QuantAwards

The Quantitative Finance Competition

2023 winners

- 1st prize: Latent Factors in Private Markets - Bakken & E. Vorpenes - BI Norwegian Business School
- 2nd prize: Interest rate sensitivity of risk measures in European and exotic options - Nicolas Manelli - Paris Dauphine University
- 3rd prize: Single factor modeling, building a financial stress index - Louis Briens - Natixis Investment Managers International

Background information

More than a century after the seminal work of Louis Bachelier, the quantitative approach to financial markets has become omnipresent.

Nowadays, many investment outfits specialize in research, development, and implementation of systematic trading strategies, while other active asset managers have added quantitative strategies to their business lines. Individual clients may also now delegate management of their portfolios to robo-advisors.

These are just a few manifestations of the changing landscape, and we believe that quantitative portfolio management will become ever more important because of the discipline offered by the scientific approach and full automation of the investment process.

The QuantAwards competition offers students a unique opportunity to showcase their creativity and their understanding of this highly timely subject.

CFA Quant Awards 2023

Latent Factors in Private Markets

Abstract

This thesis explores an innovative methodology for identifying latent risk factors in illiquid markets. By uncovering the true underlying volatility and solving for an optimal number of clusters, we identify four latent risk factors. These private factors consist of fund groupings that can only be partially explained by conventional fund characteristics. Furthermore, our study reveals that private markets offer unique exposures, constituting an additional source of factor risk premiums beyond those found in public asset pricing factors. Investors can achieve exposure to these risk premiums and the associated diversification benefits by constructing portfolios based on certain fund characteristics such as geographical focus, investment strategy, and fund size.

1 Introduction

The popularity of private market funds (PMFs) has surged in recent years, with total assets under management (AUM) globally reaching $11.7 trillion as of 2022, an annual growth rate of nearly 20% since 2017 (McKinsey, 2023). Despite the observed decline in actual returns relative to public equity markets, private equity continues to rise in favor. Some investors argue this popularity is due to investors' preference for the low volatility stemming from the return-smoothing properties of illiquid assets, in general, referred to as "volatility laundering" (Asness, 2023). The lack of mark-to-market accounting influenced by valuation changes with a partial lag, results in a smoothing effect that causes the quarterly appraised net asset values (NAVs) to be autocorrelated. The returns understate conventional risk and performance measures, leading to an overstatement of their potential for diversification or a misleadingly high alpha. As a result, the true economic risks and unsmoothed returns are necessary for evaluating risk exposures and risk-adjusted performances.

This working paper is a part of our Master's thesis, in which we leverage machine learning literature to extract underlying asset pricing factors from nonlinear and unbalanced return data that is relevant to private capital. This data has been sourced from Preqin. In the following sections, we offer a concise overview of the topic's significance for the private equity industry, outline the methodology employed, detail the data under examination, and present the obtained results.

2 Literature

Private capital funds allocate their investments to illiquid securities, resulting in delayed or "stale" returns. This can anchor returns to outdated valuation estimates, distorting fund movement with the market, beta, and alphas. Adjusting for illiquidity is essential, with methods like autoregressive unsmoothing (Geltner, 1991, 1993), utilized extensively in real estate. Hedge funds show increased market exposure due to stale prices (Asness, Krail, & Liew, 2001), and hedge fund returns are modeled as a finite moving average of economic returns (Getmansky, Lo, & Makarov, 2004). This approach extends to private equity (Franzoni, Nowak, & Phalippou, 2012), revealing lower diversification benefits and significant exposure to liquidity risk. Our thesis contributes by recovering economic return estimates, distinguishing fund-specific and systematic autocorrelation components, and accurately assessing illiquidity risk. We incorporate autocorrelation through time- and fund-specific variables to reflect true illiquidity risk. This prevents misattribution to factors like audit assumptions, enhancing risk measurement.

The surge of interest in private equity has spurred both investors and researchers to delve into the drivers of risk premiums. Existing debates encompass topics such as basis assets, style analysis, and broader asset pricing factors, aiming to optimize return variance across portfolios. A pivotal advancement traces back to Sharpe (1992), who introduced a quantitative style analysis technique for mutual funds. Our approach builds on the machine learning k-means clustering algorithm (Sebestyen, 1962; MacQueen, 1967), enhancing it for panel data. Meanwhile, the notion of *Basis Assets* has been pivotal,

with Brown and Goetzmann (1997) suggesting an alternative strategy immune to benchmark manipulation. Ahn, Conrad, and Dittmar (2009) proposed a correlation-based method. However, our study deals with autocorrelation and evolving patterns. Leveraging Bonhomme and Manresa's (2015) framework, Goetzmann, Gourier, and Phalippou (2019) presented a flexible econometric technique for latent factor construction in private markets. Our econometric technique adapts k-means clustering for autocorrelation, heteroskedasticity, and unbalanced panels. By categorizing funds into factors that share a systematic component, we offer a novel perspective on private fund alphas and unique factor risk premia.

Overall, the literature on evaluating private equity returns suggests that it is a critical task for both investors and private equity firms. By allowing the benchmarks to be based on true underlying volatility rather than traditional subdivisions, it helps to ensure that investors are receiving a fair return on their investments and that private equity firms are acting in the best interests of their investors. Moreover, it also enables investors and private equity firms to enhance portfolio construction and risk management. As the latest research on private factor performance is a rather unexplored topic, it remains an open question of what drives these risk premiums and whether the higher fees of private equity are justified by higher expected returns over public equity counterparts.

3 Hypothesis

We formulate hypotheses aligning with our research question from Section 1. We hypothesize that autocorrelation coefficients from observed returns are zero for up to 4 lags when categorizing funds by strategy category. We also address the fund-specific and systematic components of observed autocorrelation, enhancing Goetzmann, Gourier, and Phalippou's (2019) method. We then assess private factors' explainability by public factors such as Fama and French (2015) [FF] 5-factor model, the AQR model of Asness, Frazzini, and Pedersen (2018), and the 7-factor model of Fung and Hsieh (2001) (FH). We expect private factors to carry distinct risk premia. We analyze correlations between private factors and macroeconomic variables, examining their unconditional hedging benefits. Our hypotheses cover unsmoothing accuracy, risk factor explainability, and macroeconomic correlations.

4 Methodology

The sample selection procedure yields a final sample of 104,700 quarterly observed returns over 1985-2022 from 2,938 unique funds. The data is summarised in Table 1, and Appendix G provides a description of the data selection. The dataset includes quarterly cash flows and Net Asset Values (NAVs) for each fund. We use observed returns based on the Internal Rate of Return (IRR), which is a commonly used metric. The IRR captures the return earned by investors on average in their PMFs, based on the actual amount of money invested in private capital partnerships. We denote by $R^o_{i,t}$ the observed (smoothed) quarterly return (IRR) process.

4.1 Autocorrelation Structure

We compute the average fund-level autocorrelations for PMF returns across different strategies, for 1 to 4 lags (quarters). As discussed, previous research has found that lagged observed returns in private markets remain significant for up to four quarters. Table 2 presents our findings and contains average autocorrelations for observed returns. The analysis uncovers notable autocorrelations in observed returns. Real estate funds show strong smoothing with substantial autocorrelations for up to two-quarters, revealing valuations updated every six months or audited annually. Private debt strategies also exhibit significant autocorrelations, with distinct patterns reflecting illiquidity challenges and timing of exits. Private equity funds

display less pronounced smoothing, primarily for one and two quarters, potentially influenced by regulatory shifts towards fair value accounting from Q2 2005 (Mathonet and Monhanel, 2006). The average fund-level autocorrelation for the full sample over a single quarter is 0.04, denoting that 4% of the change in a fund's returns in any given quarter could be predicted from the fund's returns in the previous quarter. This is found to be statistically significant for 23% of the funds. When extended to two quarters, the autocorrelation is stronger at 0.07, a pattern that is statistically significant for 20% of the funds. As we proceed to three and four quarters, we notice a decrease in smoothing. Interestingly, when examining four-quarter lags, they show the highest percentage of statistically significant autocorrelations at 28%. This is predominantly attributed to the illiquid assets of Real Estate and Private Debt, suggesting that the valuations in these asset classes are less influenced by short-term market movements. It's worth noting that these findings shed light on the potential influence of asset illiquidity and the reporting cycle on the persistence of fund returns.

4.2 Unsmoothing Returns

Our method assumes $R^o_{i,t}$ to be a weighted average of current and past observed returns but does not impose constraints on weights and mandates a limited number of smoothing lags. In practice, the true economic returns, R_t, are not directly observable. Instead, we have access to the reported or observed returns, denoted as R^o_t. The observed returns are modeled as a weighted average of the fund's true returns over the most recent $k+1$ periods, including the current period. In our empirical analysis of PMFs, we use the model of Goetzmann, Gourier, and Phalippou (2019) to mitigate biases resulting from the NAV appraisal process introducing autocorrelation. The underlying concept is to enable the aggregate and fund-specific elements of returns to be smoothed using distinct weights, thus unsmoothing the systematic component of returns. The objective of this model is to capture the striking balance between the unobserved heterogeneity in unsmoothed returns. In this framework, assumptions regarding the autocorrelation structure are introduced. Each dummy variable is included in the vector $\mathbf{D}_{i,t} \in \mathbb{R}^d$, which collects the relevant dummy variables for fund i at time t. Incorporating dummy variables allows us to find time dependence in autocorrelation by the inclusion of fund-specific characteristics and market conditions. The set of constants that are common across funds and multiplied by d time- and fund-specific dummy variables are collectively represented by the matrix $\boldsymbol{\delta} \in \Theta^{d\times k}$ and are estimated from the cross-section of funds [1]. To express the relationship between the lagged observed returns and the dummy structure, we define the vector of lagged returns as $x_{i,t} = \left(R^o_{i,t-1}, \ldots, R^o_{i,t-k}\right)'$. We can then express the weighted sum of lagged observed returns as:

$$\sum_{j=1}^{k} \theta_{i,j,t} \cdot R^o_{i,t-k} = \mathbf{D}'_{\boldsymbol{i,t}} \boldsymbol{\delta} \boldsymbol{x_{i,t}} \tag{1}$$

We employ a penalized maximum likelihood approach to select the covariates for the intersection terms. This method is applied to the unbalanced panel of unsmoothed returns and aims to determine the optimal loadings on the covariates. The objective is to make the unsmoothed resemble independent and identically distributed (i.i.d.) normally distributed random variables while simultaneously driving insignificant loadings toward zero by applying a penalty to their norm. We utilize both Lasso penalization (Tibshirani, 1996) and elastic net (EN) regularization (Zou and Hastie, 2005) where the regularization parameter is selected through cross-validation. The regularization paths of Lasso (and also Elastic Net) are piecewise linear, which means that a slight increase from the optimal regularization parameter would give us a simpler model with some coefficients becoming zero. This helps us avoid overfitting by making the model less complex. We effectively determine the relevant covariates for modeling the dummies intersection in an efficient and data-driven manner. The EN cost function to

[1] Note that this doesn't include the dummies themselves as we follow the approach of Bonhomme and Manresa (2015) who doesn't allow for additive time-invariant fixed effects. The returns are, therefore, as described in the next section, demeaned

be minimized combines Lasso and Ridge regularization.

