

# ASSET PRICING - EMPIRICAL APPLICATION 1

## FACTORIAL MODEL AND RISK PREMIUM DECOMPOSITION - APT

Luis Miguel Fonseca  
Stéphane Eloundou Mvondo  
Natalia Cárdenas Frías

November 29, 2023

### Introduction

Focusing on recent data from the French equity market, we want to better comprehend how the market prices systemic, non-diversifiable risk embedded in the risk premium of stocks, i.e. any expected compensation beyond the risk-free return. We base our analysis on a linear decomposition of said premium on different *factors* of risk in the spirit of the Arbitrage Pricing Theory (APT) pioneered by Ross [1976]. Unlike the CAPM model which considers a unique risk premium in the market, the Ross model gives a more detailed description of the pricing of aggregate risk by decomposing the contributions of different sources of risk. Here, a risky portfolio of  $j$  stocks<sup>1</sup> is compensated with  $k$  risk premia associated with the  $k$  *common factors* that the portfolio is exposed to.

## 1 Data

### 1.1 French stock market data

We decided to build a case study of the French market because it is a liquid and matured market, central in Europe. In the case of this analysis, we had trouble getting the data needed to perform it for other countries<sup>2</sup> and the fact that France has more publicly available data helped us choose it as our market of study.

We built a portfolio with 30 French stocks that we got from Yahoo Finance. For simplicity, the synthetic portfolio is composed of one stock of each company and its composition does not change during the period studied. Table 1 shows the companies that we used to create this portfolio, they are all publicly traded companies in France since the early 2000's in Euronext Paris. Importantly, we tried to have a certain diversity in the sectors

---

<sup>1</sup>Let  $j \in \{1, \dots, J\}$  with  $J$  sufficiently large so that all idiosyncratic risk can be fully diversified. We better explain the difference between idiosyncratic and aggregate in the context of the Ross model in Section 1.1

<sup>2</sup>Initially we thought about using data from the German market but we couldn't for instance find data for their inflation-linked bond yield that we use as an endogenous factor in Section 2.1.2

represented to be able to capture some diversification to risk even if the portfolio is too small and we do not reweight it. However, because we try to implement a version of the [Fama and French \[1992\]](#) factor analysis, we restrain ourselves from choosing financial companies as the authors do due to their high leverage. Other than these two conditions, the choice of the companies was mainly restricted to data availability on public 'long' series on firm-level data, notably on market capitalization and book-to-market ratio to be able to incorporate the [Fama and French \[1992\]](#) factors to our analysis.

Because we are interested in the underlying determinants of the risk premia, and due to data availability issues, we decided to have a broad analysis with monthly data for our selection of stocks for the period 2005-2022. Monthly data is for instance used by [Chen et al. \[1986\]](#). While this is not a very long period, it encompasses important moments in the financial markets in particular the Great Financial Crisis, the subsequent European Debt Crisis, and the Covid years. Also importantly, during most of this period (following the GFC), monetary policy fixed interest rates were extremely low driving down the return of sovereign debt for countries like France and Germany<sup>3</sup> that could have been assimilated to the risk-free rate. This means that for an investor to get any returns, it had to hold risky assets. Moreover, the period following the GFC and up to 2021 was also characterized by extremely low inflation in the Euro Zone. This is interesting because the APT usually incorporates inflation risk as market risk, yet inflation was nowhere to be found for more than a decade. Seeing how the market incorporated this monetary reality is by itself an intriguing question.

## 1.2 Data description and sources

The other series that we use are the following and its sources, how they are used in the context of the analysis is described in subsequent sections.

- As a proxy for the free rate of the market we consider two measures:
  - The yield of short-term OAT, i.e. French treasuries taken from [Banque de France's website](#). As for most developed, stable countries, short-term sovereign bonds are taken as the risk-free asset as Governments are supposed to be more solvent than other agents in the economy, after all, they decide their income and could seize resources via taxes to meet their obligations.
  - The spot yield curve spot rate, for 3-month maturity of all government bonds rated triple A in the Euro Area, retrieved from the [ECB webpage](#). On top of the fact that this is a measure for short-term sovereign bonds, we consider this to be a relevant proxy for the French market due to the strong integration within the European capital market. If an investor decides that the French market becomes risky, she can easily move her investments to another European capital market that looks safer.
- To get the market rate, we use the return of the main index of the country, the CAC40 also taken from Yahoo Finance as for the components of our synthetic portfolio. In hindsight, we are not sure of the pertinence of comparing our portfolio to this index. While the composition of our portfolio is not the

---

<sup>3</sup>They were negative for certain maturities in real terms for a part of the time frame analyzed, pretty much since the European Debt Crisis until the inflation surge after Covid.

Company Name	Ticker	Industry
Accor	AC.PA	Hospitality
Air Liquide	AI.PA	Industrial Gases
Air France-KLM	AF.PA	Airlines
Airbus	AIR.PA	Aerospace
Biomerieux	BIM.PA	Biotechnology
BIC	BB.PA	Consumer Goods
Bouygues	EN.PA	Construction
Capgemini	CAP.PA	Information Technology
Carrefour	CA.PA	Retail
Casino	CO.PA	Retail
Dassault Aviation	AM.PA	Aerospace
Danone	BN.PA	Food and Beverage
Hermes International	RMS.PA	Fashion and Luxury
JCDecaux	DEC.PA	Advertising
Kering	KER.PA	Fashion and Luxury
L'Oreal	OR.PA	Cosmetics
LVMH	MC.PA	Fashion and Luxury
Michelin	ML.PA	Automotive
Nexans	NEX.PA	Electrical Equipment
Orange	ORA.PA	Telecommunications
Renault	RNO.PA	Automotive
Saint-Gobain	SGO.PA	Manufacturing
Sanofi	SAN.PA	Pharmaceuticals
Sodexo	SW.PA	Food Services
TF1	TFL.PA	Broadcasting
Thales	HO.PA	Aerospace and Defense
TotalEnergies	TTE.PA	Energy
Ubisoft	UBI.PA	Video Games
Vinci	DG.PA	Construction
Vivendi	VIV.PA	Entertainment

Table 1: Synthetic portfolio: Companies, Tickers, and Industries

same as the CAC40<sup>4</sup>, due to the data availability issues we've been mentioning, we see that our choices are heavily biased towards 'big name' companies that are those belonging to the index.

- We got the series of the exchange rate between the Euro and the US dollar from Yahoo Finance. It is read as the amount of USD needed to get one euro.
- The GDP series is taken from the [ECB webpage](#). It is available at a quarterly frequency and is available at market prices.
- The harmonized headline inflation rate is taken from the [INSEE webpage](#).
- For the market inflation expectation in a 10-year horizon, we use the break-even inflation rate published by [Agence France Trésor](#) online. Sadly, data is only available from 2013.
- For the implementation of the two additional [Fama and French \[1992\]](#) factors, we took different routes
  - We found the estimation of the factors published by K. French in his [online Data Library](#) that are

<sup>4</sup>Not all the stocks we chose are necessarily part of the CAC40 at every period studied, and the CAC40 is a weighted index that evolves over time.

constantly updated. Their estimations start in the 1990s and are made for different markets using the comprehensive CRSP dataset that is not freely available. He has an estimation for the European market that we downloaded to use but it is not clear which stocks are used to replicate their portfolio.

- To try to build these estimations ourselves for our portfolio meaning that a minima we need data on the market capitalization of each company during the time frame studied and its *book*. This information is hardly available without having access to platforms like Bloomberg or CRSP. The best information that we could find comes from [this](#) website that publishes the market capitalization and the price-to-book (the inverse of the book-to-market ratio) annual series for several stocks. The data is however cannot be directly downloaded from the site so we scrap it to get the series (see Code Appendix B). Our biggest fear with this source is that it is not clear at all where the information comes from even if they mention several quality [data providers](#) as their partners. Since it is the only source that resembles what is needed for this part we used it but we are not confident about it.

## 2 Empirical strategy and implementation

We implement a minimal approach to [Ross \[1976\]](#), namely using fewer factors than in [Chen et al. \[1986\]](#). We include both exogenous and endogenous macroeconomic risk factors as well as an approximation to implement [Fama and French \[1993\]](#) three-factor model which used stock-specific data.

Let the return  $R_j$  of the  $j$ -th component of her portfolio can be described by the following expression  $\forall j \in \{1, \dots, N\}$ :

$$R_j = \mathbb{E}[R_j] + \underbrace{\sum_{k=1}^K \beta_{j,k} f_k}_{\text{Systemic risk}} + \overbrace{u_j}^{\text{Idiosyncratic risk}} \quad (1)$$

Where  $\mathbb{E}[R_j]$  is the expected return of asset  $j$  and  $R_j$  its return without dividend i.e. the first difference of the stock price. This is the so-called *factorial model* that includes two sources of risk:

- The investor faces centered idiosyncratic risks  $u_j$ ,  $\mathbb{E}[u_j] = 0$  that are assumed to be completely diversifiable with a portfolio "large enough" ( $N$  big) because they are independent of each other  $u_j \perp u_{j'} \forall j \neq j'$ .
- She also faces  $k$  different sources of aggregate risk, modeled by the linear combination of  $f_k$  centered *shocks* that influence all  $R_j$  with a sensitivity  $\beta_{j,k}$ . By definition, these risks cannot be diversified because they affect the returns of all asset and thus has to be compensated which is the focal point of our study.

In order to implement a regression analysis and estimation as the one that follows in this section, we shall also assume that the idiosyncratic risks are uncorrelated with aggregate risk  $\text{corr}(u_j, f_k) = 0, \forall j, k$ .

This section is organized as follows. Section 2.1 identifies the aggregate market risks  $(f_k)_k$  that we are going to consider as the ones priced by the market. Section 2.2 runs a first regression analysis to identify the sensitivity of each return to each factor of risk, i.e. to identify the  $(\beta_{j,k})_{j,k}$  from the factorial model (Eq. 1). It also implements an approximation to a parallel model that decomposes the return of stocks: the [Fama and French \[1992\]](#) that

also will lead us to estimate the sensitivity of the return of each return to a series of factors. Finally, section 2.3 uses the series of  $(\hat{\beta}_{j,k})_{j,k}$  estimated before to implement an estimation of the *multibeta relationship* which is the regression that will allow us to get how much the market is remunerating the exposition to a market factor risk.

## 2.1 Identify the risk factors

We decided to explore the role of the following sources of risk for a first model inspired by Ross [1976] and Chen et al. [1986].

- The activity risk, measured by changes in GDP
- Inflation risk, measured both by the HICP of France for the short term and by the market inflation expectation in a 10-y horizon.
- Devaluation risk measured by the exchange rate between the Euro and the US dollar.

We then consider the factors of a complementary factor model, Fama and French [1992] which are not directly linked to risks.

