf{Z}^\top \in \mathbb{R}^{N\bar{p}\times TN}$, and $\varepsilon(\boldsymbol{B}) \in \mathbb{R}^{TN}$.

We define the matrix $\mathbf{H} \in \mathbb{R}^{N\bar{p}\times NT}$ as the Kronecker product below:

$$\mathbf{H} := \mathbf{I}_N \otimes \mathbf{Z}^\top. \tag{4.77}$$

so that:

$$\hat{\mathbf{g}} = \frac{1}{T}\mathbf{H}\varepsilon(\boldsymbol{B})$$

Differentiating, we obtain the Jacobian with respect to $\text{vec}(\boldsymbol{B})$:

$$\hat{\mathbf{G}} = \frac{\partial \hat{\mathbf{g}}}{\partial \,\text{vec}(\boldsymbol{B})^\top} = -\frac{1}{T}\mathbf{H}\,(\mathbf{I}_N \otimes \mathbf{X})$$

Substituting $\hat{\mathbf{g}} = \frac{1}{T}\mathbf{H}(\mathbf{y} - (\mathbf{I}_N \otimes \mathbf{X})\,\text{vec}(\boldsymbol{B}))$ and $\hat{\mathbf{G}} = -\frac{1}{T}\mathbf{H}(\mathbf{I}_N \otimes \mathbf{X})$ into Equation (4.75), we obtain the GMM estimator for a system of linear equations [11]:

$$\widehat{\text{vec}(\boldsymbol{B})}_{GMM} = \left((\mathbf{I}_N \otimes \mathbf{X})^\top \mathbf{H}^\top \mathbf{W}\, \mathbf{H}(\mathbf{I}_N \otimes \mathbf{X})\right)^{-1} (\mathbf{I}_N \otimes \mathbf{X})^\top \mathbf{H}^\top \mathbf{W}\, \mathbf{H}\mathbf{y}. \tag{4.78}$$

where $\mathbf{H} := \mathbf{I}_N \otimes \mathbf{Z}^\top$, with $\mathbf{Z} \in \mathbb{R}^{T\times\bar{p}}$ and $\mathbf{X} \in \mathbb{R}^{T\times\bar{\ell}}$.

Substituting $\mathbf{H} = \mathbf{I}_N \otimes \mathbf{Z}^\top$, we obtain:

$$\begin{aligned}\widehat{\text{vec}(\boldsymbol{B})}_{GMM} = &\left((\mathbf{I}_N \otimes \mathbf{X})^\top (\mathbf{I}_N \otimes \mathbf{Z})\, \mathbf{W}\, (\mathbf{I}_N \otimes \mathbf{Z}^\top)(\mathbf{I}_N \otimes \mathbf{X})\right)^{-1} \\ &\times (\mathbf{I}_N \otimes \mathbf{X})^\top (\mathbf{I}_N \otimes \mathbf{Z})\, \mathbf{W}\, (\mathbf{I}_N \otimes \mathbf{Z}^\top)\mathbf{y}.\end{aligned}$$

We now define the corresponding stacked matrices in Wooldridge's notation:

- $Y := \mathbf{y} \in \mathbb{R}^{TN}$: stacked outcome vector, $Y = \begin{bmatrix} y_1^\top & \cdots & y_N^\top \end{bmatrix}^\top$.

[11] For nonlinear models, it is typically not possible to solve for $\hat{\boldsymbol{\theta}}_{GMM}$ analytically. Numerical optimisation methods are used to find the parameter vector that minimizes the first-order conditions.

- $X := \mathbf{I}_N \otimes \mathbf{X} \in \mathbb{R}^{TN \times N\bar{\ell}}$: block-diagonal regressor matrix.
- $Z := \mathbf{I}_N \otimes \mathbf{Z}^\top \in \mathbb{R}^{N\bar{p} \times TN}$: block-diagonal instrument matrix.
- $\boldsymbol{b} := \mathsf{vec}(\boldsymbol{B}) \in \mathbb{R}^{N\bar{\ell}}$: stacked parameter vector.

We keep the bold vector convention for moments:

$$\hat{\mathbf{g}}(\boldsymbol{b}) = \frac{1}{T} H \boldsymbol{\varepsilon}(\boldsymbol{b}) \quad \text{with} \quad H := I_N \otimes Z^\top.$$

(If Wooldridge's non-bold notation appears, $\hat{g}_T \equiv \hat{\mathbf{g}}$.)

Using these definitions, we rewrite the GMM estimator without Kronecker notation:

$$\hat{\boldsymbol{b}}_{GMM} = (X^\top Z W Z^\top X)^{-1} X^\top Z W Z^\top Y \tag{4.79}$$

which is the matrix form used in Wooldridge (Equation 8.24).

The following formulas are shown in Wooldridge notation.

If $\bar{g} = \bar{k}$, just identified case, $X^\top Z$ is a square matrix, then we obtain:

$$(X^\top Z W Z^\top X)^{-1} = (Z^\top X)^{-1} W^{-1} (X^\top Z)^{-1}$$

The IV estimator, which is independent of W, can be derived from the GMM estimator:

$$\hat{\boldsymbol{b}}_{IV} = (Z^\top X)^{-1} W^{-1} (X^\top Z)^{-1} X^\top Z \, W \, Z^\top Y = (Z^\top X)^{-1} Z^\top Y \tag{4.80}$$

We can derive the OLS estimator by setting $Z = X$:

$$\hat{\boldsymbol{b}}_{OLS} = (X^\top X)^{-1} X^\top Y. \tag{4.81}$$

Conversely, the GMM estimator depends on the choice of the weighting matrix. If there is over identification, $\bar{g} > \bar{k}$, and the rank of $\mathbb{E}[(\mathbf{I}_N \otimes \mathbf{Z}^\top)(\mathbf{I}_N \otimes \mathbf{X})] = N\bar{\ell} = \bar{k}$ then the matrix of moment conditions $\mathbb{E}[\mathbf{z} \otimes \boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}_{N\bar{p}}$ implies consistency of $\hat{\boldsymbol{\theta}}_{GMM}$: as the sample size $T \to \infty$, the estimator converges in probability to the true parameter value. [12]