4.3 Grouped Patterns of Heterogeneity in Private Markets

To estimate our model, we adopt the time-varying grouped fixed effects model proposed by Bonhomme and Manresa (2015). Fixed-effects models treat heterogeneity as unit-specific and time-invariant. However, in datasets with a large number of parameters, the estimation of common parameters can be biased and the fixed-effects model assumes constant unobserved heterogeneity over time, which may be overly restrictive. Their model incorporates clustered time patterns of unobserved heterogeneity shared within groups of individuals (funds in our case). The time patterns specific to each group and individual fund are left unrestricted and estimated from the data. This approach allows for endogeneity in covariates. Consider the simple linear model with grouped patterns of unit-specific heterogeneity of the following form:

$$R_{i,t} = x'_{i,t}\theta + \alpha_{g_i,t} + \eta_i + v_{i,t},\ \ i = 1, ..., N\ t = 1, ..., T \tag{2}$$

where η_i are N unrestricted parameters. Letting $\bar{w}_i = \frac{1}{T}\sum_{t=1}^{T} w_{i,t}$, the following equation in deviations to the mean:

$$R_{i,t} - \bar{R}_i = (x_{i,t} - \bar{x}_i)'\,\theta + \alpha_{g_i t} - \bar{\alpha}_{g_i} + v_{it} - \bar{v}_i, \tag{3}$$

Thus resulting in the group fixed effect specification of our unsmoothed returns:

$$R_{i,t} = x_{i,t}\theta + \alpha_{g_i,t} + v_{i,t} \tag{4}$$

And our minimization model for the group fixed effects is defined as:

$$(\hat{\theta}, \alpha, \hat{\gamma}) = \arg \min_{(\Theta,\alpha,\gamma)} \sum_{i=1}^{N}\sum_{t=1}^{T} v_{i,t}^2 \tag{5}$$

Equation (2) uses a within transformation involving subtraction of the time-mean of each variable from their individual fund returns. Funds in the same cluster share the same time profile $\alpha_{g,t}$ (e.g., all i such that $g_i = 1$ share the profile α_{1t}). The number of clusters G is set to be estimated. The proposed estimator is based on an optimal grouping of the cross-sectional funds. This grouping is determined by a least squares criterion. The purpose of the group-fixed effects model (4) is to estimate parameters of individual PMFs within the context of commonly shared characteristics or 'clusters'. The unique aspect of this method is that these shared characteristics are dynamic, changing with time, and not fixed. This approach reflects the fact that the environment in which these funds operate is continuously evolving due to multiple factors like changing economic conditions, market dynamics, and regulatory changes among others. This allows the model to capture both the individual uniqueness of each fund as well as shared characteristics within a specific timeframe. This effect is represented by the latent group returns $\alpha_{g_i,t}$, which can be perceived as the latent factor that drives the returns of these funds. This can be compared to a latent factor model where each group represents a factor, and all funds within a group share a uniform factor load beta β of one. Appendix F provides further information on the machine learning algorithm, as well as Monte Carlo simulations demonstrating the robustness of the estimators in smaller sample sizes.

5 Results

In our study, we use fund return data to test various hypotheses concerning autocorrelation patterns. We construct interaction covariates that combine the first two lagged returns with dummy variables representing fund attributes (vintage year, size quartile, strategy, geography) and economic conditions (recession, extraordinary public market returns), along with the fourth lagged return for year-end audits. The resulting common covariates are depicted in Table 3 after employing Lasso penalization and elastic net regularization. The loadings, estimated via panel regressions without penalization, are shown, as well as the common model parameters ($\hat{\boldsymbol{\theta}}$). The analysis reveals a link between public markets and private fund returns' autocorrelation, with larger first lagged returns during market downturns. US-focused and Large-Capitalized funds relate to the first-lagged return, albeit with minimal economic impact. The second lagged return remains significant, indicating strong autocorrelation for the full sample looking back two quarters, particularly responsive to previous abnormal returns. Surprisingly, fund strategy or age does not significantly affect autocorrelation structure, suggesting other factors shape these patterns for the overall dataset. Table 4 lays out the goodness-of-fit indicator for the ideal number based on varying cluster groups. The BIC statistic penalizes over-fitting and is minimal (signifying optimal configuration) when the cluster number is four. By estimating the cluster groups using the BIC criterion, an upper limit of the actual groups is provided, and hence the θ estimator is consistent.

5.1 Characteristics of Private Factors

In this section, we outline how investors can select funds that align with their preferred proprietary factor. In our dataset, every fund is specifically assigned to a group that shares similar return variations, which is explained by our eight private factors. Table 5 presents the percentage of funds in a particular category divided amongst the four private factors (rows sum to 100%). We also provide the average fund size (USDm) for each fund type, and how much each Factor is comprised of the strategies in weighted size (i.e. all strategy columns within a Factor sum to 100%). The findings reveal limitations in conventional strategy categorizations for capturing PMFs' distinct risk-return profiles and returns patterns. However, we've identified attributes that offer insight into fund characteristics. This enables potential investors to strategically build portfolios based on attributes such as investment region, strategy, and fund size. To summarize, Factor 1 includes Debt, Buyout, and European-focused funds, while Factor 2 comprises mainly US-focused Buyout and RE Opportunistic funds. Factor 3 is dominated by large Buyout and Asia-focused funds, showcasing opportunities in emerging markets. Factor 4 is primarily composed of Real Estate funds, including small-cap ones, indicating a widespread ownership of real estate holdings among PMFs, regardless of their focus. These attribute-based insights better capture the diverse nature of PMF strategies and assist in constructing informed portfolios.

Table 6 presents the descriptive statistics for our four private factors, while Figure 1 illustrates the time series of time-specific latent group returns ($\alpha_{g_i,t}$). These factors are designed with a zero-mean and non-tradable nature due to the methodology employed. However, by analyzing fund traits within each cluster, we can identify factor exposure. These factors encompass Debt funds, US-focused, Large Equity, and Real Estate, with the highest market capitalization (USDm) belonging to Factor 3 (Large Buyout) at 27% of the sample. The factors exhibit similar levels of volatility, with Factor 1 (Debt & BO) being the most volatile at 13%, and Factor 3 (Large PE) being the least at 10%. Despite the limited number of groups, average autocorrelations for the factors are relatively small, averaging at 19%. The Mean Squared Error (MSE), representing the minimized objective function, is small at an average of 1.26%. The Correlation matrix between the private factors is presented in Table 7.

In our subsequent analysis, we evaluate how much of the private factors can be explained by public standard equity as-

set pricing models. We include both US and Global factors as our funds are mostly US-focused. We use the risk-free rate corresponding to each dataset.The first model encompasses the standard Fama and French (FF) (2015) 5-factor model augmented with the liquidity factor of Pastor and Stambaugh (2003). The second model derived from AQR's website includes the Fama-French 3 factor model augmented with the QMJ and BAB factors of Asness, Frazzini, and Pedersen (2018). The last model, derived from David. A Hsieh's data library includes the Fung and Hsieh (2001) factors from constructing look-back straddles on five distinct option markets. Table 8 presents results for the Global FF, AQR, and FH model. Given that our private factors are non-tradable assets, the coefficients in the table should not be interpreted as conventional alphas and betas. These factors, designed with zero means, result in constants near zero. Factor loadings can serve as proxies for betas, and some factor return time series start later than others. The FF factors start in Q3 1990, while AQR factors start in Q3 1989. There's a notable correlation between the public equity risk premium (MKT) and the four private factors, particularly Factors 1 (Debt funds) and 3 (Large Buyout) with loadings of 0.26 and 0.34 respectively. Factors 2 (US-focused) and 4 (Real Estate & Small Cap) also show significant but less strong loadings. A negative loading exists between the Size (SMB) factor and private Factor 2 (US-focused and RE). Factors 2 and 4 exhibit a negative significant loading on the profitability (RMW) factor. Factor 3 (Large Buyout) has a positive significant loading on the value (HML) factor due to the presence of Large Equity funds. The augmented liquidity (LIQ) factor has a positive significant loading on Factor 3. In the Fung-Hsieh option model, the bond option lookback straddle (BD) shows a negative significant loading on Factor 3 due to leverage and refinancing behaviors. Public factors explain a significant portion of the time-series variation in private factors, most notably for Factors 3 and 4, although not all private factors are captured by these models, suggesting limitations in replicating private market returns through passive index strategies.

Previous research has emphasized the link between business cycles and private fund returns, motivating our study of unconditional hedging benefits provided by our private factors. Hence, we study the extent to which our private factors provide unconditional hedging benefits. Table 9 presents our findings. Factor 3 (Large Buyout) loads significantly positive on the S&P Infrastructure Index, possibly due to infrastructure investments by buyout funds. Factor 4 (RE and Small-Cap) loads negatively on S&P Infrastructure and positively on the VC index. Factor 4 also loads positively on the Natural Resources and Bloomberg Commodity indices. Factor 3 shows positive loadings on the Bloomberg Commodity Index and Dow Jones Real Estate Investment Trust (REIT) index. Factor 3, dominant in Large Buyout funds, serves as a significant hedge against inflation and shows sensitivity to business cycles. All factors exhibit pro-cyclical behavior, particularly Factor 3. The credit spread (BofA US High Yield Index Option-Adjusted Spread) has a negative significant impact on all factors, in line with increases in credit spreads boosting risk premiums. Large BO funds exhibit a positive loading on high-yield bonds, aligning with past research indicating BO investments are more expensive during high interest rate periods.