### 2.1.1 Exogeneous factors

These are risk factors that do not depend directly on the financial markets or more precisely that are not deduced from a linear combination of the returns of financial assets. In our case, it is mainly the activity risk, the short-term interest rate (measured by the HICP), and the devaluation risk. Importantly, the measure of these risks is not directly measured by changes or by the level of the underlying variables as markets are informational efficient [Fama, 1970], and they have already priced in all relevant information conveyed by prices. In particular, all *predictable* movements, say in inflation, have already been incorporated by the market. This means that the factors of risk are actually the *surprises* in the movements of these variables.

To extract this unpredictable part of these variables we need to use the residuals of some time series model of the (stationary) variables<sup>5</sup>, in our case ARIMA( $p, d, q$ ) models. We used the function *auto.arima* from the R *forecasts* package that set de parameters optimally by minimizing the BIC between different specifications. We ended up with:

- ARIMA(0,1,0) (i.e. a random walk<sup>6</sup>) for the series of log GDP.
- ARIMA(0,1,1) for the exchange rate series.
- ARIMA(2,1,2) for the HICP series.

We then stored the residuals of each one of these models and merge it in the data frame with the stock return. We shall note that by construction the factor (i.e. the residuals) are by construction centered around zero. They are actually the innovations of the initial series for the macroeconomic variables.

---

<sup>5</sup>Evidently, we need that the series used are stationary before conducting any TS analysis. We tested for stationarity of our series using the ADP test and then differentiate the series that were non-stationary (see Appendix A). We didn't apply the ADF test to the series in logs afterward as the R function *auto.arima* automatically differentiates the series  $d$  times until they are stationary.

<sup>6</sup>We do find this result 'weird' but didn't find any mistake in the code and it does square with the factors being centered.

### 2.1.2 Endogeneous factors

These factors are linear combinations of the returns of financial assets. We used the *Breakeven Inflation* at a 10-year horizon which is simply the yield difference between 10-y OATi 0.10% (*Obligations Assimilables du Trésor*) indexed by the HICP and non-indexed 10-y OAT. This gap is assimilated to how markets price the possibility of having future inflation<sup>7</sup>. Once again we need to extract the unforecastable movements in this indicator to have an adequate measure of risk. We follow the same methodology as with the exogenous factors and retrieve the residuals of an ARIMA(2,1,2) model.

### 2.1.3 French-Fama factors

Fama and French [1993] can also be seen as an extension of the CAPM model but their factors, while significant specially in the US market, are hardly interpretable as risk factors. The authors show that the variation of the returns of an asset can be explained not only by the exposure to market risk as in the CAPM represented by the difference of the market return and the risk-free rate  $[R_M - R_f]$ , but also by a size and value premium in the following model.

$$R_j = \alpha_j + R_f + \beta_{m,j}[R_M - R_f] + \beta_S SMB + \beta_V HML + \varepsilon_j \quad (2)$$

The size premium refers to the observation that stocks with small market capitalizations tend to outperform stocks with larger ones and it is captured by the factor SMB, *small minus big*. It is computed as the difference in average returns of the 30% stocks with the smallest market capitalization and the average returns of the 30% stocks associated with the firms with the largest market capitalization. The value premium refers to the outperformance of "value stocks" i.e. those that have high book-to-market (B/M) and it is represented by the difference in an average return of the 50% of stocks with the highest  $B/M$  ratio (value stocks) and the 50% with lowest  $B/M$  ratio (growth stocks). We use this methodology to create our own SMB and HML factors at a yearly frequency.

As we previously mentioned, K. French made public his estimation of the factors for the European model using a more complex approach with "6 value-weight portfolios formed on size and book-to-market". SMB (Small Minus Big) is the average return on the three small portfolios minus the average return on the three big portfolios, and HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios.

---

<sup>7</sup>Note that the Breakeven inflation is the inflation rate that would equalize the return of these two sovereign bonds (due to AOA).

## 2.2 Estimate the sensitivities of the factorial model

### 2.2.1 Ross model

Given that we assumed that the idiosyncratic risks  $u_j$  are uncorrelated with the aggregate factor risks  $f_k$  we can estimate the factorial model in Eq. 1 with linear regression models<sup>8</sup>. We will also pose that  $f_k \perp f_{k'} \forall k, k'$  in order to have a good identification of the parameters with linear regression estimators  $\forall k_0, \beta_{j,k_0} = \frac{cov(R_j, f_{k_0})}{Var(f_{k_0})} \equiv \hat{\beta}$ .

For this first estimation, we need to include a time dimension on the baseline factorial model  $\forall j, \forall t$ :

$$R_{j,t+1} = \mathbb{E}_t[R_{j,t+1}] + \sum_{k=1}^K \beta_{j,k} f_{k,t+1} + u_{j,t+1} \quad (3)$$

We will therefore estimate the following:

$$R_{j,t} \sim \alpha + \sum_{k=1}^K \beta_{j,k} \hat{f}_{k,t}$$

Note that for the macroeconomic risks  $\hat{f}_{k,t}$  denotes the estimated errors we found in the previous section for each series with standard TS analysis. To do so, we split the dataset by stock and performed linear regressions for each subgroup.

**Exogeneous factors only** We performed a first version of this estimation using only the exogenous factors which allows us to have an analysis of the 2005-2022 period. Results are shown in the following table and graph. The regression table is printed in the annex but it is really hard to read because of the minuscule font size, I didn't manage to split the  $\LaTeX$  table. It can however also be seen in html file in this link [HERE](#). (For all this section, regression tables are in the annex but I recommend you to check the links). We shall note that most coefficients are non-significative for every stock.

**Exogeneous and endogeneous factors** We add another specification to include the endogenous market expectation of long-term inflation. Sadly due to the short series on the Breakeven inflation, this analysis is conducted only on the period 2013-2022, still at a monthly frequency. The table can be seen [HERE](#).

**Exogeneous and French and Fama factors** The last specification includes the French and Fama factors given by K. French data library. As in the baseline, it refers to monthly data on the 2005-2022 period. This will be the specification we will move forward with because it allows us to integrate both Ross and French-Fama models in a longer period and we have more significant results. The table can be found [HERE](#). Note that in these regressions, the coefficient on inflation risk is rarely significant which is consistent with the period studied being a time of structurally low inflation but for the last two years: French companies were not exposed to rising price risk in general.

---

<sup>8</sup>The residuals will be uncorrelated with the explanatory variables

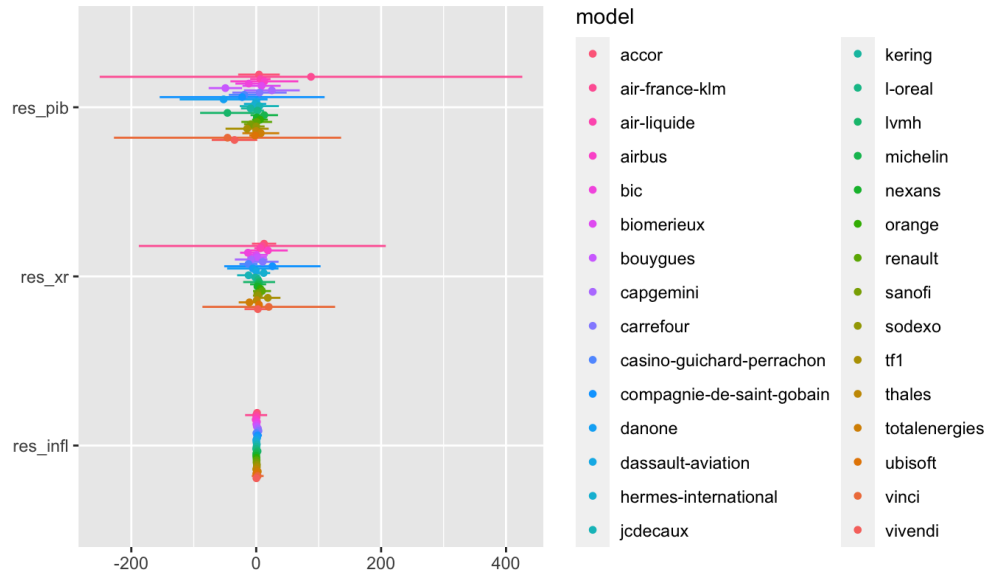


Figure 1: Betas: exogenous factors

Notes: Point estimate of the beta coefficients and its 95% confidence intervals. Note that confidence intervals close in the graph to zero might actually be significant due to the scale of the x-axis

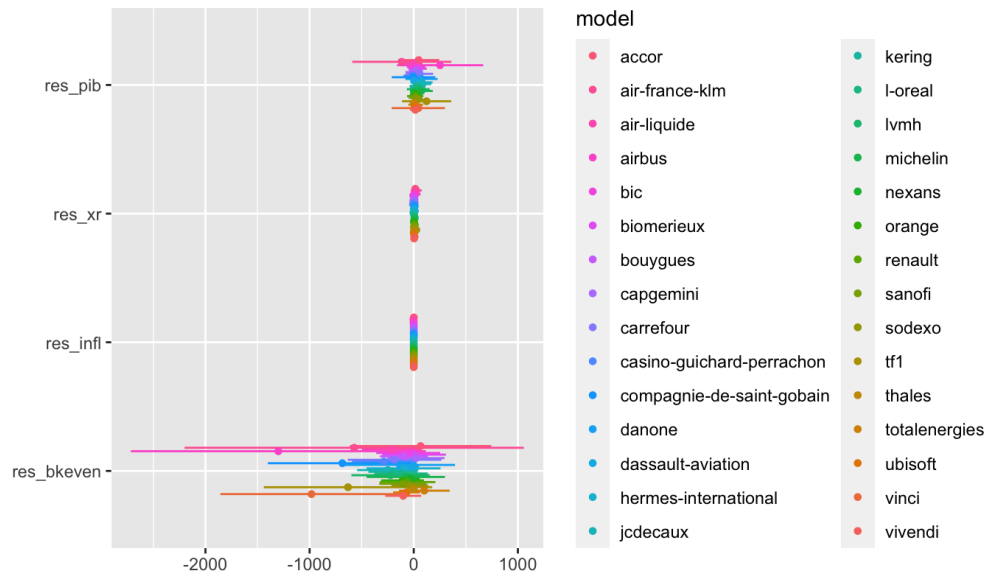


Figure 2: Betas: exogenous and endogenous factors

Notes: Point estimate of the beta coefficients and its 95% confidence intervals. Note that confidence intervals close in the graph to zero might actually be significant due to the scale of the x-axis

## 2.2.2 French and Fama model

We also wanted to try to estimate the three French and Fama with the data we collected. We used the approach mentioned in Section 2.1.3 and compared the mean return of value stocks and non-value stocks as well as of