We set $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^\top & \cdots & \mathbf{z}_T^\top \end{bmatrix}^\top$, with $\mathbf{z}_t \in \mathbb{R}^{\bar{p}}$, then we can expand the sample moment

[12] See Wooldridge page 186, System Instrumental variables (SIV) Assumption 2.

function:

$$\hat{\mathbf{g}}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}(\mathbf{z}_t, \boldsymbol{b}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{b}) \otimes \mathbf{z}_t. \tag{4.82}$$

We write the estimation error of the GMM estimator:

$$\hat{\boldsymbol{b}}_{GMM} - \boldsymbol{b}_{GMM} = \left(X^\top H^\top W H X\right)^{-1} X^\top H^\top W \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_t. \tag{4.83}$$

We want to show that the asymptotic covariance of the GMM estimator is:

$$\mathsf{Avar}(\hat{\boldsymbol{b}}_{GMM}) = \frac{1}{T}\left(G^\top W G\right)^{-1} G^\top W V W G \left(G^\top W G\right)^{-1} \tag{4.84}$$

which is equivalent to:

$$\begin{aligned}\mathsf{Avar}(\hat{\boldsymbol{b}}_{GMM}) = \frac{1}{T}&\left((HX)^\top W (HX)\right)^{-1} (HX)^\top W V W (HX) \\ &\times \left((HX)^\top W (HX)\right)^{-1}.\end{aligned} \tag{4.85}$$

We start from the asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{b}}_{GMM} - \boldsymbol{b}_{GMM})$:

$$\sqrt{T}(\hat{\boldsymbol{b}}_{GMM} - \boldsymbol{b}_{GMM}) \xrightarrow{d} \mathcal{N}\left(0, \mathsf{Avar}(\hat{\boldsymbol{b}}_{GMM})\right) \tag{4.86}$$

where $\xrightarrow{d}$ denotes convergence in distribution[13]. Depending on the source, either $\mathsf{Var}(\hat{\boldsymbol{b}}_{GMM})$ or the scaled matrix $\mathsf{Avar}(\hat{\boldsymbol{b}}_{GMM}) := T \cdot \mathsf{Var}(\hat{\boldsymbol{b}}_{GMM})$ is referred to as the asymptotic covariance of the estimator [14].

In practice, convergence does not occur at any finite sample size T, so this value is only an approximation of the true covariance of the estimator. In the limit, the covariance of the estimator satisfies $\mathsf{Var}(\hat{\boldsymbol{b}}_{GMM}) \to 0$, while the scaled quantity

$$\mathsf{Avar}(\hat{\boldsymbol{b}}_{GMM}) = \lim_{T\to\infty} T \cdot \mathsf{Var}(\hat{\boldsymbol{b}}_{GMM}) \tag{4.87}$$

represents a consistent approximation for inference when T is large [15]

Substituting $\hat{\Lambda} = \frac{1}{T}\sum_{t=1}^{T} g_t g_t^\top \in \mathbb{R}^{N\bar{p}\times N\bar{p}}$ into the covariance formula, we obtain the

[13] See Wooldridge, pp. 38 and 191, for the definition of convergence in distribution.

[14] See Wooldridge, p. 40, p. 55, where $\hat{\boldsymbol{b}}_{GMM} \sim \mathcal{N}(\boldsymbol{b}, \mathsf{Var}(\hat{\boldsymbol{b}}_{GMM})/T)$.

[15] Although $\mathsf{Var}(\hat{\boldsymbol{b}}_{GMM}) \to 0$ as $T \to \infty$, the scaled quantity $T \cdot \mathsf{Var}(\hat{\boldsymbol{b}}_{GMM})$ converges to a finite matrix, which defines the asymptotic covariance.

sample covariance estimator:

$$\widehat{\text{Avar}}(\hat{\boldsymbol{b}}_{GMM}) = \frac{1}{T}(\hat{G}^\top W \hat{G})^{-1}\hat{G}^\top W \hat{\Lambda} W \hat{G}(\hat{G}^\top W \hat{G})^{-1} \tag{4.88}$$

W is the weighting matrix, often chosen as $W = \Lambda^{-1}$ for efficiency.

From the estimation error equation 4.83, the covariance corresponds to:

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{b}}_{GMM}) &= \mathbb{E}\Big[(\hat{\boldsymbol{b}}_{GMM} - \boldsymbol{b}_{GMM})(\hat{\boldsymbol{b}}_{GMM} - \boldsymbol{b}_{GMM})^\top\Big] \\
&= (X^\top H^\top W H X)^{-1} X^\top H^\top W \cdot \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\right)\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{g}_s\right)^\top\right] \\
&\quad \cdot W H X\,(X^\top H^\top W H X)^{-1} \\
&= \frac{1}{T}\,(X^\top H^\top W H X)^{-1} X^\top H^\top W\,\hat{\Lambda}\,W H X (X^\top H^\top W H X)^{-1}.
\end{aligned} \tag{4.89}$$

where:

$$\begin{aligned}
\mathbb{E}[\hat{\mathbf{g}}_T \hat{\mathbf{g}}_T^\top] &= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t\right)\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{g}_s\right)^\top\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\mathbf{g}_t \mathbf{g}_s^\top] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\Lambda = \frac{1}{T}\,\Lambda,
\end{aligned}$$

Using the optimal weighting matrix $W = \hat{\Lambda}^{-1}$, moments with a lower covariance are assigned a greater weight as they are more informative, we obtain from the full "sandwich" form above the efficient GMM form:

$$\text{Var}(\hat{b}_{GMM}) = \frac{1}{T}\left((HX)^\top \Lambda^{-1}(HX)\right)^{-1} \tag{4.90}$$