5.2 Risk-Adjusted Performance of Private Market Funds

Our approach aims to improve performance measurement accuracy in private market funds by examining the impact of unsmoothing returns on risk-adjusted performance assessment. Unsmoothing increases average fund volatility while maintaining average returns, aligning risk assessment with that of public equities. Figure 3 shows statistics for each underlying PMF strategynbased on observed returns $R^o_{i,t}$ and unsmoothed returns $R_{i,t}$. Regression analysis using four and eight private factors reveals their explanatory power in assessing fund performance. Unsmoothing notably decreases alphas, particularly for less liquid strategies, and has minimal impact on systematic risk assessment. These findings highlight the economic significance of unsmoothing returns, offering a more informed benchmark for risk-adjusted performance evaluation across strategies, particularly for relatively illiquid funds.

6 Conclusion

The study's findings reveal the inadequacy of conventional strategy categorizations in capturing the nuanced risk-return profiles and return patterns of Private Market Funds (PMFs). Latent risk factors in PMFs are only partially explained by observable attributes such as strategy and geography, and public factors also offer partial insight. Despite challenges posed by limited data and return smoothing, private markets offer a captivating area of study due to their unique characteristics. The lack of mark-to-market accounting can understate asset volatility, inflate risk-adjusted performance, and necessitate a rigorous assessment of private fund characteristics. The research aligns with prior studies, identifying significant autocorrelations and justifying the return smoothing in illiquid assets. By introducing an innovative clustering method, the study uncovers latent factors within private markets, providing crucial investment insights. These private factors contribute to diversification, exposure to risk premia, and improved risk-adjusted performance measurement. Notably, certain factors offer unconditional inflation-hedging benefits, and most private funds display pro-cyclical behavior. While the findings inform decision-making and carry regulatory implications, the study acknowledges limitations and encourages further research to better understand private market premiums and risks.

7 Appendices

A Methodology

Table 1: Distribution of funds across categories and regions

Categorization			America		Europe	Asia	Oceania	Middle East	Africa	Multi-Regional	All Regions	All Regions
Tier 1	*Tier 2*	*Tier 3*	North	South								
Debt			**261**	**5**	**41**	**12**	**0**	**0**	**1**	**0**	**323**	**11%**
	Mezzanine	Mezzanine	113	2	13	0	0	0	1	0	129	4%
	Distressed	Distressed	83	0	12	5	0	0	0	3	103	4%
	Special Situations	Special Situations	34	3	12	7	0	0	0	0	56	2%
	Other Debt	Other Debt	31	0	4	0	0	0	0	0	35	1%
Real Assets			**357**	**22**	**57**	**38**	**0**	**0**	**1**	**0**	**475**	**16%**
		RE Opportunistic	124	17	27	30	0	0	1	0	199	7%
		RE Value Added	204	5	29	8	0	0	0	0	246	8%
		RE Distressed	16	0	0	0	0	0	0	0	16	8%
		RE Secondaries	13	0	1	0	0	0	0	0	14	0%
Equity			**1,492**	**35**	**376**	**133**	**19**	**21**	**10**	**3**	**2,089**	**71%**
	Venture Capital	Seed/Start-Up	32	0	4	2	0	4	0	0	42	1%
		Early Stage	175	1	14	5	2	7	0	0	204	7%
		General VC	318	1	35	14	0	5	0	0	373	13%
		Expansion/Late-Stage	69	1	2	3	1	0	0	0	76	3%
	Buyout	Buyout	690	20	251	52	16	4	4	1	1038	35%
	Growth	Growth	117	11	27	56	0	0	6	2	219	7%
	Turnaround	Turnaround	13	0	1	0	0	0	0	0	14	0%
	Secondaries	Secondaries	72	1	40	1	0	1	0	0	115	4%
		Direct Secondaries	6	0	2	0	0	0	0	0	8	0%
Generalist	Generalist	Generalist	30	1	14	2	0	0	0	4	**51**	**2%**
		All types	**2,140**	**63**	**488**	**185**	**19**	**21**	**12**	**10**	**2938**	**100%**
		All types	**73%**	**2%**	**17%**	**6%**	**1%**	**1%**	**0%**	**0%**	**100%**	

The table presents the distribution of funds according to the three-tier classification system as defined by Preqin. The third tier has been formulated using a mix of attributes provided by Preqin. Each classification is further detailed by the geographic areas of investment focus.

Table 2: Autocorrelations of Private Fund Returns

Strategies	Returns $R^o_{i,t}$			
	τ_1	τ_2	τ_3	τ_4
Categorization Strategies	Observed Returns $R^o_{i,t}$			
	τ_1	τ_2	τ_3	τ_4
Mezzanine	0.00 (29%)	0.06 (25%)	0.01 (19%)	−0.01 (21%)
Distressed Debt	0.05 (29%)	0.07 (20%)	−0.01 (12%)	−0.06 (26%)
Special Situations	0.03 (30%)	0.06 (11%)	0.01 (16%)	−0.06 (30%)
Other Debt	0.02 (26%)	0.05 (20%)	0.05 (23%)	−0.13 (40%)
RE Opportunistic	0.01 (28%)	0.05 (30%)	0.00 (16%)	0.09 (31%)
RE Value-Added	0.03 (28%)	0.11 (25%)	0.03 (19%)	0.05 (38%)
RE Distressed	0.15 (38%)	0.26 (38%)	0.04 (19%)	−0.07 (19%)
RE Secondaries	0.02 (57%)	0.11 (29%)	0.05 (21%)	0.11 (29%)
VC: Seed/Start-Up	0.06 (20%)	0.10 (24%)	−0.02 (12%)	−0.01 (39%)
VC: Early Stage	0.08 (26%)	0.05 (12%)	0.02 (17%)	0.02 (33%)
VC: General	0.08 (21%)	0.07 (19%)	0.05 (17%)	0.02 (29%)
Expansion / Late-Stage	0.06 (21%)	0.07 (25%)	−0.01 (16%)	−0.03 (30%)
Buyout	0.05 (20%)	0.08 (20%)	−0.00 (14%)	0.01 (25%)
Growth	0.02 (19%)	0.07 (19%)	0.00 (11%)	0.01 (25%)
Turnaround	0.16 (14%)	0.12 (36%)	0.04 (43%)	−0.06 (29%)
Secondaries	0.00 (20%)	0.04 (10%)	−0.01 (23%)	−0.00 (32%)
Generalist	0.06 (12%)	0.09 (18%)	0.05 (16%)	0.04 (24%)
Full Sample	**0.04 (23%)**	**0.07 (20%)**	**0.01 (16%)**	**0.01 (28%)**

The table shows average fund-level autocorrelations across strategies. Autocorrelations use observed returns. Percentages in brackets indicate funds where the autocorrelation is significant at 10% level. Confidence interval: $\pm 1.645 \times \frac{1}{\sqrt{T}}$. If τ_j falls outside this range, we reject $\mathcal{H}_1$ that the coefficient at lag j is zero.

B Results

Table 3: Common Covariates of the interaction terms

Selected Covariates	Coefficient	t-stat	$\hat{\boldsymbol{\theta}}$
$R^o_{i,t-1} \cdot \mathbf{1}(R^o_{i,t-1} < \mu_{S\&P} - \sigma_{S\&P})$	0.062	4.46	0.010
$R^o_{i,t-1} \cdot \mathbf{1}$(North America)	0.007	0.96	-0.018
$R^o_{i,t-1} \cdot \mathbf{1}$(Large-Cap)	0.030	2.93	0.012
$R^o_{i,t-2}$	0.015	6.14	0.004
$R^o_{i,t-2} \cdot \mathbf{1}\left(R^o_{i,t-2} > 15\%\right)$	0.165	7.28	0.137
$R^o_{i,t-2} \cdot \mathbf{1}\left(R^o_{i,t-2} < -15\%\right)$	0.155	4.52	0.103

t-stats of the hypothesis test $\mathcal{H}_{\mathbf{2}} : \delta = \mathbf{0}$ showing common covariates where we reject that the interaction coefficient is equal to zero for j lags for the full sample.

Table 4: Selection of the number of clusters

	BIC
2 clusters	1.481
4 clusters	**1.395**
6 clusters	1.434
8 clusters	1.743
10 clusters	1.795
12 clusters	1.572
14 clusters	1.947
16 clusters	1.674

Table 5: Descriptive Statistics - Mapping of funds to factors

	Factor1			Factor2			Factor3			Factor4		
	$N\%$	Size	Weight	$N\%$	Size	Weight	$N\%$	Size	Weight	$N\%$	Size	Weight
Mezzanine	28	646	2	21	861	3	14	1,661	3	14	581	3
Distressed	36	2,037	10	24	2,036	6	17	1,725	4	23	1,694	4
Special Situations	48	1,415	3	14	1,415	1	13	740	1	25	1,021	1
Other Debt	31	1,118	1	34	538	1	14	308	0	20	348	0
Debt	**35**	**1,427**	**16**	**21**	**1,330**	**10**	**14**	**1,594**	**7**	**30**	**939**	**9**
RE Opportunistic	25	1,379	7	20	2,013	9	15	1,028	3	40	1,592	13
RE Value-Added	20	571	2	33	683	5	9	999	2	38	596	5
RE Distressed	19	996	0	13	210	0	17	331	0	51	1,274	1
RE Secondaries	14	1,267	0	36	400	0	14	343	0	36	828	0
Real Estate	**22**	**1,056**	**9**	**27**	**1,117**	**14**	**12**	**961**	**5**	**39**	**1,097**	**19**
VC Early Seed/Start-Up	7	120	0	31	196	0	26	246	0	36	199	0
VC Early Stage	27	342	2	29	287	2	27	299	2	18	263	1
VC General	25	384	5	25	383	4	32	430	6	19	370	4
Expansion / Late-Stage	28	448	1	29	448	1	32	336	1	12	379	0
Buyout	25	1,904	56	25	1,832	54	23	2,772	70	27	1,604	53
Growth	25	913	5	17	605	2	28	817	6	30	817	6
Turnaround	20	462	0	25	504	1	18	417	0	37	425	0
Secondaries	22	1,393	3	24	1,397	6	22	1,264	2	33	1,342	8
Equity	**23**	**1,068**	**72**	**26**	**1,064**	**70**	**24**	**1,131**	**86**	**27**	**964**	**72**
Generalist	35	1,392	3	27	3,651	6	24	1,483	2	14	625	1
All Private funds	**25**	**1,219**	**100**	**25**	**1,236**	**100**	**22**	**1,485**	**100**	**27**	**1,129**	**100**
Africa	25	301	0	33	198	0	25	212	0	17	207	0
Asia	31	1,345	7	9	1,391	5	39	1,594	10	20	1,510	6
Europe	31	1,919	24	20	1,524	8	21	1,893	23	28	1,961	25
Multi-regional	40	780	1	10	608	0	10	630	0	10	723	0
Oceania	21	492	0	16	400	0	21	370	0	42	417	0
Middle East	14	465	0	29	537	0	33	518	0	24	537	0
North America	23	2,206	67	28	2,299	86	21	2,210	67	27	2,250	68
South America	41	537	1	16	430	0	25	444	0	17	495	1

The table provides a statistical overview of the allocation of funds across four Private Factors, based on the estimates from equation (13). It illustrates the proportion of funds allocated to each of the four private factors from different fund types. Fund types are constructed by combining fund classifications as determined by Preqin. Accompanying these proportions, the average fund size (USDm) for each corresponding subclass is also presented. Weight provides the percentage in weighted size each strategy constitutes of the private factor.