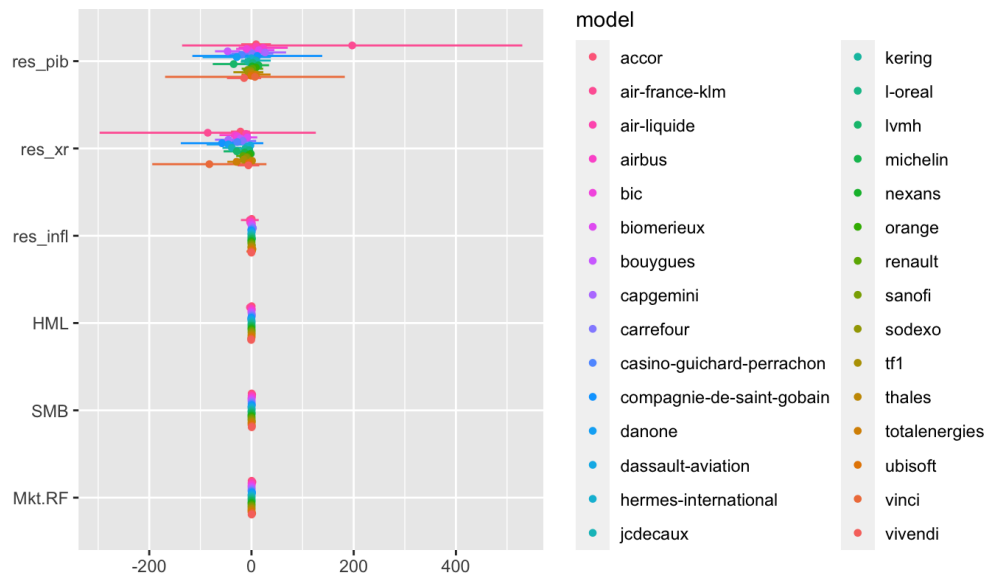


Figure 3: Betas: exogenous and French and Fama factors

*Notes:* Point estimate of the beta coefficients and its 95% confidence intervals. Note that confidence intervals close in the graph to zero might actually be significant due to the scale of the x-axis

growth vs non-growth stocks. Once again, due to data availability issues, we had to restrain our sample to study only the 2010-2022 period at a yearly frequency.

The results (see the regression table [HERE](#)) are not very significant but we think it is likely due to our data quality. We show the results for this but we will ignore this representation of the French and Fama model moving forward.

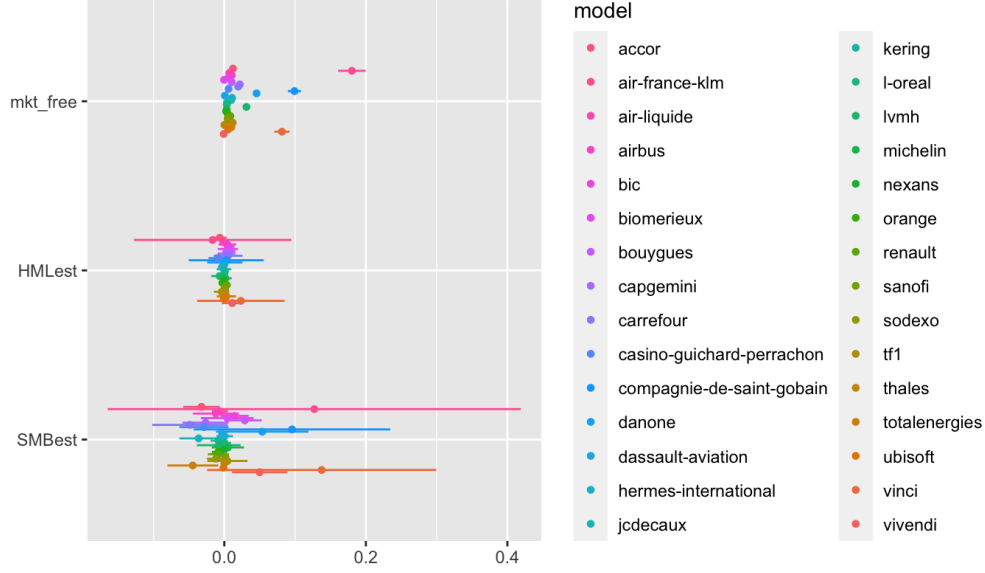


Figure 4: Betas: exogenous and French and Fama factors

Notes: Point estimate of the beta coefficients and its 95% confidence intervals. Note that confidence intervals close in the graph to zero might actually be significant due to the scale of the x-axis

## 2.3 Estimate the remuneration of risk from the multibeta relationship

Under AOA, assuming that the factorial model (Eq. 1) is an accurate depiction of how equity is price, implies that the expected returns are constrained by a multibeta relationship of the following form  $\exists \rho, \exists \lambda_1, \dots, \lambda_k$ :

$$\mathbb{E}[R_j] = \rho + \sum_{k=1}^K \lambda_k \beta_{j,k} \quad (4)$$

where each  $\lambda_k$  is the parameter representing the market price of risk (the risk premium) that the market re-tributes for being exposed to a given risk factor  $f_k$  with a sensitivity  $\beta_{j,k}$

To estimate the lambda coefficients we need to have a new dependent variable: the historical mean return of every stock in order to get a cross-sectional analysis. To do this last part of the analysis we store the sensitivities  $\hat{\beta}_{j,k}$  and the intercepts  $\hat{\rho}_j$  derived in the previous section and regress them over this historical mean  $\mathbb{E}[\hat{R}_j] - \hat{\rho} = \overline{R_{j,t}} - \hat{\rho}_j$ .

We decided to only keep the model with the exogenous factors and the French and Fama factors in monthly frequency and run the following cross-sectional regression:

$$\overline{R_{j,t}} - \hat{\rho}_j \sim \hat{\beta}_{pib,j} + \hat{\beta}_{xr,j} + \hat{\beta}_{infl,j} + \hat{\beta}_{HLM,j} + \hat{\beta}_{SMB,j} + \hat{\beta}_{Mkt.RF,j} \quad (5)$$

The coefficients on the  $(\hat{\beta}_k)_k$  will actually going to be the market payments to the investor for holding each risk in her portfolio: they are the risk-premia terms.

We first used a simple regression model to assess these coefficients. The following table summarizes it.

Table 2: Multibeta relationship with OLS

	<i>Dependent variable:</i>
	historical mean - intercept
beta_pib	0.001*** (0.000)
beta_xt	−0.001*** (0.000)
beta_infl	0.013*** (0.000)
beta_HML	−0.044*** (0.000)
beta_SMB	0.080*** (0.000)
beta_Mkt.RF	0.484*** (0.000)
Constant	−0.000 (0.000)
Observations	30
R <sup>2</sup>	1.000
Adjusted R <sup>2</sup>	1.000
Residual Std. Error	0.000 (df = 23)
F Statistic	14,020,019,119,421,070,797,960,277,655,552.000*** (df = 6; 23)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

We seem to have positive payments to bear output and inflation risk as well from the SMB factor. The contribution is negative for the devaluation risk and for the HML factor.

However, since the explanatory variables in this regression are themselves estimated from other models, a correction for potential errors needs to be done in particular to correct the standard errors.

The methodology recommended is to use a GLM model which will weight the variables by the inverse of the variance-covariance matrix of the residuals of the first estimation [[Balduzzi and Robotti, 2008](#)].

We tried to implement this in our code (see the end of Appendix C) but didn't manage to properly get the weights.

## Appendix A Additional tables and figures

### Macroeconomic factors

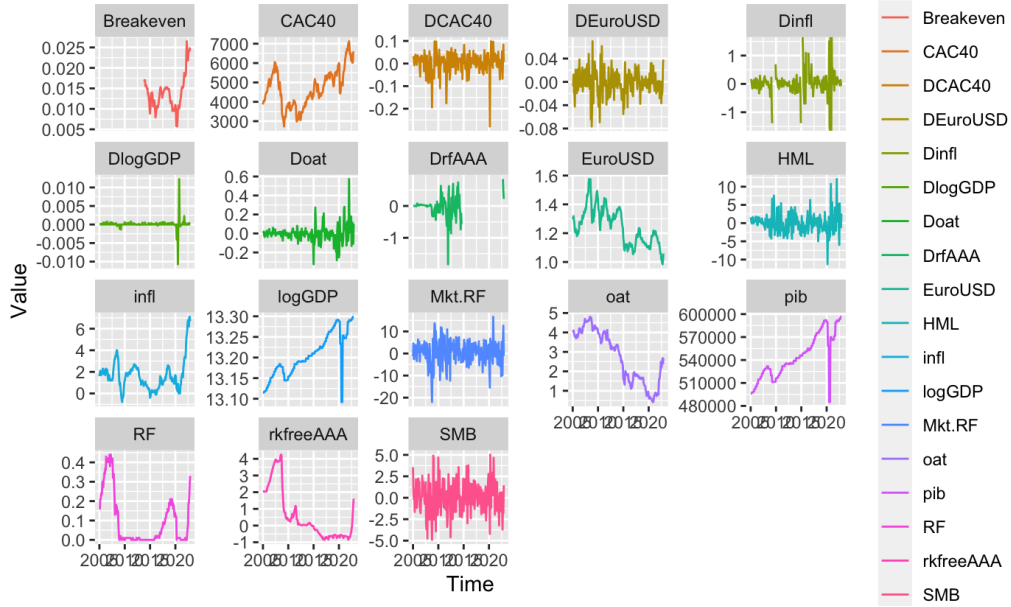


Figure 5: Factors: Series in level and in deltas

Table 3: Results ADF with trend and drift

	EuroUSD	CAC40	Inflation	PIB	RF AAA	OAT	1pct	5pct	10pct
tau3	-2.945	-2.028	-0.565	-4.716	-1.011	-0.720	-3.990	-3.430	-3.130
phi2	3.012	1.652	1.243	7.665	0.800	0.555	6.220	4.750	4.070
phi3	4.421	2.220	1.522	11.121	1.198	0.683	8.430	6.490	5.470

Table 4: Results ADF with drift

	EuroUSD	CAC40	Inflation	PIB	RF AAA	OAT	1pct	5pct	10pct
tau2	-1.841	-1.428	-0.444	-2.117	-1.545	-1.142	-3.460	-2.880	-2.570
phi1	1.791	1.277	0.438	2.589	1.195	0.803	6.520	4.630	3.810

Table 5: Results ADF with no trend nor drift

	EuroUSD	CAC40	Inflation	PIB	RF AAA	OAT	1pct	5pct	10pct
tau1	-0.618	0.414	0.353	0.721	-1.437	-0.995	-2.580	-1.950	-1.620