From equation 4.85, applying the optimal matrix we obtain:

$$\text{Avar}(\hat{b}_{GMM}) = \left((HX)^\top \Lambda^{-1}(HX)\right)^{-1} \tag{4.91}$$

For nonlinear systems the SAS GMM covariance becomes[16]:

$$\text{Var}(\hat{\boldsymbol{\theta}}_{GMM}) = ((HJ)^\top \hat{V}^{-1}(HJ))^{-1} \tag{4.92}$$

SAS moment conditions are formed as sums $T\hat{\mathbf{g}}_T = \sum_{t=1}^{T} \mathbf{g}_t$ rather than averages. Therefore, the weighting matrix is defined as $\hat{V} = \sum_{t=1}^{T} \mathbf{g}_t \mathbf{g}_t^\top = T\hat{\Lambda}$. This scaling leads to the equivalent covariance expression 4.92, written in terms of $\hat{V}$ instead of $\hat{\Lambda}$.

The choice of the weight matrix depend on the moments being independent and identically distributed (i.i.d); independent but heteroscedastic; autocorrelated and heteroscedastic.
For example, we have proven that the MM estimator and the OLS estimator are equivalent. However, it is important to note that the default covariance of the OLS estimator is based on homoscedasticity and absence of error autocorrelation assumptions. In the case of a single equation, so for time-series regression:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} \tag{4.93}$$

It means that in the presence of error autocorrelation and heteroscedasticity, we need a method to estimate the disturbances covariances. One technique to build a robust covariance matrix to account for heteroscedasticity is the White Heteroscedasticity estimator, White (1980), which is also called heteroscedastic consistent-covariance matrix estimator (HCCME). It can be derived from equation 4.53 and equation 4.27:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \tag{4.94}$$

where $\hat{\mathbf{W}}$ is a diagonal matrix with squared OLS residuals on the diagonal (the original HC0 proposed by White):

$$\hat{\mathbf{W}} = \text{diag}(\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \ldots, \hat{\varepsilon}_T^2) \tag{4.95}$$

with $\hat{\varepsilon}_t = y_t - x_t^\top\hat{\boldsymbol{\beta}}$ being the OLS residual at time t.

Another technique is the Weighted Least Squares estimator that we have shown earlier, see equation 4.57, which will change the beta loading estimation as well.

In order to match the standard errors calculated by the OLS estimator with the errors of the GMM estimator in case of linear regression one option is to apply the White robust covariance estimator. However, for statistical inference, the Generalized Method of Moments and the Generalized Least Squares method which is used by NSUR regression, are

[16] We replace the averaged Jacobian by the summed one: $HX_{\text{sum}} = T \cdot HX_{\text{avg}}$ and apply $V = T\Lambda$.

both more robust than Ordinary Least Squares.

L GMM Equivalence of Raw and Demeaned Factor

In this Appendix, we show that the GMM raw and demeaned factor models are equivalent under the mapping: $\lambda_{\tilde{f}} = \lambda_f + \bar{f}$, $\lambda_0' = \lambda_0$, $\beta_i' = \beta_i$.

We use the following notation: $R_{i,t}$ are the (excess) returns for asset i at time t, f_t is the market factor with sample mean $\bar{f} \equiv T^{-1}\sum_{t=1}^{T} f_t$. We define the demeaned factor $\tilde{f}_t \equiv f_t - \bar{f}$ and consider the stacked linear system with common prices of risk λ_0, λ_f and loadings β_i.

For the raw factor system, the residuals are:

$$\varepsilon_{i,t}(\lambda_0, \lambda_f, \beta_i) = R_{i,t} - \lambda_0 - \lambda_f\beta_i - \beta_i f_t, \quad \forall i, t$$

and the GMM/SUR moments with instruments $Z_t = [1, f_t]^\top$ are:

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t} = 0, \qquad \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}\, f_t = 0, \quad \forall i$$

For the demeaned factor system, where $\tilde{f}_t = f_t - \bar{f}$, the residuals are:

$$\varepsilon_{i,t}'(\lambda_0', \lambda_{\tilde{f}}, \beta_i') = R_{i,t} - \lambda_0' - \lambda_{\tilde{f}}\beta_i' - \beta_i'\tilde{f}_t, \quad \forall i, t$$

and the GMM/SUR moments with instruments $Z_t' = [1, \tilde{f}_t]^\top$ are:

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}' = 0, \qquad \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}'\, \tilde{f}_t = 0, \quad \forall i$$

Using $f_t = \bar{f} + \tilde{f}_t$, $\lambda_{\tilde{f}} = \lambda_f + \bar{f}, \lambda_0' = \lambda_0,$ and $\beta_i' = \beta_i$:

$$\begin{aligned}\varepsilon_{i,t}' &= R_{i,t} - \lambda_0 - (\lambda_f + \bar{f})\beta_i - \beta_i(f_t - \bar{f}) \\ &= R_{i,t} - \lambda_0 - \lambda_f\beta_i - \beta_i f_t = \varepsilon_{i,t}, \quad \forall i, t\end{aligned}$$

Since the residuals coincide, the stacked sample moments and the GMM/SUR quadratic

objective are identical under the mapping. In particular:

$$\frac{1}{T}\sum_t \varepsilon'_{i,t} = \frac{1}{T}\sum_t \varepsilon_{i,t} = 0, \qquad \frac{1}{T}\sum_t \varepsilon'_{i,t}\,\tilde{f}_t = \frac{1}{T}\sum_t \varepsilon_{i,t}\,f_t - \bar{f}\frac{1}{T}\sum_t \varepsilon_{i,t} = 0, \quad \forall i$$

M Vech Operator

The *vech* operator is used to transform a symmetric matrix or a lower triangular matrix into a vector by stacking its elements on and below the main diagonal.