Table 6: Summary statistics of Private Factors

Factor	% *N fund*	% *Capitalization*	β	σ	MSE	AR(1)
1	25%	25%	1.03	13%	1.20	0.28
2	25%	24%	1.11	11%	1.19	0.05
3	22%	27%	1.08	10%	1.37	0.16
4	27%	24%	1.08	12%	1.26	0.27

The table presents data for each private factor, detailing the proportion of funds tied to the factor (% *Nfund*), the total market capitalization these funds comprise (% *Capitalization*), the average beta for funds contributing to the factor (β), the annualized volatility (σ) of the factor returns, the overall mean square error (MSE) for funds contributing to the factor, and the first-order autocorrelation AR(1) of the factor returns.

Table 7: Correlation matrix of Private Factors

Factors	1	2	3	4
1	1			
2	0.23	1		
3	0.27	0.20	1	
4	0.45	0.10	0.26	1

Figure 1: Time-series of Factors

The plotted graphs below exhibit the four private factors. Each unsmoothed return from the respective funds is characterized by the formula $R_{i,t} = x_{i,t}\theta + \alpha_{g_i,t} + v_{i,t}$. These private factor time series reflect the latent group returns, $\alpha_{g_i,t}$, associated with each cluster of funds g_i. The factor returns are compared to the S&P 500 index.

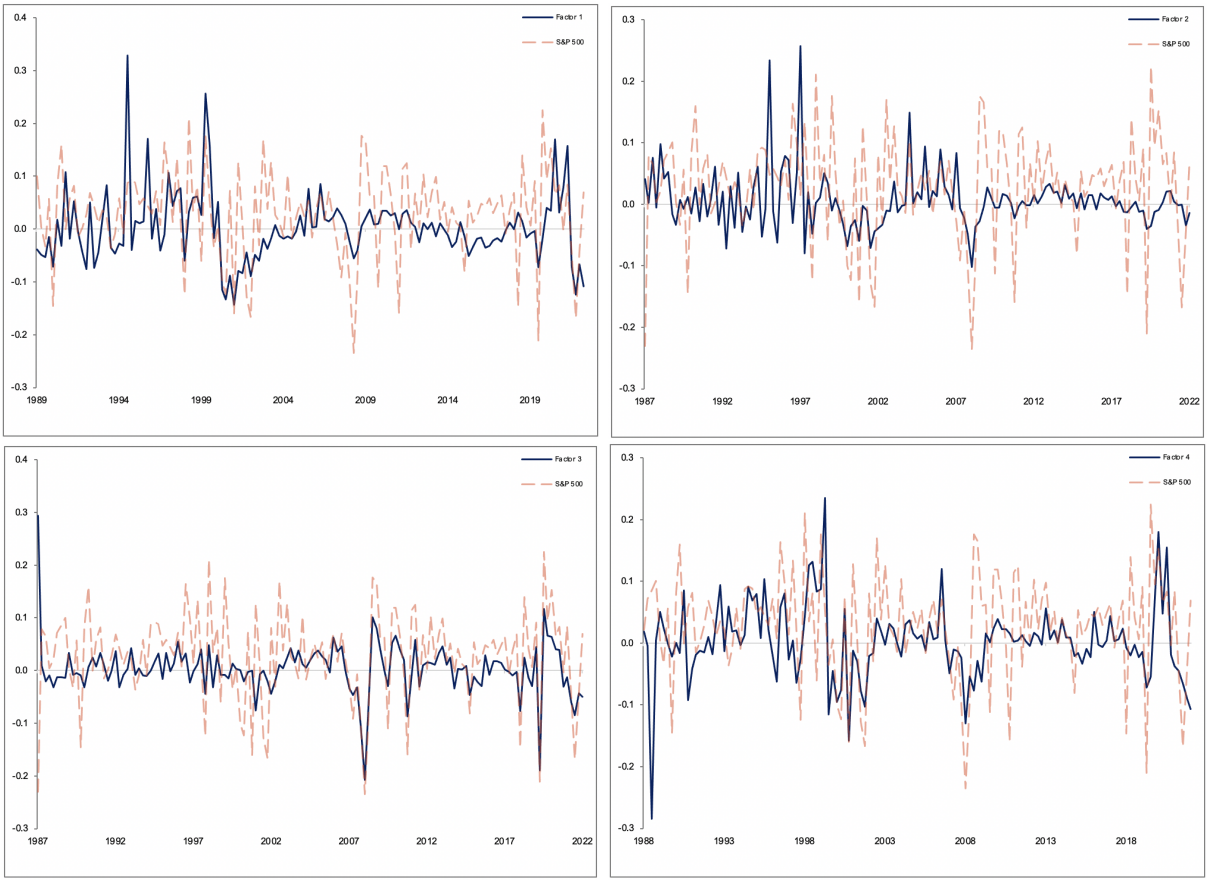


Table 8: Private factors vs. public factors

Global FF	Factor 1	Factor 2	Factor 3	Factor 4
MKT	0.259	0.112	0.342	0.146
	(*3.461*)***	(*2.193*)**	(*9.041*)***	(*2.335*)**
SMB	-0.010	-0.234	0.034	0.079
	(*-0.838*)	(*-2.644*)***	(*0.568*)	(*0.793*)
HML	0.023	0.144	0.201	0.047
	(*0.859*)	(*1.505*)	(*3.088*)***	(*0.435*)
RMW	-0.155	-0.188	0.142	-0.390
	(*-1.252*)	(*-2.038*)**	(*2.272*)**	(*-3.774*)***
CMA	-0.239	-0.043	-0.157	-0.160
	(*-1.272*)	(*-0.307*)	(*-1.651*)	(*-1.022*)
LIQ	-0.061	0.040	0.115	0.015
	(*-0.871*)	(*0.766*)	(*3.232*)***	(*0.263*)
R^2	20%	11%	55%	29%
Global AQR				
MKT	0.213	0.011	0.304	0.108
	(*2.440*)**	(*0.168*)	(*7.280*)***	(*1.484*)
SMB	-0.068	-0.425	-0.002	0.357
	(*-0.331*)*	(*-2.800*)	(*-0.019*)	(*2.088*)**
$HMLd$	-0.290	-0.083	0.018	-0.263
	(*-3.083*)***	(*-1.196*)	(*0.401*)	(*-3.349*)***
QMJ	-0.151	-0.450	-0.119	-0.260
	(*-0.687*)	(*-2.758*)***	(*-1.132*)	(*-1.412*)
BAB	-0.091	0.090	0.287	-0.071
	(*-0.885*)	(*1.175*)	(*5.812*)***	(*-0.819*)
R^2	19%	10%	59%	26%
Fung and Hsieh				
BD Option	-0.020	-0.017	-0.045	-0.007
	(*-1.156*)	(*-1.365*)	(*-4.900*)***	(*-0.477*)
FX Option	0.025	0.016	-0.012	-0.019
	(*1.175*)	(*1.059*)	(*-1.053*)	(*-1.046*)
COM Option	0.007	0.018	0.002	0.049
	(*0.583*)	(*0.88*)	(*0.124*)	(*1.968*)*
IR Option	-0.025	-0.016	-0.020	-0.034
	(*-1.756*)*	(*-1.586*)	(*-2.619*)***	(*-2.783*)***
STK Option	-0020	-0.011	-0.015	-0.013
	(*-0.888*)	(*-0.713*)	(*-1.280*)	(*-0.667*)
R^2	7%	7%	43%	14%

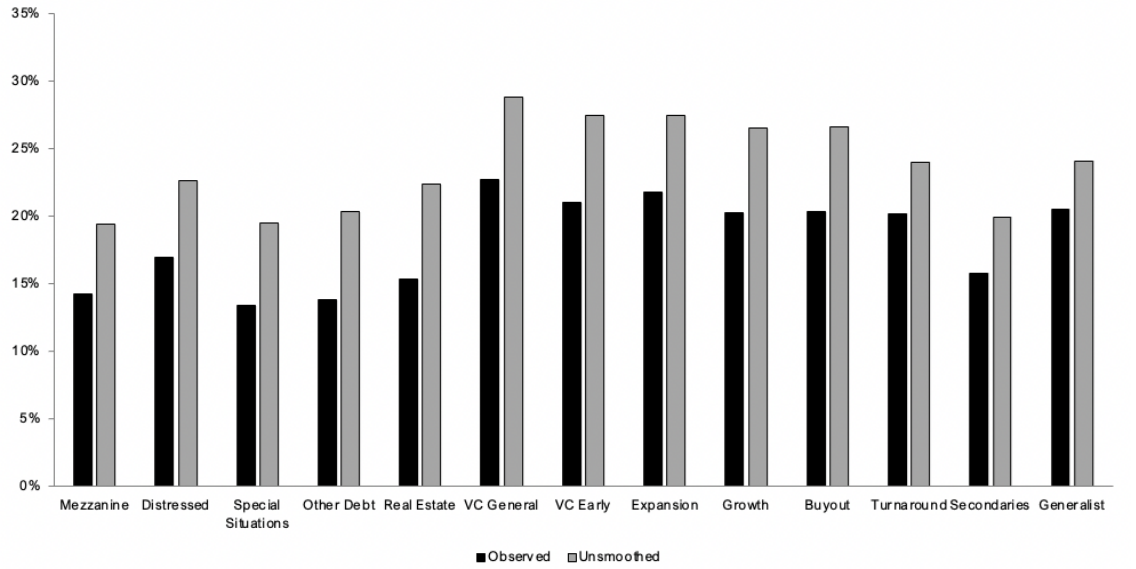
The table shows the estimated coefficients from regressions of the excess return for the four private factors on public asset pricing models. t-statistics are in parentheses of the hypothesis test $\mathcal{H}_3$: The returns of private factors can not be significantly explained by public factors. R^2 is the explanatory power from the regression. ***/**/* indicates significance at the 1%/5%/10% level.