Figure 6: Betas: exogenous factors

13

Estimate the beta coefficients for each exogenous and endogenous factor

Dependent variable: Return																																						
	score	air-france-km	air-liquide	airbus	hic	biomeux	bourgais	capgemini	carrefour	casino	gildard	perrecheu	compagnie-de-saint-gobain	danone	dassault	aviation	hermes	international	jobscaux	kering	l-oreal	l'oreal	lvb	micbelin	exams	orange	renault	renault	sand	sodexo	ifl	thales	totalenergies	ubssoft	viesi	vivendi		
res_gnb	47.225 (99.831)	-114.811 (239.823)	-0.127 (42.066)	252.848 (208.506)	22.185 (45.090)	36.520 (45.088)	21.190 (34.038)	0.267 (40.371)	55.823 (66.098)	-9.259 (45.566)	-2.114 (105.316)	74.728 (77.968)	2.966 (8.901)	77.391 (52.266)	83.035* (43.311)	39.946 (37.779)	43.700 (55.550)	73.103 (55.415)	4.857 (13.209)	22.257 (30.177)	14.026 (33.470)	16.666 (25.936)	39.553 (19.057)	123.361 (119.057)	-0.367* (0.232)	19.218 (35.851)	8.349 (18.990)	43.899 (128.795)	15.103 (25.321)									
res_xr	14.768 (11.125)	23.779 (26.725)	10.730** (4.688)	20.465 (23.255)	-4.871 (5.025)	2.910 (5.024)	7.548** (3.793)	-4.210 (4.499)	9.481 (7.366)	-7.071 (5.022)	9.502 (11.736)	5.232 (8.688)	1.309 (0.992)	10.869* (5.824)	-3.688 (4.826)	4.296 (3.866)	10.222 (4.210)	4.810 (6.168)	2.326 (1.472)	4.667 (3.363)	2.765 (4.383)	3.279 (2.890)	24.667* (13.267)	0.992*** (0.026)	1.111 (3.995)	4.166* (2.116)	9.006 (14.352)	3.633 (2.822)										
res_inf	1.153 (1.330)	-2.827 (3.196)	-0.124 (0.561)	-2.077 (2.779)	-0.075 (0.601)	-0.113 (0.454)	0.613 (0.538)	-0.321 (0.881)	0.191 (0.881)	-0.700 (0.601)	0.860 (1.403)	0.420 (1.039)	-0.048 (0.119)	-0.883 (0.697)	-0.201 (0.577)	-0.013 (0.462)	-0.333 (0.738)	0.934 (0.176)	0.028 (0.482)	0.499 (0.254)	-0.242 (0.346)	0.362 (0.446)	-0.273 (1.287)	0.005 (0.003)	-0.135 (0.478)	-0.029 (0.253)	0.119 (1.716)	-0.194 (0.337)										
res_ikeven	65.720 (341.840)	-571.630 (821.196)	-213.246 (144.043)	-1.300.032* (713.961)	-53.868 (154.397)	2.813 (154.389)	-95.837 (116.552)	21.853 (138.238)	-183.955 (226.333)	-286.148* (154.313)	-685.514* (360.618)	-132.907 (268.975)	-9.489 (30.480)	-97.899 (178.968)	-249.767* (148.303)	-199.134* (118.783)	-122.313 (129.362)	-222.781 (189.751)	-76.577 (89.751)	-30.931 (103.330)	-107.728 (134.671)	-59.027 (88.808)	-103.925 (407.672)	-36.423 (67.794)	-630.963 (122.760)	-0.756 (65.027)	-469.126 (441.015)	-981.382** (86.705)	-101.222 (86.705)									
Constant	0.084 (0.357)	1.753** (0.859)	0.212 (0.151)	-0.061 (0.347)	0.188 (0.161)	0.007 (0.122)	0.287** (0.145)	0.252* (0.145)	0.452* (0.237)	0.399** (0.161)	0.726* (0.377)	0.524* (0.279)	-0.018 (0.032)	-0.006 (0.187)	0.384** (0.155)	0.148 (0.124)	-0.043 (0.135)	0.247 (0.198)	-0.015 (0.047)	0.096 (0.108)	0.219 (0.141)	0.067 (0.093)	0.044 (0.026)	0.0005 (0.001)	0.127 (0.128)	0.080 (0.068)	0.704 (0.461)	0.129 (0.091)										
Observations	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109	109		
R <sup>2</sup>	0.036	0.018	0.064	0.061	0.013	0.011	0.076	0.014	0.035	0.080	0.041	0.021	0.020	0.071	0.073	0.048	0.040	0.048	0.050	0.032	0.057	0.064	0.017	0.060	0.069	0.939	0.010	0.050	0.051	0.035	0.010	0.050	0.051	0.035	0.010	0.050	0.051	
Adjusted R <sup>2</sup>	-0.001	-0.020	0.028	0.025	-0.025	-0.027	0.041	-0.023	-0.002	0.044	0.004	-0.017	-0.017	0.035	0.037	0.011	0.003	0.012	0.013	-0.005	0.021	0.028	-0.021	0.024	0.033	0.936	-0.028	0.013	0.015	-0.002	-0.028	0.013	0.015	-0.002	-0.028	0.013	0.015	
Residual Std. Error (df = 104)	3.619	8.095	1.255	7.559	1.635	1.635	1.234	1.464	2.396	1.634	3.818	2.827	0.323	1.895	1.570	1.258	1.370	2.007	2.009	0.479	1.094	1.426	1.213	0.940	4.316	0.008	1.300	0.688	4.669	0.918	1.300	0.688	4.669	0.918	1.300	0.688	4.669	
F Statistic (df = 4; 104)	0.965	0.482	1.768	1.700	0.349	0.300	2.145*	0.381	0.934	2.257*	1.110	0.548	0.538	1.976	2.037*	1.309	1.076	1.320	1.358	0.872	1.569	1.705	0.448	1.652	1.932	398.596***	0.269	1.367	1.465	0.952	0.269	1.367	1.465	0.952	0.269	1.367	1.465	0.952

Figure 7: Betas: exogenous and endogenous factors

Estimate the beta coefficients for each exogenous and French and Famam factors

	Dependent variable:																															
	Return																Return															
	score	air-france-km	air-liquidide	airbus	bic	biométrieux	boyauges	cagellan	carrefour	casino	gaulichard	germarchon	compagnie-de-saint-gobain	danone	danone-acton	hermes-international	leclercq	loreal	lybny	michelein	recoan	orange	renault	santefi	sodexo	tf1	thales	totalenergies	ubsanft	vinci	vivendi	
res_pjb	9.008 (14.726)	197.078 (168.809)	8.992 (10.689)	22.914 (24.569)	-0.534 (10.381)	14.176 (15.545)	-46.679*** (12.585)	26.534 (20.882)	1.860 (20.808)	-19.907 (11.956)	-19.907 (11.956)	-19.907 (11.956)	11.391 (64.447)	-28.821 (33.899)	-0.202 (1.590)	1.835 (7.509)	8.791 (14.654)	-6.690 (7.258)	2.897 (4.551)	-34.850** (20.745)	13.992 (10.496)	8.509 (5.820)	1.774 (4.929)	-7.086 (5.889)	-1.164 (4.851)	-0.083 (0.031)	-0.040 (0.031)	8.949 (14.411)	0.406 (3.942)	6.800 (89.166)	-14.479 (16.957)	
res_xr	-21.529** (9.354)	-85.560 (107.234)	-10.601** (4.463)	-32.113** (15.480)	-27.600*** (6.594)	-8.154 (9.575)	-21.439*** (7.843)	-45.238*** (11.285)	-16.934 (13.218)	-27.779*** (7.595)	-27.779*** (7.595)	-27.779*** (7.595)	-57.576 (40.939)	-45.160** (21.541)	-2.190** (1.010)	-8.181* (4.770)	-38.402*** (9.309)	-12.775*** (4.611)	-4.357 (1.455)	-28.570*** (13.778)	-13.284** (6.668)	-1.407 (1.173)	-16.448*** (3.697)	-4.815*** (1.313)	-10.265*** (3.741)	-14.905 (9.454)	0.995*** (0.019)	-29.154*** (9.154)	-5.194*** (2.594)	-82.422 (56.642)	-0.003 (10.772)	
res_inf	0.344 (0.767)	-3.228 (8.795)	-0.002 (0.366)	-2.571** (1.270)	0.792 (0.541)	-0.103 (0.410)	0.307 (0.652)	1.812* (1.088)	2.793** (1.084)	-0.238 (0.623)	-0.238 (0.623)	-0.238 (0.623)	-0.242 (3.358)	0.533 (1.767)	0.062 (0.083)	-0.273 (0.391)	0.122 (0.763)	0.198 (0.378)	-0.276 (0.258)	1.041 (1.081)	-0.323 (0.547)	-0.042 (0.096)	0.077 (0.303)	0.177 (0.257)	0.754** (0.307)	-0.274 (0.774)	0.001 (0.002)	1.931** (0.751)	-0.247 (0.205)	-0.649 (4.646)	-0.109 (0.883)	
HML	0.331*** (0.093)	-2.902*** (1.066)	0.198*** (0.044)	0.524*** (0.154)	0.034 (0.066)	-0.033 (0.098)	0.171** (0.079)	0.464*** (0.132)	0.592*** (0.131)	0.268*** (0.076)	0.268*** (0.076)	0.268*** (0.076)	-0.189 (0.407)	-0.516** (0.214)	0.026*** (0.010)	0.181*** (0.047)	0.266*** (0.093)	0.033 (0.046)	0.088** (0.031)	0.123 (0.131)	0.158** (0.066)	0.014*** (0.012)	0.055 (0.037)	0.320*** (0.088)	0.197*** (0.031)	0.344*** (0.057)	0.002 (0.002)	0.245*** (0.091)	0.079*** (0.025)	-0.709 (0.583)	-0.573*** (0.107)	
SMB	0.362*** (0.127)	1.365 (1.459)	0.261*** (0.061)	0.842*** (0.211)	-0.075 (0.090)	0.262* (0.134)	0.070 (0.108)	0.292 (0.180)	0.359** (0.180)	0.215** (0.103)	0.215** (0.103)	0.215** (0.103)	0.705 (0.557)	0.002 (0.293)	-0.030** (0.014)	0.275*** (0.065)	0.331*** (0.127)	0.006 (0.063)	0.004 (0.043)	0.552*** (0.179)	-0.005 (0.091)	0.042*** (0.016)	0.075 (0.050)	0.145*** (0.043)	0.007 (0.051)	0.823*** (0.128)	0.004 (0.003)	0.272** (0.125)	0.096*** (0.034)	0.814 (0.771)	0.424*** (0.147)	
MLRF	0.296*** (0.049)	1.888*** (0.558)	0.134*** (0.023)	0.401*** (0.080)	0.172*** (0.034)	0.095* (0.051)	0.222*** (0.041)	0.300*** (0.069)	0.139** (0.069)	0.096*** (0.039)	0.096*** (0.039)	0.096*** (0.039)	1.022*** (0.213)	0.633*** (0.112)	0.016*** (0.005)	0.175*** (0.048)	0.216*** (0.048)	0.131*** (0.024)	0.066*** (0.016)	0.330*** (0.089)	0.159*** (0.035)	0.029*** (0.006)	0.118*** (0.016)	0.086*** (0.019)	0.101*** (0.049)	0.250*** (0.0848)	0.134*** (0.019)	0.087*** (0.013)	1.384*** (0.295)	0.236*** (0.056)		
Constant	-0.211 (0.228)	5.396*** (2.614)	-0.029 (0.109)	-0.376 (0.377)	0.165 (0.161)	0.005 (0.241)	0.446** (0.194)	0.196 (0.323)	0.476 (0.322)	0.137 (0.185)	0.137 (0.185)	0.137 (0.185)	2.448** (0.998)	1.046** (0.525)	0.001 (0.025)	-0.012 (0.116)	0.191 (0.227)	0.067 (0.112)	-0.064 (0.077)	0.468 (0.321)	-0.187 (0.163)	-0.069 (0.029)	0.249 (0.060)	0.021 (0.076)	0.131 (0.091)	0.116 (0.230)	-0.003 (0.005)	0.312 (0.223)	0.040 (0.061)	1.265 (1.381)	0.242 (0.263)	
Observations	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	215	
R <sup>2</sup>	0.286	0.075	0.322	0.259	0.166	0.038	0.241	0.213	0.190	0.146	0.146	0.146	0.118	0.148	0.131	0.357	0.196	0.162	0.137	0.193	0.156	0.208	0.226	0.305	0.290	0.315	0.329	0.939	0.158	0.296	0.108	0.209
Adjusted R <sup>2</sup>	0.266	0.048	0.303	0.238	0.142	0.010	0.220	0.191	0.166	0.122	0.122	0.122	0.092	0.123	0.106	0.338	0.172	0.138	0.112	0.170	0.132	0.185	0.204	0.284	0.270	0.295	0.310	0.938	0.133	0.276	0.082	0.186
Residual Std. Error (df = 208)	3.305	37.487	1.277	5.469	2.330	3.489	2.806	4.687	4.670	2.683	2.683	2.683	14.464	7.610	0.357	1.685	3.289	1.629	1.111	4.656	2.356	0.414	1.306	2.399	1.106	1.322	3.333	0.007	3.234	0.885	20.012	3.806
F Statistic (df = 6; 208)	13.904***	2.799**	16.492***	12.121***	6.910***	1.369	11.031***	9.401***	8.114***	5.947***	5.947***	5.947***	4.622**	6.000***	5.247***	19.232***	8.432***	6.719***	5.520***	8.282***	6.413***	9.117***	10.130***	15.182***	14.174***	15.919***	16.995***	536.989***	6.487***	14.573***	4.191***	9.165***