The definition applies for any symmetric matrix. We use $\mathbf{L} \in \mathbb{R}^{n\times n}$, a lower triangular matrix:

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix}.$$

The vech operator is defined as:

$$\text{vech}(\mathbf{L}) = \big(\,\ell_{11},\,\ell_{21},\,\ell_{22},\,\ell_{31},\,\ell_{32},\,\ell_{33}\,\big)^\top.$$

where the elements are listed row by row, taking only the entries on or below the diagonal, which is particularly useful when parameterizing covariance matrices via their Cholesky factor $\mathbf{L}$, because only the entries on and below the diagonal are free parameters.

N Derivation of the Elliptical Copula Log-Density

N.1 Gaussian Copula Log-Density

Under the Gaussian copula specification, we assume the marginals are parametric Gaussians:

$$F_{X_i}(x) = \Phi\Big(\frac{x}{\sigma_i}\Big), \quad \text{where} \quad \sigma_i = \sqrt{\Sigma_{ii}}, \quad \forall i$$

where Φ denotes the standard normal cumulative distribution function. For each residual time series i, we observe T values $\varepsilon_{i,t}$. We model these as draws from a marginal cumulative distribution function F_{X_i}. The probability integral transform is then applied to each observation:

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

We write $z_{i,t} \equiv q_{i,t}$ for the Gaussian case, where $z_{i,t}$ is the standardised residual quantile realisation obtained via the probability integral transform followed by the inverse standard

normal CDF:

$$z_{i,t} = \Phi^{-1}\big(u_{i,t}\big), \quad \forall i, t$$

and:

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big) = \Phi\Big(\frac{\varepsilon_{i,t}}{\sigma_i}\Big), \quad \forall i, t$$

So plugging in:

$$z_{i,t} = \Phi^{-1}\Big[\Phi\Big(\frac{\varepsilon_{i,t}}{\sigma_i}\Big)\Big] = \frac{\varepsilon_{i,t}}{\sigma_i}, \quad \forall i, t$$

where $\sigma_i = \sqrt{\Sigma_{ii}}$ denotes the marginal standard deviation of the residuals in series i. This transformation standardizes each residual to have a marginal standard normal distribution.

If $G = \Phi$ and $g = \phi$ are the standard normal Cumulative Distribution Function (CDF) and Probability Density Function (PDF), then the copula density in equation 3.5 becomes[17]:

$$c_{\boldsymbol{X}}(u_1, \ldots, u_N) = \frac{\phi_N\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N);\ 0, \mathbf{P}\big)}{\prod_{i=1}^{N} \phi\big(\Phi^{-1}(u_i)\big)}$$

We derive the full expression below:

$$\ln c_{\boldsymbol{X}}\left(u_{1,t}, \ldots, u_{N,t}\right) = -\frac{1}{2}\Big(\mathbf{z}_t^\top \left(\mathbf{P}^{-1} - \mathbf{I}\right) \mathbf{z}_t\Big) \ - \ \frac{1}{2} \ln\big|\mathbf{P}\big| \ + \ \text{constant}, \quad \forall t$$

where

$$c_{\boldsymbol{X}}\left(u_{1,t}, \ldots, u_{N,t}\right) = \frac{\phi_N\left(\mathbf{z}_t;\ 0, \mathbf{P}\right)}{\prod_{i=1}^{N} \phi\left(z_{i,t}\right)}, \quad \forall t$$

For the derivation, we start from the density of a multivariate standard Gaussian with correlation matrix $\mathbf{P}$:

$$\phi_N(\mathbf{z}_t; 0, \mathbf{P}) = \frac{1}{(2\pi)^{N/2} \left|\mathbf{P}\right|^{1/2}} \ \exp\Big(-\frac{1}{2} \mathbf{z}_t^\top \mathbf{P}^{-1} \mathbf{z}_t\Big), \quad \forall t$$

Taking logarithms, we obtain the log multivariate normal density:

$$\ln \phi_N(\mathbf{z}_t; 0, \mathbf{P}) = -\frac{N}{2} \ln(2\pi) \ - \ \frac{1}{2} \ln\big|\mathbf{P}\big| \ - \ \frac{1}{2} \mathbf{z}_t^\top \mathbf{P}^{-1} \mathbf{z}_t, \quad \forall t$$

[17] We recall the distinction in the notation below:

- the function definition $c_{\boldsymbol{X}}(u_1, \ldots, u_N)$ is a deterministic formula mapping $[0,1]^N$ to $\mathbb{R}_+$,
- the evaluation at random variables: $c_{\boldsymbol{X}}(U_1, \ldots, U_N)$ is itself a random variable,
- the evaluation at observed data: $c_{\boldsymbol{X}}(u_{1,t}, \ldots, u_{N,t})$ is a number.

For the denominator, we use the univariate standard normal density:

$$\phi(z_{i,t}) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2} z_{i,t}^2\Big), \quad \forall i, t$$

Taking logarithms:

$$\ln \phi(z_{i,t}) = -\frac{1}{2} \ln(2\pi) \; - \; \frac{1}{2} z_{i,t}^2, \quad \forall i, t$$

Summing over all components, we obtain the log of the product of marginals:

$$\sum_{i=1}^{N} \ln \phi(z_{i,t}) = -\frac{N}{2} \ln(2\pi) \; - \; \frac{1}{2} \sum_{i=1}^{N} z_{i,t}^2, \quad \forall t$$

Recall that, by Sklar's theorem 3.4, if the joint distribution is absolutely continuous with density $f_{\boldsymbol{X}}$, and the marginals have densities f_{X_i}, this implies that the joint density can be factorized as:

$$f_{\boldsymbol{X}}(x_1, \ldots, x_N) = c_{\boldsymbol{X}}\big(F_{X_1}(x_1), \ldots, F_{X_N}(x_N)\big) \prod_{i=1}^{N} f_{X_i}(x_i).$$