Table 9: Private factors on other variables

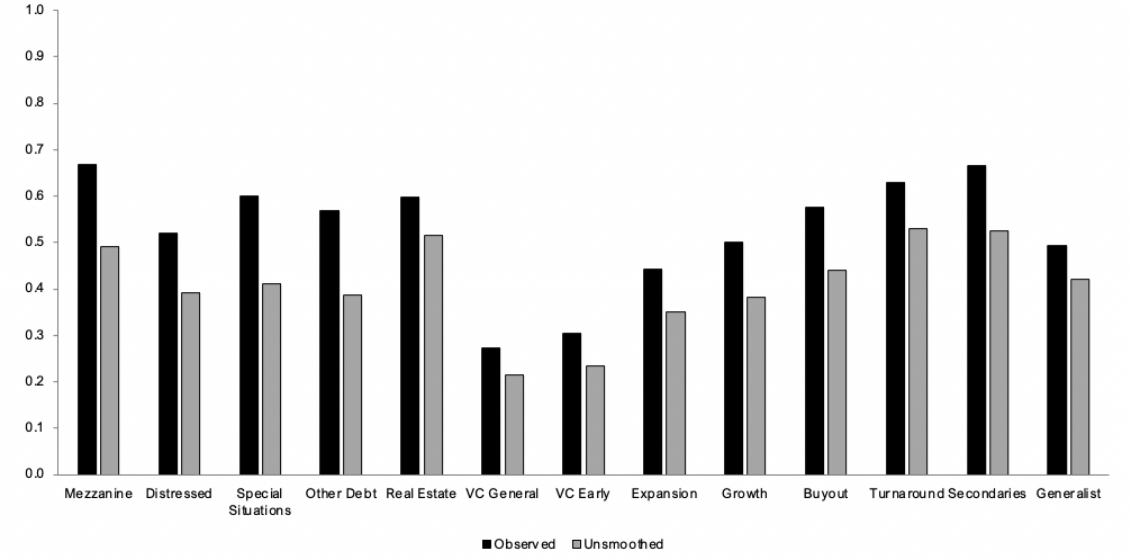
Macroeconomic Variables	Factor 1	Factor 2	Factor 3	Factor 4
US Inflation	0.330	-0.706	1.004	0.523
	(*0.53*)	(*-1.508*)	(*2.534*)***	(*0.872*)
Industrial Production	0.341	0.222	0.726	0.360
	(*0.867*)	(*0.751*)	(*2.900*)***	(*0.949*)
BofA Credit Spread	-0.663	-0.876	-0.632	-0.837
	(*-2.971*)***	(*-5.224*)***	(*-4.452*)***	(*-3.895*)***
R^2	16%	29%	42%	24%
Alternative Indices				
BB Commodity	0.095	0.085	0.172	0.120
	(*1.321*)	(*1.561*)	(*3.804*)***	(*1.790*)*
DJ US REIT	0.011	0.024	0.113	0.040
	(*0.180*)	(*0.518*)	(*2.941*)***	(*0.699*)
BofA US High Yield	0.119	0.098	0.143	-0.130
	(*0.966*)	(*1.041*)	(*1.827*)*	(*-1.132*)
R^2	3%	4%	25%	4%
Private Indices (Geo)				
Preqin Europe	-0.003	0.048	0.289	-0.225
	(*-0.030*)	(*0.671*)	(*3.190*)***	(*-2.042*)**
Preqin Asia	-0.430	0.047	0.216	0.292
	(*-3.257*)***	(*0.537*)	(*1.97*)*	(*2.180*)**
Preqin US	1.118	0.460	0.482	0.798
	(*7.616*)***	(*4.732*)***	(*3.949*)***	(*5.354*)***
R^2	49%	48%	65%	48%
Private Indices (Strategy)				
Preqin BO	0.147	0.370	0.731	0.367
	(*0.643*)	(*3.949*)***	(*5.539*)***	(*2.002*)**
Preqin VC	0.413	-0.067	0.080	0.596
	(*2.369*)**	(*-0.943*)	(*0.796*)	(*4.253*)***
Preqin Natural Resources	0.156	0.083	-0.072	0.145
	(*0.973*)	(*1.265*)	(*-0.779*)	(*1.121*)
S&P Infrastructure	-0.054	-0.045	0.230	-0.200
	(*-0.663*)	(*-1.336*)	(*4.898*)***	(*-3.065*)***
R^2	33%	49%	85%	61%

The table presents the private factor loadings on macroeconomic variables, alternative indices associated with illiquid markets, and alternative stock indices. t-statistics are in parentheses of the hypothesis test $\mathcal{H}_4 : \beta = \mathbf{0}$. R^2 is the explanatory power from the regression. ***/**/* indicates significance at the 1%/5%/10% level.

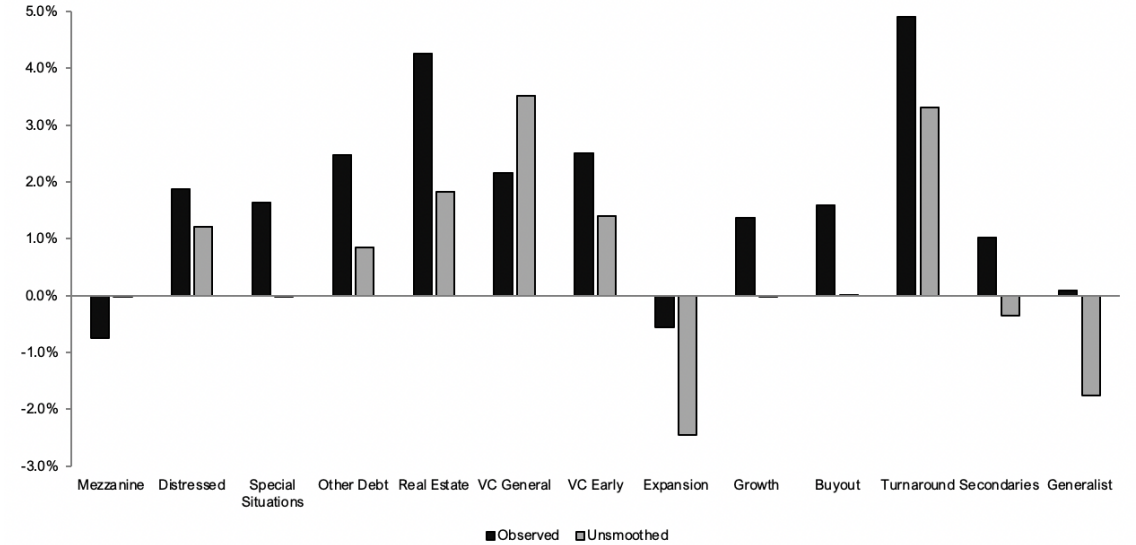
(a) Average σ



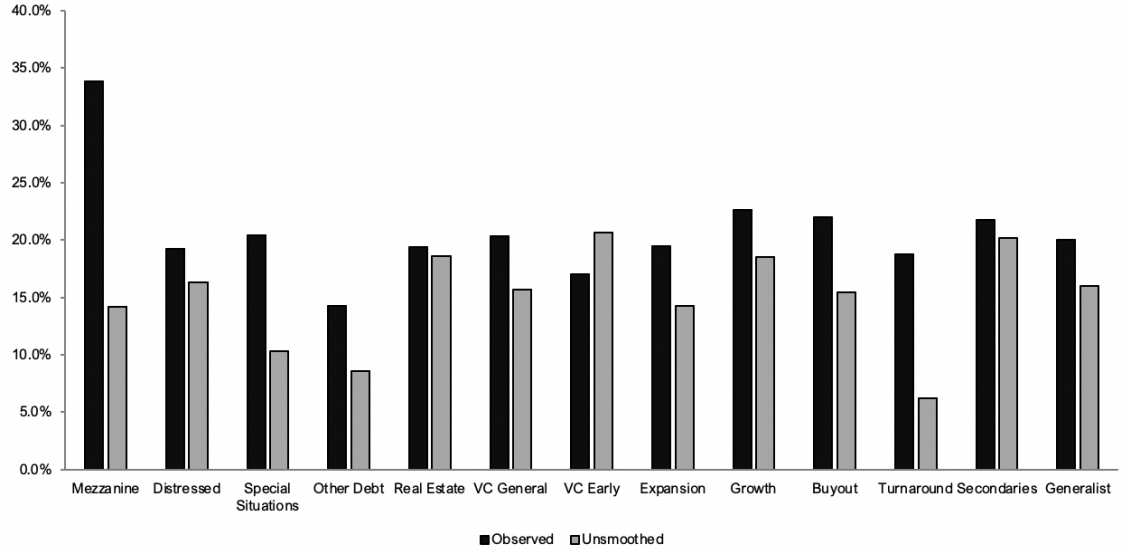
(b) Average Sharpe Ratios



(c) Average α (Private 8-Factor Model)



(d) % of Funds with significant α at 10%



(e) Average R^2 (Private 8-Factor Model)

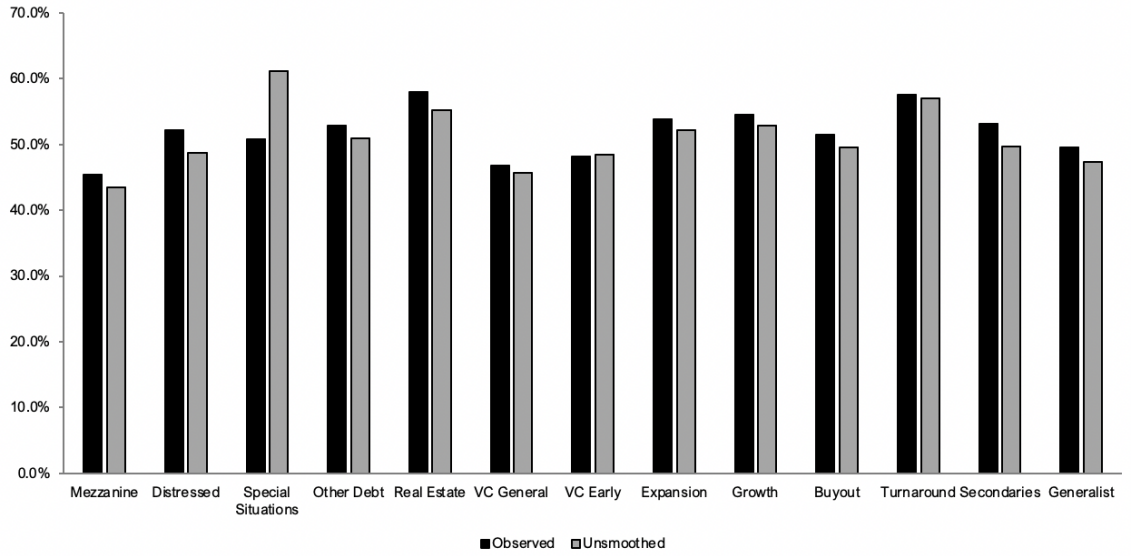


Figure 3: PMF Risk and Performance by Strategy

The figures plots average fund-level results by strategies using observed and unsmoothed returns from January 1995 to December 2022.