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Figure 8: Betas: exogenous and French and Famam factors

Estimate the beta coefficients for French and Fam factors																																	
Dependent variable:																																	
Return																																	
	accor	air	franc	km	air	liquide	air	bus	hic	biomoteurs	biogaz	capitaux	carrefour	cinéma	richard	perche	bus	compagne	de	sauir	gobain	diapase	danse	vision	hermes	international	le	jeune	l'escal	l'escal	l'escal	l'escal	l'escal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)
mL_fine	0.012*** (0.001)	0.186*** (0.010)	0.007*** (0.0004)	0.010*** (0.001)	0.006*** (0.001)	-0.0004 (0.001)	0.010*** (0.001)	0.022*** (0.001)	0.020*** (0.001)	0.006*** (0.001)	0.099*** (0.005)	0.046*** (0.002)	0.001*** (0.0001)	0.011*** (0.0004)	0.010*** (0.001)	0.004*** (0.0003)	0.003*** (0.001)	0.031*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.008*** (0.0005)	0.004*** (0.0002)	0.005*** (0.0004)	0.005*** (0.0004)	0.012*** (0.001)	0.005*** (0.0001)	0.010*** (0.001)	0.005*** (0.0001)	0.008*** (0.0001)	0.005*** (0.0001)	0.005*** (0.0001)	
HMLent	-0.006 (0.005)	-0.017 (0.057)	-0.001 (0.002)	0.004 (0.006)	0.007 (0.004)	0.005 (0.007)	0.007 (0.005)	0.003 (0.006)	0.005 (0.010)	-0.009 (0.007)	0.003 (0.027)	0.001 (0.013)	-0.001 (0.001)	-0.003 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	-0.005 (0.001)	
SNBest	-0.035*** (0.013)	0.127 (0.148)	-0.006 (0.006)	-0.012 (0.016)	0.014 (0.010)	0.004 (0.019)	0.029*** (0.012)	-0.027 (0.016)	-0.049* (0.027)	-0.029 (0.018)	0.096 (0.071)	0.054 (0.033)	-0.003* (0.002)	0.001 (0.006)	-0.006 (0.014)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	
Constant	0.133 (0.121)	1.124 (1.374)	0.008 (0.053)	-0.145 (0.152)	-0.011 (0.094)	0.010 (0.176)	-0.003 (0.113)	0.166 (0.152)	0.297 (0.251)	0.229 (0.163)	0.260 (0.654)	0.136 (0.307)	0.015 (0.016)	0.010 (0.055)	0.223* (0.128)	0.068 (0.053)	-0.056 (0.043)	0.235 (0.145)	-0.107 (0.107)	-0.010 (0.011)	0.044 (0.043)	0.116* (0.068)	-0.012 (0.025)	0.086 (0.059)	0.049 (0.134)	-0.001 (0.001)	0.327* (0.168)	0.020 (0.020)	-0.242 (0.263)	-0.122 (0.184)			
Observations	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390	
R <sup>2</sup>	0.349	0.469	0.467	0.192	0.156	0.002	0.307	0.513	0.241	0.072	0.540	0.530	0.123	0.673	0.238	0.208	0.248	0.703	0.048	0.741	0.386	0.438	0.446	0.267	0.289	0.005	0.165	0.786	0.368	0.018			
Adjusted R <sup>2</sup>	0.344	0.465	0.463	0.186	0.150	-0.006	0.302	0.509	0.235	0.065	0.536	0.526	0.116	0.670	0.232	0.202	0.242	0.701	0.041	0.739	0.381	0.433	0.443	0.262	0.283	-0.003	0.158	0.784	0.363	0.011			
Residual Std. Error (df = 386)	1.673	18.957	0.731	2.102	1.301	2.429	1.553	2.102	3.465	2.248	9.021	4.237	0.214	0.761	1.763	0.729	0.598	2.006	1.481	0.50	0.587	0.944	0.339	0.809	1.850	0.016	2.324	0.271	10.321	2.535			
F Statistic (df = 3; 386)	68.960***	113.475***	112.669***	30.571***	23.869***	0.212	57.024***	335.308***	40.808***	10.055***	150.916***	145.158***	18.029***	264.673***	40.173***	33.871***	42.392***	304.287***	6.483***	368.330***	80.744***	100.115***	46.930***	82.302***	0.669	25.354***	472.475***	4.965***	2.388*				
Note: *** p < 0.01, ** p < 0.05, * p < 0.1.																																	

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Figure 9: Betas: exogenous and French and Famamfactors



## Appendix B Code - Data gathering and cleaning in Python

```
1 #!/usr/bin/env python3
2 # -*- coding: utf-8 -*-
3 """
4 AP1 - Data gathering and Data Cleaning
5
6 Scrapping - Firm level data for French and Fama factors
7 Use Yahoo Finance API to get the financial data for all the stocks
8 Use Eurostat API to get the macro data
9 Two data sets (French-Fama factors and long term inflation expectation) are found online and
   have been downloaded in CSV file beforehang
10
11 Merge and clean the dataset
12
13 @author: nataliacardenasf
14 """
15
16 import pandas as pd
17 import numpy as np
18 import os
19
20 import requests
21 from bs4 import BeautifulSoup
22
23 #import pandas_datareader.data as web
24 import yfinance as yf
25 #from eurostatapiclient import EurostatAPIClient
26 import datetime
27
28
29 os.chdir('/Users/nataliacardenasf/Documents/GitHub/PROJECTS_AP_FE/AP 1')
30
31 company_names_lower = [
32     'air-liquide', 'airbus', 'bouygues', 'capgemini', 'carrefour', 'casino-guichard-perrachon',
33     , 'vivendi',
34     'kering', 'l-oreal', 'lvmh', 'michelin', 'orange', 'renault', 'sanofi', 'thales',
35     'totalenergies', 'vinci', 'compagnie-de-saint-gobain', 'ubisoft', 'tfl', 'danone',
36     'dassault-aviation', 'air-france-klm', 'accor', 'bic', 'hermes-international',
37     'jcdcaux', 'nexans', 'sodexo', 'biomerieux', "CAC40", "EuroUSD"]
38
39 tickers = [
40     'AI.PA', 'AIR.PA', 'EN.PA', 'CAP.PA', 'CA.PA', 'CO.PA', 'VIV.PA', 'KER.PA', 'OR.PA', 'MC.
41     PA',
42     'ML.PA', 'ORA.PA', 'RNO.PA', 'SAN.PA', 'HO.PA', 'TTE.PA', 'DG.PA', 'SGO.PA', 'UBI.PA', '
43     TFI.PA',
44     'BN.PA', 'AM.PA', 'AF.PA', 'AC.PA', 'BB.PA', 'RMS.PA', 'DEC.PA', 'NEX.PA', 'SW.PA', 'BIM.
45     PA', "^FCHI", 'EURUSD=X']
```

```

42
43
44
45 ### Scrap firm level data for French and Fama factors
46 # List of companies' URLs
47 mktcap_urls = ['https://companiesmarketcap.com/' + x + '/marketcap/' for x in company_names_lower
   [:-2]]
48 pricebook_urls = ['https://companiesmarketcap.com/' + x + '/pb-ratio/' for x in
    company_names_lower[:-2]]
49
50
51 # Scrapping functions
52 def scrape_market_cap(url, company_name):
53     response = requests.get(url)
54     if response.status_code == 200:
55         soup = BeautifulSoup(response.content, 'html.parser')
56         table_body = soup.find('table', class_='table').find('tbody')
57         if table_body:
58             data = []
59             rows = table_body.find_all('tr')
60             for row in rows:
61                 cols = row.find_all('td')
62                 if len(cols) >= 2:
63                     year = cols[0].text.strip()
64                     market_cap = cols[1].text.strip()
65                     #variation = cols[2].text.strip()
66                     data.append({'Year': year, 'MarketCap': market_cap, 'Company':
company_name})
67             return pd.DataFrame(data)
68     return None
69
70 def scrape_price_book(url, company_name):
71     response = requests.get(url)
72     if response.status_code == 200:
73         soup = BeautifulSoup(response.content, 'html.parser')
74         table_body = soup.find('table', class_='table').find('tbody')
75         if table_body:
76             data = []
77             rows = table_body.find_all('tr')
78             for row in rows:
79                 cols = row.find_all('td')
80                 if len(cols) >= 2:
81                     year = cols[0].text.strip()
82                     pricebook = cols[1].text.strip()
83                     #variation = cols[2].text.strip()
84                     data.append({'Year': year, 'PriceBook': pricebook, 'Company': company_name
    })
85     return pd.DataFrame(data)