Therefore, the log copula density is given by subtracting the log of the product of the marginal densities from the log of the joint density:

$$\ln c_{\boldsymbol{X}} = \ln f_{\boldsymbol{X}} \; - \; \sum_{i=1}^{N} \ln f_{X_i}.$$

We obtain:

$$\ln c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) = \ln \phi_N(\mathbf{z}_t; 0, \mathbf{P}) - \sum_{i=1}^{N} \ln \phi(z_{i,t}), \quad \forall t$$

Expanding the expressions:

$$\ln c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) = -\frac{1}{2} \ln\big|\mathbf{P}\big| - \frac{1}{2} \mathbf{z}_t^\top \mathbf{P}^{-1} \mathbf{z}_t + \frac{1}{2} \sum_{i=1}^{N} z_{i,t}^2, \quad \forall t$$

We note that:

$$\sum_{i=1}^{N} z_{i,t}^2 = \mathbf{z}_t^\top \mathbf{I} \mathbf{z}_t, \quad \forall t$$

Therefore:

$$\ln c_{\boldsymbol{X}}\left(u_{1,t},\ldots,u_{N,t}\right) = -\frac{1}{2}\,\mathbf{z}_t^{\top}\left(\mathbf{P}^{-1}-\mathbf{I}\right)\mathbf{z}_t - \frac{1}{2}\,\ln\left|\mathbf{P}\right| + \text{constant}, \quad \forall t$$

where the quadratic form is defined as:

$$Q_t = \mathbf{z}_t^{\top}\left(\mathbf{P}^{-1}-\mathbf{I}\right)\mathbf{z}_t.$$

and the constant arises from the cancelled normalising terms.

N.2 Student-t Copula Log-Density

Under the Student-t copula specification, we assume the marginals are parametric univariate Student-t distributions with ν degrees of freedom and unit scale:

$$F_{X_i}(x) = t_\nu\big(x\big), \quad \forall i$$

where $t_\nu(\cdot)$ denotes the cumulative distribution function.

The probability integral transform is:

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big) = t_\nu\big(\varepsilon_{i,t}\big), \quad \forall i,t$$

Then, $z_{i,t}$ is the standardised residual quantile realization:

$$z_{i,t} = t_\nu^{-1}\big(u_{i,t}\big), \quad \forall i,t$$

Substituting the expression for $u_{i,t}$, we obtain:

$$z_{i,t} = t_\nu^{-1}\Big[t_\nu\big(\varepsilon_{i,t}\big)\Big] = \varepsilon_{i,t}, \quad \forall i,t$$

Under this parametric marginal assumption: each residual $\varepsilon_{i,t}$ already has a marginal Student-t distribution with ν degrees of freedom.

If $G = t_\nu$ and $g = t'_\nu$ denote the univariate Student-t cumulative distribution function and

probability density function, then the copula density in equation 3.5 becomes:

$$c_{\boldsymbol{X}}(u_1, \ldots, u_N) = \frac{t_N\Big(\, t_\nu^{-1}(u);\ \nu,\, \mathbf{P}\Big)}{\prod_{i=1}^{N} t'_\nu\Big(\, t_\nu^{-1}(u_i)\Big)},$$

where:

- $t_\nu^{-1}(\cdot)$ denotes the quantile function (inverse CDF) of the univariate Student-t distribution with ν degrees of freedom: $t_\nu^{-1}(u)$ is defined as the value x such that $\mathbb{P}(T \leq x) = u$.

- $t_N(\,\cdot\,;\, \nu, \mathbf{P})$ denotes the density of the N-dimensional Student-t distribution with correlation matrix $\mathbf{P}$, given by:

$$t_N\big(\mathbf{z}_t;\, \nu,\, \mathbf{P}\big) = \frac{\Gamma\big(\frac{\nu+N}{2}\big)}{\Gamma\big(\frac{\nu}{2}\big)\, \nu^{N/2}\, \pi^{N/2}\, \big|\mathbf{P}\big|^{1/2}} \Big(1 + \frac{1}{\nu}\, \mathbf{z}_t^\top \mathbf{P}^{-1}\, \mathbf{z}_t\Big)^{-\frac{\nu+N}{2}}, \quad \forall t$$

- $t'_\nu(\cdot)$ denotes the probability density function of the univariate Student-t distribution with ν degrees of freedom, given by:

$$t'_\nu\big(z_{i,t}\big) = \frac{\Gamma\big(\frac{\nu+1}{2}\big)}{\sqrt{\nu\pi}\, \Gamma\big(\frac{\nu}{2}\big)} \Big(1 + \frac{z_{i,t}^2}{\nu}\Big)^{-\frac{\nu+1}{2}}, \quad \forall i, t$$

The copula density can be written in expanded form as:

$$c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) = \frac{\dfrac{\Gamma\big(\frac{\nu+N}{2}\big)}{\Gamma\big(\frac{\nu}{2}\big)\, \nu^{N/2}\, \pi^{N/2}\, \big|\mathbf{P}\big|^{1/2}} \Big(1 + \frac{1}{\nu}\, \mathbf{z}_t^\top \mathbf{P}^{-1}\, \mathbf{z}_t\Big)^{-\frac{\nu+N}{2}}}{\prod_{i=1}^{N} t'_\nu\big(z_{i,t}\big)}, \quad \forall t$$

Taking the logarithms, we obtain:

$$\begin{aligned} \ln\, c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) = \ln\, \Gamma\Big(\frac{\nu+N}{2}\Big)\ -\ \ln\, \Gamma\Big(\frac{\nu}{2}\Big)\ -\ \frac{N}{2}\, \ln(\nu\pi)\ -\ \frac{1}{2}\, \ln\big|\mathbf{P}\big| \\ -\ \frac{\nu+N}{2}\, \ln\Big(1 + \frac{1}{\nu}\, \mathbf{z}_t^\top\, \mathbf{P}^{-1}\, \mathbf{z}_t\Big)\ -\ \sum_{i=1}^{N} \ln\, t'_\nu\big(z_{i,t}\big), \quad \forall t \end{aligned}$$

where the constants are not relevant for the optimisation and the quadratic form is defined

as:

$$Q_t = \mathbf{z}_t^\top \, \mathbf{P}^{-1} \, \mathbf{z}_t.$$

In our implementation, we replace the exact t-likelihood criterion with the Gaussian quadratic form, i.e. a Quasi Maximum Likelihood (QML) GLS approach, which is consistent for the mean parameters when paired with robust (sandwich) standard errors, where $\mathbf{e}_t^\top \boldsymbol{\Sigma}^{-1} \mathbf{e}_t \equiv Q_t$ for fixed $(\mathbf{P}, \nu, \boldsymbol{\Sigma})$.