C References

Agarwal, V., T. C. Green, and H. Ren, 2018, "Alpha or beta in the eye of the beholder: What drives hedge fund flows?" *Journal of Financial Economics* 127(3), 417–434.

Ahn, D.-H., Conrad, J., and Dittmar, R. F., 2009, "Basis Assets," *The Review of Financial Studies*, 22(12), 5133–5174.

Aigner, P., Albrecht, S., Beyschlag, G., Friederich, T., Kalepky, M., and Zagst, R., 2008, "What Drives PE: Analyses of Success Factors for Private Equity Funds," *The Journal of Private Equity*, 11(4), 63-85.

Amin, G. S., and, Kat, H. M., 2003, "Hedge Fund Performance 1990-2000: Do the "Money Machines" Really Add Value?, " *The Journal of Financial and Quantitative Analysis*, 38(2), 251-274.

Ang, A., D. Papanikolaou, and M. M. Westerfield, 2014, "Portfolio Choice with Illiquid Assets," *Management Science*, 2737 – 2761.

Anson, M. 2017. "Measuring Liquidity Premiums for Illiquid Assets," *The Journal of Alternative Investments* 20(2), 39–50.

Aragon, G., 2007, "Share restrictions and asset pricing: Evidence from the hedge fund industry," *Journal of Financial Economics* 83, 33–58.

Aragon, G. O. and V. Nanda, 2011, "Tournament behavior in hedge funds: High-water marks, fund liquidation, and managerial stake". *The Review of Financial Studies* 25(3), 937–974.

Asness, C., Krail, R., and Liew, J.. 2001, "Do Hedge Funds Hedge?", *Journal of Portfolio Management.*

Asness, C. S., 2023, "Volatility Laundering," AQR Capital Management.
https://www.aqr.com/Insights/Perspectives/Volatility-Laundering.

Asness, C. S, and A. Frazzini, 2013, "The devil in HML's details," *The Journal of Portfolio Management*, 39(4), 49–68.

Asness, C. S., A. Frazzini, and L. H. Pedersen, 2018, "Quality minus junk," *Review of Accounting Studies*, 1–79.

Barber, B. M., and Yasuda, A., 2017, "Interim fund performance and fundraising in private equity," *Journal of Financial Economics*, 124(1), 172–194.

Barkham, R. and D. Geltner, 1995, "Price discovery in American and British property markets," *Real Estate Economics* 23(1), 21–44.

Barth, D. and P. Monin, 2018, "Illiquidity in Intermediary Portfolios: Evidence from Large Hedge Funds," Working Paper.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon, 2012, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics* 104(3), 535–559.

Bollen, N. P., and Pool, V. K., 2008, "Conditional Return Smoothing in the Hedge Fund Industry," *The Journal of Financial and Quantitative Analysis*, 43(2), 267–298.

Bonhomme, S., and Manresa, E., 2015, "Grouped Patterns Of Heterogeneity in Panel Data," *Econometrica*, 83(3), 1147–1184.

Brown, S. J., and Goetzmann, W. J., (1997), "Mutual fund styles," *Journal of Financial Economics*, 43(3), 373-399.

Brown, G. W., O. Gredil, and S. N. Kaplan, 2017, "Do Private Equity Funds Manipulate Reported Returns?," *Journal of Financial Economics*.

Campbell, R., 2008, "Art as a financial investment," *Journal of Alternative Investments*, 10, 64–81.

Cao, C., Chen, Y., Liang, B., and Lo, A. W., 2013, "Can hedge funds time market liquidity?," *Journal of Financial Economics*, 109(2), 493-516.

Cassar, G., and Gerakos, J., 2011. "Hedge Funds: Pricing Controls and the Smoothing of Self-reported Returns," *Review of Financial Studies, Society for Financial Studies*, 24(5), 1698-1734.

Chakraborty., and Ewens. M, 2018, "Managing Performance Signals Through Delay: Evidence from Venture Capital," *Management Science, INFORMS*, vol. 64(6), 2875-2900.

Crain, N. G., and K. K. F. Law, 2016, "Spreading Sunshine: Fair Value Accounting and Information Production," Working Paper.

Couts, S., Goncalves, A., and Rossi, A., 2020, "Unsmoothing Returns of Illiquid Funds," *Kenan Institute of Private Enterprise Research Paper*, 20(5), USC Lusk Center of Real Estate Working Paper Series.

Dimson, E., 1979, "Risk measurement when shares are subject to infrequent trading," *Journal of Financial Economics*, 7(2), 197-226.

Dimson, E., and C. Spaenjers, 2011, "Ex post: The investment performance of collectible stamps," *Journal of Financial Economics*, 100(2), 443–458.

Fama, E. F., and K. R. French, 2015, "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fisher, J., D. Gatzlaff, D. Geltner, and D. Haurin, (2003).,"Controlling for the impact of variable liquidity in commercial real estate price indices, *Real Estate Economics*.

Fisher, J. D., D. M. Geltner, and R. B. Webb (1994). "Value indices of commercial real estate: a comparison of index

construction methods". *The Journal of Real Estate Finance and Economics* 9(2), 137–164.

Fortado, L., and Wigglesworth, R., 2018, "Man Group and AQR try to take aim at private equity industry," Financial Times, https://www.ft.com/content/d352d308-6d83-11e8-92d3-6c13e5c92914.

Franzoni, F., Nowak, E., and Phalippou, L., 2012, "Private Equity Performance and Liquidity Risk," *The Journal of Finance*, 67(6), 2341-2373.

Frazzini, A., and L. H. Pedersen, 2014, "Betting against beta," *Journal of Financial Economics*, 111(1), 1–25.

Fung, W., and Hsieh, D. A., 2001, "Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases," *Journal of Financial and Quantitative Analysis*, 35(3), 291–307.

Fung, W., Hsieh, D. A., Naik, N. Y., and Ramadorai, T., 2008, "Hedge Funds: Performance, Risk, and Capital Formation," *The Journal of Finance*, 63(4), 1777–1803.

Getmansky, M., Lo, and A. W., Makarov, I., 2004. "An econometric model of serial correlation and illiquidity in hedge fund returns," *Journal of Financial Economics*, 74(3), 529-609.

Geltner, D., 1991, "Smoothing in appraisal-based returns," *Journal of Real Estate Finance and Economics*, 4, 327–345.

Geltner, D., 1993, "Estimating Market Values from Appraised Values without Assuming an Efficient Market," *Journal of Real Estate Research* 8(3), 325–345.

Goetzmann, W., Gourier, Elise., and Phalippou, L., 2019, "How Alternative are Private Markets?," Working Paper.

Gresch, N., and von Wyss, R., 2011, "Private Equity Funds of Funds vs. Funds: A Performance Comparison," The Journal of Private Equity, 14(2), 43-58.

Jagannathan, R., A. Malakhov, and D. Novikov, 2010, "Do hot hands exist among hedge fund managers? An empirical evaluation," *Journal of Finance* 65(1), 217–255.

Jenkinson, T., Sousa, M., and Stucke, R., 2013, "How Fair are the Valuations of Private Equity Funds?

J.P. Morgan., 2009, "Hard-to-value assets in uncertain times: Fair value reporting best practices for limited partners."

Jurek, J., and Stafford, E., 2015, "The Cost of Capital for Alternative Investments," *The Journal of Finance* 70(5), 2185-2226.

Kang, B. U., F. In, G. Kim, and T. S. Kim, 2010, "A longer look at the asymmetric dependence between hedge funds and the equity market," *Journal of Financial and Quantitative Analysis* 45(3), 763–789.

Kat, H., and C. Brooks, 2002, "The statistical properties of hedge fund index returns and their implications for investors," *Journal of Alternative Investments*, 5, 26–44.

Khandani, A. E. and A. W. Lo, 2011, "Illiquidity Premia in Asset Returns: An Empirical Analysis of Hedge Funds, Mutual Funds, and US Equity Portfolios," *Quarterly Journal of Finance* 1(2), 205–264.

Kosowski, R., N. Y. Naik, and M. Teo, 2007, "Do hedge funds deliver alpha? A Bayesian and bootstrap analysis," *Journal of Financial Economics* 84(1), 229–264.

Lhabitant, F. and Learned, M, 2002, "Hedge Fund Diversification: How Much is Enough?," *FAME Research Working Paper* 52.

Li, H., Y. Xu, and X. Zhang, 2016, "Hedge Fund Performance Evaluation under the Stochastic Discount Factor Framework," *Journal of Financial and Quantitative Analysis* 51(1), 231–257.

Liao, T., 2005, "Clustering of Time Series Data - A Survey," *Pattern Recognition* 38, 1857–1874.

Lo, A. and C. MacKinlay, 1990, "An Econometric Analysis of Nonsynchronous Trading," *Journal of Econometrics* 45, 181–212.

MacQueen, J., 1967, "Some methods of classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, ed. by L. M. L. C. . J. Neyman, vol. 1, pp. 281–297. Berkeley, CA: University of California Press.

Mannix, R., 2019, "Machine learning study identifies eight risk factors in private equity," Risk.net.
https://www.risk.net/investing/7030386/machine-learning-study-identifies-eight-risk-factors-in-private-equity

Mathonet, P., and Monhanel. G, 2006, "Valuation guidelines for private equity and venture capital funds: A survey," *Journal of Alternative Investments* 59-70.

McKinsey, 2023, "McKinsey Global Private Markets Review 2023: Private markets turn down the volume".

Pagliari Jr, J. L., K. A. Scherer, and R. T. Monopoli, (2005), "Public versus private real estate equities: a more refined, long-term comparison," *Real Estate Economics*, 33(1), 147–187.

Pastor, L., and R. F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111, 642–685.