```

```

86     return None
87
88
89 # Scraping market cap data
90 dfmktcap = pd.DataFrame()
91 for url, company in zip(mktcap_urls, company_names_lower[:-2]):
92     data = scrape_market_cap(url, company)
93     if data is not None:
94         dfmktcap = pd.concat([dfmktcap, data])
95
96 #Scap price book
97 dfpricebook = pd.DataFrame()
98 for url, company in zip(pricebook_urls, company_names_lower[:-2]):
99     data = scrape_price_book(url, company)
100    if data is not None:
101        dfpricebook = pd.concat([dfpricebook, data])
102
103 #indexes
104 dfmktcap['Year'] = pd.to_datetime(dfmktcap['Year'])
105 dfmktcap['Year'] = pd.DatetimeIndex(dfmktcap['Year']).year
106
107 dfpricebook['Year'] = pd.to_datetime(dfpricebook['Year'])
108 dfpricebook['Year'] = pd.DatetimeIndex(dfpricebook['Year']).year
109
110
111 ##Merge datasets
112 final_firm = dfmktcap.copy()
113 final_firm = final_firm.merge(dfpricebook, how='outer', on=['Year', 'Company'])
114
115
116 del dfmktcap, dfpricebook, mktcap_urls, pricebook_urls, url, data, company
117
118
119 missing = final_firm[final_firm.isna().any(axis=1)]
120 missing = missing.sort_values(by=['Year'])
121 missing = missing.reset_index()
122 #have both data points for all firms for 2010-2022
123 # in 09 only missing data is from BIC, Carrefour, Ubisolt, AirFrance
124
125 #remove 2023
126 final_firm = final_firm[final_firm.Year != 2023]
127
128 #Get book to market ratio = inverse of price-book ratio
129 final_firm['PriceBook'] = pd.to_numeric(final_firm['PriceBook'], errors='coerce')
130 final_firm['PriceBook'].replace('nan', np.nan, inplace=True)
131 final_firm['BookMarket'] = final_firm['PriceBook'].apply(lambda x: x ** -1 if not pd.isnull(x)
132                  ) else np.nan)

```

```

133 #Clean MarketCap
134 final_firm['MarketCap'] = (final_firm['MarketCap'].replace({'\$': '', ' B': ''}, regex=True).
    astype(float) * 1_000) # Clear the letters, convert to float and scale to millions
135
136 #Date format
137 final_firm['Year'] = pd.to_datetime(final_firm['Year'], format='%Y')
138
139 #===Get monthly data
140 monthly_data = pd.DataFrame()
141 # Repeat the yearly data for each month and each firm
142 for index, row in final_firm.iterrows():
143     firm_data = pd.DataFrame()
144     monthly_year = pd.date_range(start=row['Year'], periods=12, freq='MS')
145     firm_data['Date'] = monthly_year
146     firm_data['Company'] = row['Company']
147     firm_data['MarketCap'] = row['MarketCap']
148     firm_data['BookMarket'] = row['BookMarket']
149     firm_data['PriceBook'] = row['PriceBook']
150     monthly_data = pd.concat([monthly_data, firm_data])
151
152 del index, monthly_year, row, firm_data
153
154 firms_year = final_firm.copy()
155 firms_month = monthly_data.copy()
156
157 del final_firm, monthly_data, missing
158
159 #%%Get return data with Yahoo finance
160
161 start = datetime.datetime(2002, 1, 1)
162 end = datetime.datetime(2022, 12, 31)
163
164 #Get all data
165 data = yf.download(tickers, start=start,
166                    end=end)
167
168 #Focus on adjusted closed values only
169 adjclose=data['Adj Close']
170 adjclose = adjclose.set_axis(company_names_lower, axis=1)
171
172 #Use monthly data: mean of the months value
173 adjclose = adjclose.resample('1M').mean(numeric_only=True)
174 adjclose_y = adjclose.resample('1Y').mean(numeric_only=True)
175
176 ## extract CAC40 and exchange rate
177 cac_xrate_month = adjclose.loc[:, ["CAC40", "EuroUSD"]]
178 cac_xrate_year = adjclose_y.loc[:, ["CAC40", "EuroUSD"]]
179 adjclose = adjclose.drop(columns=["CAC40", "EuroUSD"])

```

```

180 adjclose_y = adjclose_y.drop(columns=["CAC40", "EuroUSD"])
181
182 #Reshape
183 prices_monthly = pd.melt(adjclose, value_vars=company_names_lower[0:-2], ignore_index=False)
184 prices_yearly = pd.melt(adjclose_y, value_vars=company_names_lower[0:-2], ignore_index=False)
185
186
187 del data, adjclose, adjclose_y
188
189 #I'm not getting right values for xrate when downloading in bulk
190 data = yf.download(['EURUSD=X', '^FCHI'], start=start, end=end)
191 adjclose=data['Adj Close']
192 adjclose = adjclose.set_axis(['EuroUSD', 'CAC40'], axis=1)
193 #adjclose = pd.DataFrame(adjclose, columns=['Date', 'EuroUSD'])
194
195 cac_xrate_month = adjclose.resample('1M').mean(numeric_only=True)
196 cac_xrate_year = adjclose.resample('1Y').mean(numeric_only=True)
197
198 del data, adjclose, start, end
199
200
201 #%% Endogeneous factor: long term inflation expectation from external file
202
203 #upload the Agence France Tresor data
204 pi_endo= pd.read_excel('2023_11_01_rend_tit_ref_oatei.xls', skiprows=[0,1,2,3,4], usecols
    =[0,3])
205 pi_endo.columns = ['Date', "Breakeven"]
206
207 pi_endo["Date"] = pd.to_datetime(pi_endo["Date"])
208 pi_endo= pi_endo.set_index(pi_endo["Date"])
209 pi_endo.drop(columns=['Date'])
210
211 #get monthly data
212 piendo_month = pi_endo.resample('1M').mean(numeric_only=True)
213
214 #get yearly data
215 piendo_year = pi_endo.resample('1Y').mean(numeric_only=True)
216
217 del pi_endo
218
219 #%% French and Fama - their data
220
221 df = pd.read_csv('Europe_3_Factors.csv', skiprows=[0,1,2])
222
223 #montly data, need to fix dates
224 frenchfama_month = df.iloc[:399,:]
225
226 frenchfama_month['Unnamed: 0'] = frenchfama_month['Unnamed: 0'].astype(str) # Convert to

```

```

    string for manipulation
227 frenchfama_month['Year'] = frenchfama_month['Unnamed: 0'].str[:4] # Extract year from the
    encoded date
228 frenchfama_month['Month'] = frenchfama_month['Unnamed: 0'].str[4:] # Extract month from the
    encoded date
229 frenchfama_month['Date'] = pd.to_datetime(dict(year=frenchfama_month['Year'], month=
    frenchfama_month['Month'], day=1))
230 frenchfama_month.drop(['Year', 'Month', 'Unnamed: 0'], axis=1, inplace=True)
231 frenchfama_month = frenchfama_month.set_index(frenchfama_month['Date'])
232 frenchfama_month = frenchfama_month.drop(columns=["Date"])
233 frenchfama_month = frenchfama_month.loc['2002-01-01':]
234 frenchfama_month = frenchfama_month.astype(float)
235
236
237 #yearly data
238 frenchfama_year = df.iloc[402:,:]
239 frenchfama_year["Unnamed: 0"] = pd.to_datetime(frenchfama_year['Unnamed: 0'])
240 frenchfama_year.rename(columns={"Unnamed: 0": 'Date'}, inplace=True)
241 frenchfama_year = frenchfama_year.set_index(frenchfama_year['Date'])
242 frenchfama_year = frenchfama_year.drop(columns=["Date"])
243 frenchfama_year = frenchfama_year.loc['2002-01-01':]
244 frenchfama_year = frenchfama_year.astype(float)
245
246 del df
247
248
249 ### Macro data
250 #APIs didn't work as planned
251
252 ## PIB Q
253 pib = pd.read_csv("ECB_PIB.csv")
254 pib.columns = ['Date', 'Q', 'pib']
255 pib['Date'] = pd.to_datetime(pib['Date'])
256 pib = pib.drop(columns=['Q'])
257 pib = pib.set_index(pib['Date'])
258 pib = pib.loc['2002-01-01':]
259 pib = pib.drop(columns=['Date'])
260
261 #monthly
262 pib_monthly = pib.resample('MS').ffill()
263 #yearly
264 pib_year = pib.resample('1Y').last()
265
266 del pib
267
268 ##risk free AAA
269 rkfreeAAA = pd.read_csv('ECB_yield.csv')
270 rkfreeAAA.columns=['Date', "time" , "rkfreeAAA"]

```

```

271 rkfreeAAA['Date'] = pd.to_datetime(rkfreeAAA['Date'])
272 rkfreeAAA = rkfreeAAA.set_index(rkfreeAAA['Date'])
273 rkfreeAAA = rkfreeAAA.drop(columns=['time', "Date"])
274
275 rkfreeAAA_monthly = rkfreeAAA.resample("1M").mean(numeric_only=True)
276 rkfreeAAA_year = rkfreeAAA.resample("1Y").mean(numeric_only=True)
277
278 del rkfreeAAA
279
280 ##HICP
281 pi_month = pd.read_csv("HICP.csv", sep=';', encoding = 'latin1', skiprows=[0,1,2,3], usecols
    = [0,1], header=None)
282 pi_month.columns = ['Date', "infl"]
283 pi_month["Date"] = pd.to_datetime(pi_month['Date'], format = "%Y-%m")
284 pi_month=pi_month.set_index(pi_month["Date"])
285 pi_month = pi_month.drop(columns=['Date'])
286
287 #yearly
288 pi_year = pi_month.resample("1Y").mean(numeric_only=True)
289
290
291 ##OAT
292 oat = pd.read_csv('OAT.csv', sep=';', skiprows=[0,1,2,3,4,5], usecols=[0,5], header=None)
293 oat.columns = ['Date', 'oat']
294 oat['Date'] = pd.to_datetime(oat["Date"])
295 oat = oat.set_index(oat["Date"])
296 oat = oat.drop(columns=["Date"])
297 oat = oat.loc["2002-01-01":]
298
299 oat['oat'] = oat['oat'].replace("-", np.nan)
300 oat['oat'] = oat['oat'].str.replace(',', '.').astype(float)
301
302 oat_month = oat.resample('M').mean(numeric_only=True)
303 oat_year = oat.resample('1Y').mean(numeric_only=True)
304
305 del oat
306
307
308 ### Merge the dataframes
309
310 #===MONTHLY
311 #macro stuff, only date index matter
312 monthly = pd.concat([cac_xrate_month, frenchfama_month, oat_month, pi_month, pib_monthly,
    piendo_month, rkfreeAAA_monthly], axis=1)
313 monthly= monthly.resample('M').last() #some tables encoded end of month, others on the 1st
314
315 monthly.to_csv('Monthly_series.csv')
316