N.3 Optimisation

The steps of our algorithm are:

1. Compute residuals:

$$\varepsilon_{i,t} = R_{i,t} - \lambda_0 - \lambda_f \, \beta_i - \beta_i \, \tilde{f}_t, \quad \forall i, t$$

2. Compute the parametric ranking (PIT) of the residuals via the univariate marginal CDF:

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

   where F_{X_i} denotes the CDF of the univariate marginal distribution.

3. Compute the quadratic form Q_t (see Gaussian N.1, and Student-t N.2)

4. Compute the marginal log densities of the residuals:

$$\sum_{i=1}^{N} \ln \, f_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall t$$

   where $f_{X_i}(\cdot)$ denotes the marginal density of the elliptical distribution, and

$$u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

5. Compute the log copula density:

$$\ln \, c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) = \ln \, f_N\big(\mathbf{z}_t \,\big|\, \nu, \, \mathbf{P}\big) \; - \; \sum_{i=1}^{N} \ln \, f_{Z_i}\big(z_{i,t}\big), \quad \text{for all } t$$

   where f_N denotes the multivariate elliptical density and f_{Z_i} the univariate elliptical marginal density.

6. Sum over all t to obtain the total log-likelihood:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \Big[\ln\, c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) + \sum_{i=1}^{N} \ln\, f_{X_i}\big(\varepsilon_{i,t}\big) \Big].$$

The parameter vector collects:

$$\boldsymbol{\theta} = \big(\, \lambda_0,\, \lambda_f,\, \beta_1, \ldots, \beta_N,\, \mathsf{vech}(\mathbf{L}),\, \nu \big),$$

where ν denotes the degrees of freedom or other tail parameter. Here, $\mathbf{L}$ is the lower triangular Cholesky factor of a covariance matrix $\mathbf{S}$, and the correlation matrix is defined as:

$$\mathbf{P} = \frac{\mathbf{S}}{\sqrt{\mathsf{diag}(\mathbf{S})\, \mathsf{diag}(\mathbf{S})^{\top}}}, \quad \text{with} \quad \mathbf{S} = \mathbf{L}\,\mathbf{L}^{\top}.$$

The vech operator (see Appendix M) is defined as:

$$\mathsf{vech}(\mathbf{L}) = \big(\, \ell_{11},\, \ell_{21},\, \ell_{22},\, \ell_{31},\, \ell_{32},\, \ell_{33},\, \ldots \big)^{\top}.$$

The parameter estimation proceeds by minimizing the negative log-likelihood with respect to $\boldsymbol{\theta}$. The optimisation is performed using BFGS or L-BFGS-B algorithms in Python (scipy.optimize minimize routine).

O Robust Inference for Gaussian Copula MLE

We define the parameter vector $\boldsymbol{\theta} = \big(\lambda_0,\, \lambda_f,\, \beta_1, \ldots, \beta_N,\, \mathsf{vech}(\mathbf{L})\big)$, where $\mathbf{L}$ is the lower-triangular Cholesky factor of the covariance matrix $\boldsymbol{\Sigma}$.

At each time t, the joint likelihood is expressed in terms of the copula density and marginal densities:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \Big[\ln\, c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) + \sum_{i=1}^{N} \ln\, f_{X_i}\big(\varepsilon_{i,t}\big) \Big].$$

where:

$$\varepsilon_{i,t} = R_{i,t} - \lambda_0 - \lambda_f\, \beta_i - \beta_i\, f_t, \quad u_{i,t} = F_{X_i}\big(\varepsilon_{i,t}\big), \quad \forall i, t$$

We define the score vector at time t as:

$$\mathbf{s}_t = \nabla_{\boldsymbol{\theta}} \Big[\ln\, c_{\boldsymbol{X}}\big(u_{1,t}, \ldots, u_{N,t}\big) + \sum_{i=1}^{N} \ln\, f_{X_i}\big(\varepsilon_{i,t}\big) \Big], \quad \forall t$$

Stacking all T contributions, the empirical outer product of scores is:

$$\widehat{\Omega}^{18} = \sum_{t=1}^{T} \mathbf{s}_t \, \mathbf{s}_t^\top.$$

and the Hessian $\widehat{\mathbf{H}}$ is the observed Hessian of the negative log-likelihood:

$$\widehat{\mathbf{H}} = -\nabla^2 \Big[\ln \mathcal{L}(\boldsymbol{\theta}) \Big].$$

Under correct specification of the copula and marginals, the covariance matrix of $\widehat{\boldsymbol{\theta}}$ is:

$$\widehat{\mathrm{Var}}_{\mathrm{MLE}}(\widehat{\boldsymbol{\theta}}) = \widehat{\mathbf{H}}^{-1}.$$

To obtain inference robust to possible model misspecification, we use the robust MLE estimator (*sandwich estimator*):

$$\widehat{\mathrm{Var}}_{\mathrm{robust}}(\widehat{\boldsymbol{\theta}}) = \widehat{\mathbf{H}}^{-1} \, \widehat{\Omega} \, \widehat{\mathbf{H}}^{-1}. \tag{96}$$

The robust standard errors, for $j = 1 \dots J$, are:

$$\widehat{\mathrm{SE}}(\widehat{\theta}_j) = \sqrt{\big[\widehat{\mathrm{Var}}_{\mathrm{robust}}(\widehat{\boldsymbol{\theta}})\big]_{jj}}.$$

The associated t-statistics are: $t_j = \frac{\widehat{\theta}_j}{\widehat{\mathrm{SE}}(\widehat{\theta}_j)}$. Assuming asymptotic normality, the two-sided p-values are: $p_j = 2\left(1 - \Phi\big(|t_j|\big)\right)$, where $J = N + N(N+1)/2 + 2$ is the total number of estimated parameters in $\hat{\boldsymbol{\theta}}$.