Patton, A. J., 2008, "Are market neutral hedge funds really market neutral?" *Review of Financial Studies* 22(7), 2495–2530.

Phalippou, L. and Gottschalg, O., 2005, "Performance of Private Equity Funds.

Phalippou, L., and Zollo, M., 2005, "What Drives Private Equity Performance?

Preqin, 2023, "Preqin Private Capital Performance Data Guide".

Rehring, C, 2012, "Real Estate in a Mixed-Asset Portfolio: The Role of the Investment Horizon," *Real Estate Economics* 40(1) pp. 65–95.

Ross, S., and R. Zisler, 1991, "Risk and return in real estate," *Journal of Real Estate Finance and Economics*, 4, 175–190.

Sebestyen, G. S., 1962, Decision making processes in pattern recognition. New York: Macmillan.

SEC, 2023, "SEC Adopts Amendments to Enhance Private Fund Reporting", Press Release May 3, 2023. Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, 58(1), 267–288.

Sharpe, W., 1992, "Asset allocation: Management style and performance measurement," *The Journal of Portfolio Management*, 18(2), 7–19.

Teo, M., 2009, "The geography of hedge funds", *Review of Financial Studies* 22(9), 3531–3561.

Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, 58(1), 267–288.

Titman, S. and C. Tiu, 2011,. "Do the Best Hedge Funds Hedge?" *Review of Financial Studies* 24(1), 123–168.

Weidig, T., and Mathonet, P., 2004, "The Risk Profile of Private Equity".

Zou, H., and T. Hastie, 2005, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society*, 67(2), 301–320.

D Assumptions and properties of model

D.1 Assumptions

The Assumptions follow as in Bonhomme and Manresa (2015) (a-g) and in Goetzmann, Gourier, and Phalippou (2018) (h-i). Assumption g ensures consistency and introduces the grouping structure. Assumptions h and i introduce the properties in the covariate structure first introduced by the autocorrelation covariates. Collectively all assumptions contribute to a well-behaved and specified model.

a) Θ and A are compact subsets of $\mathbb{R}$. This assumes the compactness of the parameter spaces defining the parameter vectors containing the common parameters shared across all individual finds, and group-specific time effects, respectively. This ensures stationarity in the factor returns.

b) $E(|X_{it}|^2) \leq M$, where $|\cdot|$ denotes the Euclidean norm. This bound on the second moment on factor returns ensures bounded variability, ruling out non-stationary covariates.

c) $E[v_{it}] = 0$ and $E[v_{it}^4] \leq M$. The weak error dependence condition also ensures non-stationary covariates are excluded.

d) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E(v_{it}v_{js}X_{it}X_{js}) \leq M$. Bounds the average covariance between error terms and factor returns, inducing independence, and places restrictions on the interaction of errors, covariates, and group returns. Inducing independence

e) $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} E(v_{it}v_{jt}) \leq M$. Bounds and limits cross dependence between the error terms, minimizing dependence between error terms toward zero.

f) $\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}(v_{it}v_{jt}, v_{is}v_{js}) \leq M$. Bounds the cross-covariance, again restricting interaction between errors and returns.

g) For $d \in \{1, \ldots, D\}$, let $\bar{X}_{d,g_i,t}$ denote the mean of X_{it} over i, t such that $D_{it} = 1$, in the intersection of groups $g_{0i} = g$ and $g_i = \bar{g}$. Let $\hat{\rho}d$ be the minimum eigenvalue of the following matrix, where the infimum is taken over all possible groupings $\gamma = g_1, \ldots, g_N$: $\hat{\rho}d = \inf_{\gamma\in\Gamma_G}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}(X_{it}-\bar{X}_{d,g_i,t})(X_{it}-\bar{X}_{d,g_i,t})'$. Then $\mathrm{plim}_{N,T\to\infty}\hat{\rho}_d = \rho_d > 0$. The assumption introduces the concept of groups (or clusters) and ensures that as the sample size and time period increase, the limit of $\hat{\rho}_d$ exists and is positive when approaching infinity, denoted by $\rho_d > 0$. It however requires that the dummies times the lagged returns exhibit sufficient variation over time and across individuals to identify the components of .

h) There exists b such that $\forall i, t, \left|\sum_{l=1}^{L}\theta_{il,t}\right| \leq b < 1$. Bound absolute parameter values. Prohibits any individual parameter to become excessive. This condition is necessary to be able to recover residual errors from observed returns, conditionally on the parameters.

i) For all i, $\mathrm{rank}\left(\frac{1}{T}\sum_{t=1}^{T} X_{it}X_{it}'\right) = L$. The rank of the matrix is constant, thus the covariates are not linearly dependent and have sufficient variability. This ensures the avoidance of multicollinearity and requires that the lagged returns exhibit sufficient variation over time, for each fund, to identify the fund-specific autocorrelation component.

D.2 Theorem

The proof follows Bonhomme and Manresa (2015) and the adaption made by Goetzmann, Gourier, and Phalippou (2018). The most important part of the adaption is the demeaned observed returns are expected at 0, which is what the unsmoothing process tries to ensure. A simple explanation of the proof is provided for the explanation of the models validity and

consistency:

Proof 1. Under the assumptions, as T and N tend to infinity, $\hat{\delta} \xrightarrow{p} \delta_0$ (parameter vector convergence), for all i, $\hat{\theta i} \xrightarrow{p} \theta 0 i$ (individual-specific parameter convergence) and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\hat{\alpha}\hat{g}i,t - \alpha_0 g0i,t| \xrightarrow{p} 0$ (group-specific convergence). The parameters become more accurate with more data (consistency).

Proof 2. We first define functions $\hat{Q}(\theta,\eta,\alpha,\gamma)$ and $\tilde{Q}(\theta,\eta,\alpha,\gamma)$, and we observe that $\hat{Q}$ and $\tilde{Q}$ differ by a term that converges to zero, i.e. $\hat{Q}(\theta,\eta,\alpha,\gamma) - \tilde{Q}(\theta,\eta,\alpha,\gamma) \xrightarrow{p} 0$. Meaning that despite different specifications of the formulas, they are asymptotically equivalent and their difference converges towards zero.

Next, we apply the Cauchy-Schwarz inequality to bound the expectations, using assumptions to establish that certain quantities are bounded. We find that the first term in the expectation converges to zero as N and T go to infinity, i.e., $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}[D_{it}^{0}(\delta_0 - \delta)]X_{it}\frac{1}{1-\sum_{l=1}^{L}\theta_{il,t}} \xrightarrow{p} 0$.

Finally, applying the Cauchy-Schwarz inequality again, we find that the second term in the expectation also converges to zero as N and T go to infinity, i.e., $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}(\alpha_0 g_{it} - \alpha g_{it}) \xrightarrow{p} 0$.

E Grouped Patterns of Heterogeneity in Private Markets

Model (4) contains three types of parameters: the parameter vector $\theta \in \Theta$, which is common across individual funds; the group-specific time effects $\alpha_{g,t} \in \mathcal{A}$, for all $g \in \{1,...,G\}$ and all $t \in \{1,...,T\}$; and the cluster variables g_i for all $i \in \{1,...,N\}$, which map individual funds into groups. The parameter spaces Θ and $\mathcal{A}$ are subsets of $\mathbb{R}^{\mathbb{K}}$ and $\mathbb{R}$, respectively. We denote as α the set of all $\alpha_{g,t}$'s, and as γ the set of all g_i's. Thus, $\gamma \in \Gamma$ denotes a particular grouping (i.e., partition) of the N funds, where Γ_G is the set of all groupings of $[1,...,N]$ into at most G groups. The covariates vector $x_{i,t}$ may include strictly exogenous regressors and lagged outcomes. This component captures the autocorrelation pattern specific to each fund and is identified through the analysis of their individual time series of returns. The model allows for time-invariant regressions under certain support conditions. Moreover, $x_{i,t}$ and $\alpha_{g_i,t}$ are allowed to be arbitrarily correlated, where we introduce a set of constants that are common across funds and multiplied by d time- and fund-specific dummy variables. These constants are collectively represented by the matrix $\boldsymbol{\delta} \in \Theta^{d\times k}$ and are estimated from the cross-section of funds.

The full model we minimize wrt. simplifies to:

$$v_{i,t} = \frac{R_{i,t}^{o} - \left(D_{i,t}'\boldsymbol{\delta}x_{i,t}\right)}{1-\sum_{l=1}^{L}\theta_{i,l,t}} - \alpha_{gi,t} \tag{6}$$

After adjusting for the unsmoothing parameter in the denominator we get the grouped fixed-effects estimator model in (4) which is defined as the solution to the following minimization problem:

$$(\hat{\theta},\hat{\alpha},\hat{\gamma}) = \underset{(\theta,\alpha,\gamma)\in\Theta\times A^{GT}\times\Gamma_G}{\arg\min}\sum_{i=1}^{N}\sum_{t=1}^{T}(R_{i,t} - x_{i,t}'\theta - \alpha_{g_i,t})^2 \tag{7}$$

For given values of θ and α, the optimal group assignment for each individual fund is:

$$\hat{g_i}(\theta,\alpha) = \underset{g\in\{1,...,G\}}{\arg\min}\sum_{t=1}^{T}(R_{i,t} - x_{i,t}'\theta - \alpha_{g_i,t})^2 \tag{8}$$

The group fixed-effects estimator of $(\hat{\theta}, \hat{\alpha})$ can be written as:

$$(\hat{\theta}, \hat{\alpha}) = \underset{(\theta,\alpha)\in\Theta\times A^{GT}\times\Gamma_G}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(R_{i,t} - x'_{i,t}\theta - \alpha_{\hat{g}_i(\theta,\alpha),t})^2 \tag{9}$$

To find the optimal amount of clusters we optimize by minimizing the following Bayesian Information Criterium (BIC):

$$BIC(G) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t} - r_t - \hat{\theta}_t^{(G)}\left(R_{i,t-\Delta} - r_{t-\Delta}\right) - \hat{\alpha}_{i,t}^{(G)}\right)^2 + \hat{\sigma}^2\frac{(GT+N+K)}{NT}\ln(NT) \tag{10}$$

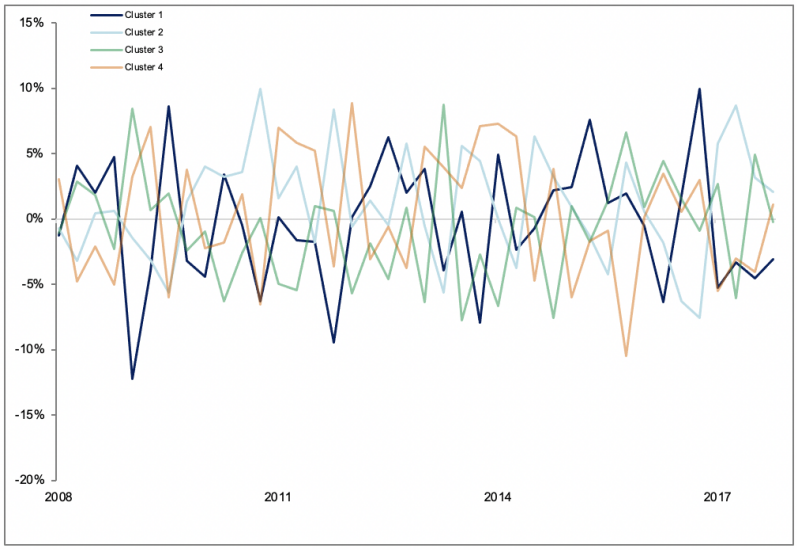
Where the $\hat{\sigma}^2$ is estimated by overfitting the model on a large number of groups.