```

```

317
318 #firm specific data
319 firms_month['Date'] = firms_month['Date'] + pd.offsets.MonthEnd(0) #all other df have eomonth
    date
320 firms_month = firms_month.set_index(['Date'])
321 firms_month = firms_month.set_index('Company', append=True)
322
323 prices_monthly.columns = ['Company', 'value']
324 prices_monthly.set_index('Company', append=True)
325
326 monthly_stock = pd.merge(prices_monthly.reset_index(), firms_month.reset_index(), on=["Date",
    "Company"], how='outer').set_index(["Date", "Company"])
327 monthly_stock.to_csv("Firm_monthly.csv")
328
329 #merge the two
330 merged_monthly = pd.merge(monthly, monthly_stock, left_index=True, right_index=True, how='
    right')
331
332 #Export df in csv
333 merged_monthly.to_csv("DATA_month.csv")
334
335
336
337 #=== YEARLY
338 yearly = pd.concat([cac_xrate_year, frenchfama_year, oat_year, pi_year, pib_year, piendo_year,
    rkfreeAAA_year], axis=1)
339 yearly = yearly.resample('Y').last()
340
341 yearly.to_csv('Yearly_series.csv')
342
343 # =====
344 # #firm specific
345 # firms_year.rename(columns={"Year":'Date'}, inplace=True)
346 # firms_year['Date'] = firms_year['Date'] + pd.offsets.MonthEnd(0)
347 # firms_year = firms_year.set_index(['Date'])
348 # firms_year = firms_year.set_index('Company', append=True)
349 #
350 # prices_yearly.columns = ['Company', 'value']
351 # prices_yearly = prices_yearly.set_index('Company', append=True)
352 #
353 # yearly_stock = pd.merge(prices_yearly.reset_index(), firms_year.reset_index(), on=["Date",
    "Company"], how='outer').set_index(["Date", "Company"])
354 # =====
355
356 #yearly_stock = monthly_stock.groupby('Company').resample('Y').mean()
357
358 monthly_data_reset = monthly_stock.reset_index()
359 yearly_stock = monthly_data_reset.groupby('Company').resample('Y', on='Date').mean()

```



```
360 #yearly_stock = yearly_stock.set_index(["Date", "Company"])
361
362 yearly_stock.to_csv("Firm_yearly.csv")
363
364
365 #merge the two
366 merged_year = pd.merge(yearly, yearly_stock, left_index=True, right_index=True, how='right')
367
368 merged_year.to_csv("DATA_yearly.csv")
```

## Appendix C Code - Analysis in R

```
1 #%% AP 1
2 #%% ncardenasfrias
3
4 pacman::p_load(data.table, urca, dotwisker, tidyverse, ggplots, xts, stargazer, forecast, plm,
5                 ggplot2, tidyr)
6 library(dplyr)
7 setwd('/Users/nataliacardenasf/Documents/GitHub/PROJECTS_AP_FE/AP 1')
8
9 df <- fread("DATA_month.csv")
10 as.data.table(df)
11 #####
12 ## Identify the risk factors
13 #####
14
15 #Need to remove the predicable part to the endogeneous and exogeneous series
16
17 #upload the monthly data with the yearly factorsand transform into a list of TS
18 data = read.csv('Monthly_series.csv')
19 data$Date <- as.Date(data$Date)
20
21 filtered_data <- data %>% #Filter rows between 2005 and 2022
22   filter(Date >= as.Date("2005-01-01") & Date <= as.Date("2022-12-31"))
23 time_series_cols <- filtered_data %>%
24   select(-Date)
25 time_series_list <- lapply(time_series_cols, function(col) {
26   ts_values <- ts(col, start = c(year(min(filtered_data$Date)), month(min(filtered_data$Date))
27     ), frequency = 12)
28   return(ts_values)
29 })
30 names(time_series_list) <- names(time_series_cols)
31
32 #plot series
33 time_series_df <- as.data.frame(time_series_list)
34 time_series_df$Date <- time(time_series_list[[1]])
35 time_series_long <- pivot_longer(time_series_df, cols = -Date, names_to = "Series", values_to
36   = "Value")
37 ggplot(time_series_long, aes(x = Date, y = Value, color = Series)) +
38   geom_line() +
39   facet_wrap(~ Series, scales = "free_y") +
40   labs(x = "Time", y = "Value")
41
42 ##check stationarity
43
44 lv = filtered_data[,c("EuroUSD", "CAC40", "infl", "pib", "Breakeven", "rkfreeAAA", 'oat')]
```

```

44
45 lv.adf.ln.trend = list(
46   XR = ur.df(lv$EuroUSD, type='trend', selectlags = c('BIC')),
47   cac = ur.df(lv$CAC40, type='trend', selectlags = c('BIC')),
48   infl = ur.df(lv$infl, type='trend', selectlags = c('BIC')),
49   pib = ur.df(lv$pib, type='trend', selectlags = c('BIC')),
50   #breakeven = ur.df(lv$Breakeven_no_na, type='trend', selectlags = c('BIC')),
51   rfAAA = ur.df(lv$rkfreeAAA, type='trend', selectlags = c('BIC')),
52   oat = ur.df(lv$oat, type='trend', selectlags = c('BIC'))
53 )
54
55 summary(lv.adf.ln.drift$XR)
56 print("levelVariable with drift and trend")
57 test = cbind(t(lv.adf.ln.trend$XR@teststat), t(lv.adf.ln.trend$cac@teststat),
58             t(lv.adf.ln.trend$infl@teststat), t(lv.adf.ln.trend$pib@teststat),
59             t(lv.adf.ln.trend$rfAAA@teststat), t(lv.adf.ln.trend$oat@teststat),
60             lv.adf.ln.trend$XR@cval)
61 #stargazer(test, out = 'Tables/trend_macro.tex')
62
63 lv.adf.ln.drift = list(
64   XR = ur.df(lv$EuroUSD, type='drift', selectlags = c('BIC')),
65   cac = ur.df(lv$CAC40, type='drift', selectlags = c('BIC')),
66   infl = ur.df(lv$infl, type='drift', selectlags = c('BIC')),
67   pib = ur.df(lv$pib, type='drift', selectlags = c('BIC')),
68   #breakeven = ur.df(lv$Breakeven_no_na, type='drift', selectlags = c('BIC')),
69   rfAAA = ur.df(lv$rkfreeAAA, type='drift', selectlags = c('BIC')),
70   oat = ur.df(lv$oat, type='drift', selectlags = c('BIC'))
71 )
72 print("levelVariable with drift ")
73 stat_macro_drift = cbind(t(lv.adf.ln.drift$XR@teststat), t(lv.adf.ln.drift$cac@teststat),
74                          t(lv.adf.ln.drift$infl@teststat), t(lv.adf.ln.drift$pib@teststat),
75                          t(lv.adf.ln.drift$rfAAA@teststat), t(lv.adf.ln.drift$oat@teststat),
76                          lv.adf.ln.drift$XR@cval)
77 #stargazer(stat_macro_drift, out='Tables/drift_macro.tex')
78
79 lv.adf.ln.none = list(
80   XR = ur.df(lv$EuroUSD, type='none', selectlags = c('BIC')),
81   cac = ur.df(lv$CAC40, type='none', selectlags = c('BIC')),
82   infl = ur.df(lv$infl, type='none', selectlags = c('BIC')),
83   pib = ur.df(lv$pib, type='none', selectlags = c('BIC')),
84   #breakeven = ur.df(lv$Breakeven_no_na, type='none', selectlags = c('BIC')),
85   rfAAA = ur.df(lv$rkfreeAAA, type='none', selectlags = c('BIC')),
86   oat = ur.df(lv$oat, type='none', selectlags = c('BIC'))
87 )
88 print("levelVariable with none ")
89 stat_macro_none = cbind(t(lv.adf.ln.none$XR@teststat), t(lv.adf.ln.none$cac@teststat),
90                        t(lv.adf.ln.none$infl@teststat), t(lv.adf.ln.none$pib@teststat),
91                        t(lv.adf.ln.none$rfAAA@teststat), t(lv.adf.ln.none$oat@teststat),

```

```

92         lv.adf.ln.none$XR@cval)
93 #stargazer(stat_macro_none, out='Tables/drift_none.tex')
94
95
96
97
98 #differentiate
99 filtered_data$logGDP = log(filtered_data$pib)
100
101 filtered_data$logGDP <- c(NA, diff(filtered_data$logGDP))
102 filtered_data$DCAC40 <- c(NA, diff(filtered_data$CAC40))
103 filtered_data$DEuroUSD <- c(NA, diff(filtered_data$EuroUSD))
104 filtered_data$Dinfl <- c(NA, diff(filtered_data$infl))
105 filtered_data$Doat <- c(NA, diff(filtered_data$oat))
106 filtered_data$DrfAAA <- c(NA, diff(filtered_data$rkfreeAAA))
107
108 time_series_cols <- filtered_data %>%
109   select(-Date)
110 time_series_list <- lapply(time_series_cols, function(col) {
111   ts_values <- ts(col, start = c(year(min(filtered_data$Date)), month(min(filtered_data$Date))
112     ), frequency = 12)
113   return(ts_values)
114 })
115 names(time_series_list) <- names(time_series_cols)
116
117 #plot series
118 time_series_df <- as.data.frame(time_series_list)
119 time_series_df$Date <- time(time_series_list[[1]])
120 time_series_long <- pivot_longer(time_series_df, cols = -Date, names_to = "Series", values_to
  = "Value")
121 ggplot(time_series_long, aes(x = Date, y = Value, color = Series)) +
122   geom_line() +
123   facet_wrap(~ Series, scales = "free_y") +
124   labs(x = "Time", y = "Value")
125
126 # Auto arima differentiates as much as needed to get stationary variables
127
128 logGDP_arima = auto.arima(filtered_data$logGDP) #ARIMA 0,1,0
129 logGDP_arima
130 EuroUSD_arima = auto.arima(filtered_data$EuroUSD) #ARIMA 0,1,1
131 EuroUSD_arima
132 infl_arima = auto.arima(filtered_data$infl) #ARIMA 1,1,1
133 infl_arima
134 breakeven_arima = auto.arima(filtered_data$Breakeven) #ARIMA 2,1,2
135 breakeven_arima
136
137 timestamps <- filtered_data$Date

```

```

138 res_pib = resid(logGDP_arima)
139 res_xr = resid(EuroUSD_arima)
140 res_infl = resid(infl_arima)
141 res_bkeven = resid(breakeven_arima)
142 min_length <- min(length(res_pib), length(res_xr), length(res_infl))
143 # Create a dataframe with aligned timestamps and residuals padded with NA for breakeven
144 residuals_df <- data.frame(
145   Date = timestamps[1:min_length],
146   res_pib = c(res_pib[1:min_length], rep(NA, times = max(0, length(timestamps) - min_length)))
147   ,
148   res_xr = c(res_xr[1:min_length], rep(NA, times = max(0, length(timestamps) - min_length))),
149   res_infl = c(res_infl[1:min_length], rep(NA, times = max(0, length(timestamps) - min_length)
150   )),
151   res_bkeven = c(res_bkeven[1:min_length], rep(NA, times = max(0, length(timestamps) - min_
152   length)))
153 )
154 #####
155 ## Estimate the beta coefficients
156 #####
157
158 data_firm = read.csv('Firm_monthly.csv')
159 data_firm$Date <- as.Date(data_firm$Date)
160
161 filtered_data_firm <- data_firm %>% #Filter rows between 2005 and 2022
162   filter(Date >= as.Date("2005-01-01") & Date <= as.Date("2022-12-31"))
163 time_series_cols_firm <- filtered_data_firm %>%
164   select(-Date)
165 time_series_list_firm <- lapply(time_series_cols_firm, function(col) {
166   ts_values_firm <- ts(col, start = c(year(min(filtered_data_firm$Date)), month(min(filtered_
167   data_firm$Date))), frequency = 12)
168   return(ts_values_firm)
169 })
170 names(time_series_list_firm) <- names(time_series_cols_firm)
171
172 ## to get return, differenciate value of the stock (within each company)
173 filtered_data_firm <- filtered_data_firm[order(filtered_data_firm$Company, filtered_data_firm$
174   Date), ]
175 filtered_data_firm$Return <- with(filtered_data_firm, ave(value, Company, FUN = function(x) c(
176   NA, diff(x))))
177
178 finalmonthly =merge(filtered_data_firm, filtered_data, by = "Date")
179 finalmonthly =merge(finalmonthly, residuals_df, by = "Date")
180
181 #split df by company to get a regression per stock