In our implementation, the log-likelihood is computed directly from the multivariate Gaussian distribution:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln \phi_N\big(\boldsymbol{\varepsilon}_t; \mathbf{0}, \boldsymbol{\Sigma}\big),$$

where $\phi_N(\cdot; \mathbf{0}, \boldsymbol{\Sigma})$ denotes the N-dimensional Gaussian density with mean zero and covariance matrix $\boldsymbol{\Sigma}$, and

$$\boldsymbol{\varepsilon}_t = \boldsymbol{R}_t - \lambda_0 \mathbf{1} - \lambda_f \boldsymbol{\beta} - \boldsymbol{\beta} f_t, \quad \forall t$$

This formulation implicitly assumes both the copula and the marginals are Gaussian. As

[18] This use of $\widehat{\boldsymbol{\Omega}}$ refers to the outer product of score vectors in the robust (sandwich) variance estimator. It is unrelated to the $\boldsymbol{\Omega}$ used earlier to denote the error covariance structure in the context of non-spherical disturbances.

a result, the full joint density coincides with the log-likelihood of a Gaussian copula with Gaussian marginals.

In the canonical copula decomposition, the log-likelihood is typically written as:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left[\ln c_{\boldsymbol{X}}(u_{1t}, \ldots, u_{Nt}) + \sum_{i=1}^{N} \ln f_{X_i}(\varepsilon_{i,t}) \right],$$

where $u_{i,t} = F_{X_i}(\varepsilon_{i,t})$ and $c_{\boldsymbol{X}}$ is the copula density.

When both the copula and marginals are Gaussian, this decomposition simplifies to the multivariate normal log-density.

We compute the finite-difference Hessian $\widehat{\mathbf{H}}$ from the Gaussian (or t) log-likelihood, and the outer product of per-time scores $\widehat{\boldsymbol{\Omega}}$ for robust inference. The robust (sandwich) covariance matrix is then given by equation 96.

References

Anatolyev, S. and Mikusheva, A. (2022). Factor models with many assets: Strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics*, 229(1): 103–126. (cited on pp. 7 and 52)

Apostol, T. M. (1969). *Calculus, Volume II: Geometry*. Wiley India Private Limited, 2nd edition. (cited on p. 33)

ARPM (2025). Copulas. https://www.arpm.co/lab/. Last accessed January 2025. (cited on pp. 75 and 76)

Asghari, M., Fathollahi-Fard, A. M., Mirzapour, A., and Dulebenets, M. A. (2022). Transformation and linearization techniques in optimization: A state-of-the-art survey. *Mathematics (Basel)*, 10(2): 283. (cited on p. 16)

Ballestra, L. V. and Tezza, C. (2025). A multi-factor model for improved commodity pricing: Calibration and an application to the oil market. *10.48550/arxiv.2501.15596*. (cited on p. 123)

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., and Kato, K. (2018). High-dimensional econometrics and regularized GMM. Technical report, Cornell University Library, arXiv.org. (cited on p. 128)

Black, F., Jensen, M. C., and Scholes, M. (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets*, pages 79–121. (cited on pp. 7 and 8)

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge. (cited on p. 17)

Brooks, R. and Iorio, A. D. (2009). Testing the integration of the US and Chinese stock markets in a fama-french framework. *Journal of Economic Integration*, 24: 435–454. (cited on pp. 31, 32, 33, 51, 52, 123, and 126)

Bruner, R. F., Li, W., Kritzman, M., Myrgren, S., and Page, S. (2008). Market integration in developed and emerging markets: Evidence from the CAPM. *Emerging markets review*, 9(2): 89–103. (cited on pp. 32, 33, and 85)

Bryzgalova, S. (2015). *Essays in empirical asset pricing*. PhD thesis. London School of Economics

and Political Science (University of London), http://etheses.lse.ac.uk/id/eprint/3204. (cited on pp. 86 and 125)

Cochrane, J. H. (2005). *Asset Pricing*. Princeton University Press, Princeton, NJ, 2nd edition. (cited on pp. 6, 7, 9, 10, 12, 29, 31, 33, 48, 68, 70, 75, 82, and 83)

Connor, G. (1984). A unified beta pricing theory. *Journal of Economic Theory*, 34(1): 13–31. (cited on p. 33)

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2): 255–289. (cited on p. 7)

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). *Correlation and Dependence in Risk Management: Properties and Pitfalls*, pages 176–223. Cambridge University Press, United Kingdom. (cited on p. 75)

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1): 3–56. (cited on pp. 7 and 12)

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1): 1–22. (cited on p. 36)

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of political economy*, 81(3): 607–636. (cited on pp. 6, 7, 9, 10, 29, 30, and 76)

Fernandez, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441): 359. (cited on p. 91)

French, K. (2025). Fama French Data. https://mba.tuck.dartmouth.edu/. Last accessed January 2025. (cited on pp. 36 and 88)

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premia in large cross-sectional equity data sets. *Econometrica*, 84(3): 985–1046. (cited on p. 7)

Gibbons, M. R. (1982). Multivariate tests of financial models: A new approach. *Journal of Financial Economics*, 10(1): 3–27. (cited on p. 12)

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5): 2223–2273. (cited on p. 8)

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. (cited on p. 36)

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4): 1029–1054. (cited on pp. 7, 30, and 76)

Hansen, L. P. and Richard, S. F. (1987). The role of conditioning information in deducing testable. *Econometrica*, 55(3): 587–613. (cited on p. 68)

Harvey, C. R. and Bekaert, G. (1994). Time-varying world market integration. Technical report, National Bureau of Economic Research. (cited on p. 85)

Harvey, C. R. and Liu, Y. (2021). Lucky factors. *Journal of Financial Economics*, 141(2): 413–435. ID: 271671. (cited on p. 63)

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55–67. (cited on p. 84)

IMF (2025). Commodity Prices. https://www.imf.org/en/Research/commodity-prices. Last accessed 2025. (cited on pp. 60 and 109)

Jagannathan, R. and Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51(1): 3–53. (cited on p. 6)

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London. (cited on p. 76)

Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25(5): 827–853. (cited on p. 75)

Jorion, P. and Schwartz, E. (1986). Integration vs. segmentation in the Canadian stock market. *The Journal of finance (New York)*, 41(3): 603–614. (cited on pp. 31, 32, 34, and 51)

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *The Journal of Economic Perspectives*, 15(4): 143–156. (cited on p. 93)

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408): 881–896. (cited on pp. 91 and 93)

Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2): 289–314. (cited on p. 7)

Litzenberger, R. H. and Ramaswamy, K. (1979). The effect of personal taxes and dividends on

capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7(2): 163–195. (cited on p. 7)

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I—convex underestimating problems. *Mathematical Programming*, 10(1): 147–175. (cited on p. 15)

Miffre, J. and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking and Finance*, 31(6): 1863–1886. ID: 271679. (cited on p. 130)

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International economic review (Philadelphia)*, 47(2): 527–556. (cited on p. 75)

Patton, A. J. (2013). *Copula Methods for Forecasting Multivariate Time Series*, volume 2, pages 899–960. Elsevier B.V, Amsterdam. (cited on p. 75)

Penco, V. and Lucas, C. (2024). Financial integration of the european, north america, asiatic and japanese stock markets from 2003 to present times. *Journal of Economic Integration*. (cited on pp. ii, 2, 5, 31, 37, 51, and 72)

Pericoli, M. and Taboga, M. (2012). Bond risk premia, macroeconomic fundamentals and the exchange rate. *International review of economics and finance*, 22(1): 42–65. (cited on p. 147)

Rayens, B. and Nelsen, R. B. (2000). An introduction to copulas. *Technometrics*, 42(3): 317. (cited on p. 76)

Richardson, M. and MacKinlay, A. (1991). Using generalized method of moments to test mean-variance efficiency. *Journal of Finance*, 46(2): 511–27. (cited on p. 52)

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3): 341–360. (cited on p. 12)

Sakkas, A. (2024). Risk premia in the commodity market. *Social Science Research Network, SSRN*. https://papers.ssrn.com/abstract=4796343. (cited on p. 70)

Sakkas, A. and Tessaromatis, N. (2020). Factor based commodity investing. *Journal of Banking and Finance*, 115: 105807. (cited on p. 130)

Sarisoy, C., de Goeij, P., and Werker, B. J. M. (2024). Linear factor models and the estimation of expected returns. *Finance and Economics Discussion Series*, I(2024-014). (cited on pp. 7 and 52)

SAS (2018). *SAS/ETS® 15.1 User's Guide: Model Procedure*. Cary, NC: SAS Institute Inc. (cited on p. 25)

Shanken, J. (1992). On the estimation of beta-pricing models. *The Review of financial studies*, 5(1): 1–33. (cited on pp. 7, 67, and 137)

Shanken, J. and Roll, R. (1985). Multivariate tests of the zero-beta CAPM/a note on the geometry of shanken's CSR t (superscript 2) test for mean/variance efficiency. *Journal of Financial Economics*, 14(3): 327. (cited on p. 75)

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3): 425–442. (cited on pp. 8 and 11)

Shi, Q. and Li, B. (2019). Evaluating alternative methods of asset pricing based on the overall magnitude of pricing errors. *Finance research letters*, 29: 125–128. (cited on p. 52)

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universit'e de Paris, 8*, pages 229–231. (cited on p. 77)

Solnik, B. H. (1974). The international pricing of risk : An empirical investigation of the world capital market structure. *Journal of Finance*, 29(2): 365–378. (cited on p. 32)

Stambaugh, R. F. (1982). On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics*, 10(3): 237–268. (cited on p. 12)

Stehle, R. (1977). An empirical test of the alternative hypotheses of national and international pricing of risky assets. *The Journal of finance (New York)*, 32(2): 493–502. (cited on pp. 32 and 33)

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5): 1055–1096. (cited on p. 7)

Stultz, R. M. (1984). Pricing capital assets in an international setting: An introduction. *Journal of International Business Studies*, 15(3): 55–73. (cited on p. 32)

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society.Series B, Statistical methodology*, 73(3): 273–282. (cited on p. 84)

Vasicek, O. A. (1973). A note on using cross-sectional information in bayesian estimation of security betas. *Journal of Finance*, 28(5): 1233–1239. (cited on p. 7)

Vielma, J. P. (2015). Mixed integer linear programming formulation techniques. *SIAM Review*, 57(1): 3–57. (cited on p. 18)

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4): 817–838. (cited on pp. 47 and 160)

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge. (cited on pp. 27 and 28)

Yahoo Finance (2025). Oil ETFs: Ex. USO. https://it.finance.yahoo.com/quote/USO/. Last accessed January 2025. (cited on pp. 60 and 109)

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and test of aggregation bias. *Journal of the American Statistical Association*, 57(298): 348–368. (cited on pp. 6, 30, and 52)

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society.Series B, Statistical methodology*, 67(2): 301–320. (cited on p. 84)