```

```

180 dta_bystock = split(finalmonthly, finalmonthly$Company)
181
182
183 #Exogeneous factors only
184 fit_lm <- function(data) {
185   lm(Return ~ res_pib + res_xr + res_infl, data = data)
186 }
187 regs_beta <- lapply(dta_bystock, function(subset) fit_lm(subset))
188
189 summary(regs_beta)
190 stargazer(regs_beta,
191           title = "Estimate the beta coefficients for each exogeneous factor",
192           column.labels = names(regs_beta),
193           out="Tables/betas_exo.tex")
194 stargazer(regs_beta,
195           title = "Estimate the beta coefficients for each exogeneous factor",
196           column.labels = names(regs_beta),
197           type = "html", out = "betas_exo.html")
198
199 library(dotwhisker)
200
201 dwplot(regs_beta)
202
203 #Include endofactor -> start in 2013
204 fit_lm_endo <- function(data) {
205   lm(Return ~ res_pib + res_xr + res_infl+ res_bkeven, data = data)
206 }
207 regs_beta_endo <- lapply(dta_bystock, function(subset) fit_lm_endo(subset))
208 stargazer(regs_beta_endo,
209           title = "Estimate the beta coefficients for exogeneous and endogeneous factors
210           (2013-2022)",
211           column.labels = names(regs_beta),
212           out="Tables/betas_exo_endo.tex")
213 stargazer(regs_beta_endo,
214           title = "Estimate the beta coefficients for each exogeneous and endogeneous factor "
215           ,
216           column.labels = names(regs_beta),
217           type = "html", out = "betas_exo_endo.html")
218 dwplot(regs_beta_endo)
219
220 #Include FF factors
221 fit_lm_exoff <- function(data) {
222   lm(Return ~ res_pib + res_xr + res_infl+ HML+SMB+Mkt.RF, data = data)
223 }
224 regs_beta_exoff <- lapply(dta_bystock, function(subset) fit_lm_exoff(subset))
225 summary(regs_beta_exoff)

```

```

226 stargazer(regs_beta_exoff,
227           title = "Estimate the beta coefficients for each exogeneous factor and French and
                Fama factors",
228           column.labels = names(regs_beta),
229           out="Tables/betas_exo_ff.tex")
230 stargazer(regs_beta_exoff,
231           title = "Estimate the beta coefficients for each exogeneous and French and Fama
                factor",
232           column.labels = names(regs_beta),
233           type = "html", out = "betas_exo_ff.html")
234
235 dwplot(regs_beta_exoff)
236
237 ##### FF model with annual data and our estimation for this sample
238 Firm_yearly <- read_csv("Firm_yearly.csv")
239 Firm_yearly = merge(Firm_yearly,
240                     Firm_yearly%>%
241                       filter(!is.na(MarketCap))%>%
242                       group_by(Date)%>%
243                       dplyr::summarise(MedianMCt=median(MarketCap)),
244                     by="Date")
245 DiffMC = merge(
246   Firm_yearly%>%
247     filter(!is.na(MarketCap) & MarketCap>MedianMCt & !is.na(value))%>%
248     group_by(Date)%>%
249     dplyr::summarise(AvgTop = mean(value)),
250   Firm_yearly%>%
251     filter(!is.na(MarketCap) & MarketCap<MedianMCt & !is.na(value))%>%
252     group_by(Date)%>%
253     dplyr::summarise(AvgBot = mean(value)),
254   by = "Date")
255 DiffMC$SMBest = DiffMC$AvgTop-DiffMC$AvgBot
256
257 Firm_yearly = merge(Firm_yearly,
258                     Firm_yearly%>%
259                       filter(!is.na(BookMarket))%>%
260                       group_by(Date)%>%
261                       dplyr::summarise(MedianBMT=median(BookMarket)),
262                     by="Date")
263 DiffBM = merge(
264   Firm_yearly%>%
265     filter(!is.na(BookMarket) & BookMarket>MedianBMT & !is.na(value))%>%
266     group_by(Date)%>%
267     dplyr::summarise(AvgTop = mean(value)),
268   Firm_yearly%>%
269     filter(!is.na(BookMarket) & BookMarket<MedianBMT & !is.na(value))%>%
270     group_by(Date)%>%
271     dplyr::summarise(AvgBot = mean(value)),

```

```

272   by = "Date")
273 DiffBM$HMLest = DiffBM$AvgTop-DiffBM$AvgBot
274
275 Firm_yearly =merge(Firm_yearly, DiffBM, by = "Date")
276 Firm_yearly =merge(Firm_yearly, DiffMC, by = "Date")
277
278 Data_year = read_csv("DATA_yearly.csv")
279 Firm_yearly =merge(Firm_yearly, Data_year, by = "Date")
280 Firm_yearly$Date = as.Date(Firm_yearly$Date)
281
282 Firm_yearly <- Firm_yearly[order(Firm_yearly$Company.x, Firm_yearly$Date), ]
283 Firm_yearly$Return <- with(Firm_yearly, ave(value.x, Company.x, FUN = function(x) c(NA, diff(x
    ))))
284
285 Firm_yearly$DCAC40 = c(NA, diff(Firm_yearly$CAC40))
286 Firm_yearly$Doat = c(NA, diff(Firm_yearly$Doat))
287 Firm_yearly$mkt_free = Firm_yearly$DCAC40 - Firm_yearly$Doat
288
289 filtered_Firm_yearly <- Firm_yearly %>% #Filter rows between 2010 and 2022 (no missing values)
290   filter(Date >= as.Date("2010-01-01") & Date <= as.Date("2022-12-31"))
291
292 fit_lm_myFFonly <- function(data) {
293   lm(Return ~ mkt_free+HMLest+SMBest, data = data)
294 }
295 year_by_stock_FF = split(filtered_Firm_yearly, filtered_Firm_yearly$Company.x)
296 regs_beta_myFF <- lapply(year_by_stock_FF, function(subset) fit_lm_myFFonly(subset))
297
298 stargazer(regs_beta_myFF,
299           title = "Estimate the beta coefficients for computed French and Fama factors
    (2010-2022)",
300           column.labels = names(regs_beta_myFF),
301           out="Tables/betas_myff.tex")
302 stargazer(regs_beta_myFF,
303           title = "Estimate the beta coefficients for French and Fama factors",
304           column.labels = names(regs_beta),
305           type = "html", out = "betas_ff.html")
306 dwplot(regs_beta_myFF)
307
308
309
310
311
312
313 #####
314 ## Estimate the lambdas
315 #####
316
317

```



```

318 #Extract the beta coefficients of reg with exo and FF factors
319 # Initialize an empty dataframe to store coefficients and model names
320 coefficients_df <- data.frame(Model = character(), Intercept = numeric(),
321                               beta_pib = numeric(), beta_xr = numeric(),
322                               beta_infl = numeric(), beta_HML = numeric(),
323                               beta_SMB = numeric(), beta_Mkt.RF = numeric(),
324                               stringsAsFactors = FALSE)
325
326 # Iterate through subsets of data and fit models
327 for (i in seq_along(dta_bystock)) {
328   subset <- dta_bystock[[i]]
329   model <- fit_lm_exoff(subset)
330   coefficients <- c(model$coefficients[1], model$coefficients[-1])
331   row <- data.frame(Model = paste0("Model_", i),
332                     Intercept = coefficients[1],
333                     beta_pib = coefficients[2],
334                     beta_xt = coefficients[3],
335                     beta_infl = coefficients[4],
336                     beta_HML = coefficients[5],
337                     beta_SMB = coefficients[6],
338                     beta_Mkt.RF = coefficients[7])
339   coefficients_df <- rbind(coefficients_df, row)
340 }
341
342 coefficients_df
343 coefficients_df$Company = names(regs_beta_exoff) #added column with company name
344
345
346
347 #Get the historical mean return of every stock
348
349 historical_mean_returns <- finalmonthly %>%
350   group_by(Company) %>%
351   summarise(mean_return = mean(Return, na.rm = TRUE))
352
353 #merge two data sets
354 multibeta = merge(coefficients_df, historical_mean_returns, by="Company")
355
356 ## Get return - intercept
357 multibeta$mean_intercept = multibeta$mean_return - multibeta$Intercept
358
359 ## Run std linear model
360 model_multibeta = lm(mean_intercept ~ beta_pib+beta_xt+beta_infl+beta_HML+beta_SMB+beta_Mkt.RF
361                       , data= multibeta)
362 summary(model_multibeta)
363 stargazer(model_multibeta, out='Tables/LM_multibeta.tex')
364

```

```

365 ##### Use GLS to correct the regression
366 library(nlme)
367
368 #get the var-cov matrix of the initial regression to use as weights
369 residuals_list =list()
370 for (i in seq_along(regs_beta)) {
371   residuals <- residuals(regs_beta[[i]])
372   residuals_list[[i]] <- residuals
373 }
374
375 combined_residuals <- do.call(rbind, residuals_list)
376
377 var_cov_matrix_pooled <- cov(combined_residuals)
378
379 inv_var_cov = solve(var_cov_matrix_pooled)
380 corr_struct = corMatrix(inv_var_cov)
381
382 model_multibeta2 = gls(mean_intercept ~ beta_pib+beta_xt+beta_infl+beta_HML+beta_SMB+beta_Mkt.
  RF, data= multibeta, weights=corr_struct)

```

## References

- Pierluigi Balduzzi and Cesare Robotti. Mimicking Portfolios, Economic Risk Premia, and Tests of Multi-Beta Models. *Journal of Business & Economic Statistics*, 26(3):354–368, 2008. ISSN 0735-0015.
- Nai-Fu Chen, Richard Roll, and Stephen A. Ross. Economic Forces and the Stock Market. *The Journal of Business*, 59(3):383–403, 1986. ISSN 0021-9398.
- Eugene F. Fama and Kenneth R. French. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2):427–465, 1992. ISSN 0022-1082. doi: 10.2307/2329112.
- Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993. ISSN 0304-405X. doi: 10.1016/0304-405X(93)90023-5.
- Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, December 1976. ISSN 0022-0531. doi: 10.1016/0022-0531(76)90046-6.