E.1 Robustness Test for Algorithm

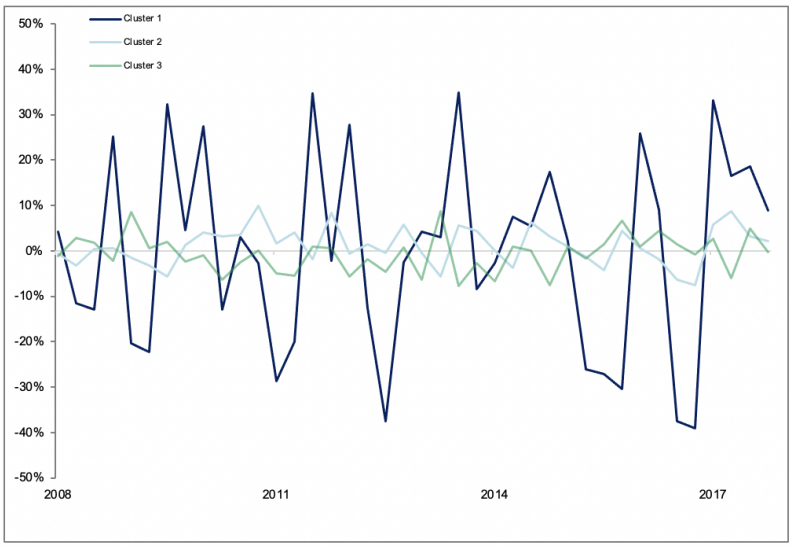
The model is evaluated using synthetic datasets derived from Monte Carlo simulations. Each dataset comprises three "true" groups, with 500 funds per group. Every fund contributes a 40-quarter time series, collectively resulting in 60,000 data points per dataset. Distinct assumptions characterize each group. In the first dataset, returns follow independent standard Brownian motions with quarterly volatility parameters of 10%, 20%, and 30%. The second dataset introduces a correlation structure where the second group correlates with the first and the third with the second. Because of the multivariate structure, the correlation structure is created by a Cholesky decomposed covariance matrix. In this instance, all groups share a quarterly volatility parameter of 0.3 in their stochastic return generation. For the third dataset, the volatility parameter in the single standard Brownian motion is set to 20%, followed by scaling of the groups by factors of 10%, 20%, and 30%. Residuals abide by a normal distribution around 0, with residual volatility set to 0.2. Aside from not introducing autocorrelation this mirrors the structures of synthetic data used in Goetzmann, Gourier, and Phalippou (2018). We intentionally set autocorrelation to zero to facilitate model training on an unsmoothed dataset. The synthetic data does exhibit traces of autocorrelation due to the probabilistic characteristics of the Brownian motion process employed in return generation. The model, free of dummies and autocorrelation, simplifies to a k-means clustering. The descriptive statistics of the simulated clusters are presented in Table 10, while the time series of the time-specific latent group returns $(\alpha_{g_i,t})$ are shown in Figure 4 It works well for finding groups in datasets 2 (Correlated Random Variables) and 1 (Independent Standard Brownian Motions). However, it struggles for dataset 3 (Single Standard Brownian Motion) as the SSE decrease due to a larger group rarely outperforming the BIC increase. We've also noticed a relationship between the magnitude of the Mean Squared Error (MSE) and the optimal groups, indicating that the model struggles for small sample sizes if the differences are small between the true groups. The optimal groupings of Datasets 1 and 3 score MSE values less than 0.05 and 0.002 respectively, whereas the optimal cluster for Dataset 2 frequently corresponds to an MSE around 0.65 level.

Figure 4: Time-series of Clusters

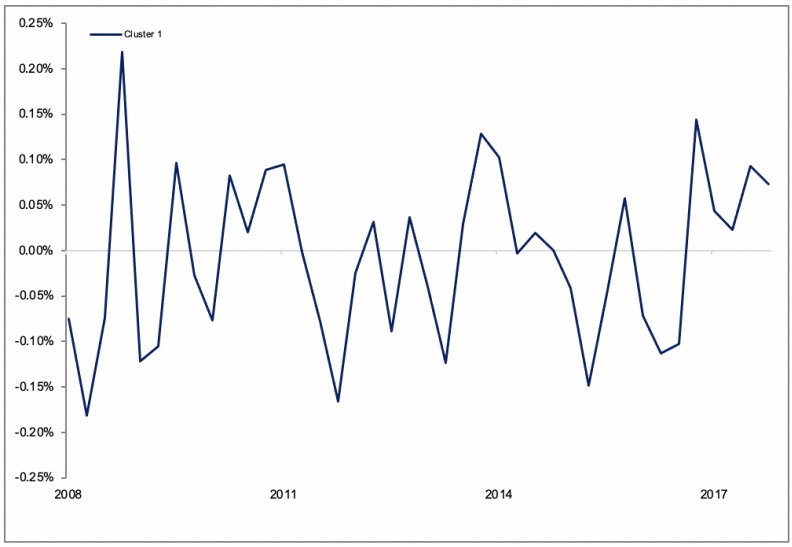
The plotted graphs below exhibit the four private factors. Each return from the respective simulated funds is characterized by the formula $R_{i,t} = x_{i,t}\theta + \alpha_{g_i,t} + v_{i,t}$. These simulated factor time series reflect the latent group returns, $\alpha_{g_i,t}$, associated with each cluster of funds g_i.



(a) Independent Standard Brownian Motions



(b) Correlated Random Variables



(c) Single Standard Brownian Motion

Table 10: Selection of the number of simulated clusters

BIC (%)	Case 1	Case 2	Case 3
1 cluster	6.310	93.806	**0.251**
2 clusters	6.451	96.880	0.256
3 clusters	6.554	**92.788**	0.260
4 clusters	**6.243**	100.042	0.264
5 clusters	6.832	102.296	0.266
MSE (%)			
1 cluster	4.629	69.473	0.186
2 clusters	4.554	68.228	0.184
3 clusters	4.486	67.190	0.185
4 clusters	4.431	66.493	0.184
5 clusters	4.440	65.749	0.183

F Sample Selection and Data Description

The initial dataset includes 4676 private investment vehicles. We apply some standard screens before including observations in the sample. We start by excluding funds with a fund size of fewer than 10 million USD (28 funds), and funds with fixed-income or fund-of-funds profiles (1049 funds) discussed in Section 4.2. Following this initial filtration, 3599 funds persist in our database. Subsequently, we remove NAVs exceeding 15 years of the fund lifespan after the vintage year (6,448 cash flows). We winsorize quarterly returns outside the 1% and 99% percentiles (11,570 cash flows) and retain only funds with a minimum of 12 consecutive quarterly returns. The filtered dataset results in a total omission of 1,738 funds and 54,623 cash flows. Table 11 shows the effects of each filtering process on the remaining quantity of funds, returns, and cash flows in the database.

Table 11: Sample Selection and Data Preprocessing

	Funds	Returns	# Cash Flows
Initial database	4676		159,323
Filters on fund characteristics			
Fund size (USD) > 10m	4648		158,673
Exclusion of strategies	3599		125,389
Total returns		125,389	
Filters on NAV and returns			
Return < 15y Vintage	3529	118,941	
Winsorizing	3465	107,371	
Filtered dataset	**2938**	**104,700**	

The table provides an overview of the effects that various filters have on the data, specifically in terms of the quantity of funds examined, returns, and cash flows. With respect to NAV, any NAV 15 years after vintage year. The constraints applied to returns fall at the 1% and 99% percentiles of the return distribution, which correspond to -0.41 and 0.63, respectively. A fund is only included if it has a minimum of 12 consecutive quarterly returns after applying our screening.

Interest Rate Sensitivity of Risk Measures in European and Exotic Options: A Study of Delta, Vega, and Breach Probability

Abstract

The recent changes in interest rates have greatly influenced how we approach investment strategies and evaluate financial risks. This study aims to examine the impact of interest rate changes on the risk analysis of European and Exotic options. Our goal is to assess how fluctuations in interest rates affect the Delta and Vega of European options, and then investigate their impact on the probability of breaching a barrier. First, we derive the sensitivity of Delta and Vega to the interest rate and analyze their behavior. Secondly, we compute the probability that a stock crosses a barrier B over $t \in [0, T]$, which can be seen as the price of a one-touch option, before deriving its sensitivity to the interest rates. The computation of this probability is done by finding the Cumulative Distribution Function of the maximum of a Geometric Brownian Motion and validated by a Monte Carlo method. At the same time, the sensitivity to the interest rates is derived using a forward differentiation. This research also includes a pricing application for one-touch options and European binary options.

In the current financial landscape, neglecting the influence of interest rate changes on conventional risk assessment could render option risk management ineffective. Our findings spotlight the significant impact of changes in interest rates on Delta and Vega, and emphasize the pronounced effect on the probability of barrier breaches, especially when the option is at the money. These results attest to the necessity of incorporating more interest rate sensitivity into options risk management strategies to yield a more accurate and robust risk evaluation process.

Keywords: Barrier Breaching, Interest rates sensitivity, reflection principle, Monte Carlo Simulation, pricing

1 Introduction

The recent surge in interest rates raised challenges not observed since the onset of the Great Financial Crisis. In the last 15 years, interest rates were close to zero, and traders and financial institutions did not take them into account for short to mid-term horizons. However, since the COVID crisis and the beginning of the war in Europe, interest rates raised which led financial institutions to face forgotten risks. The emerging risks already impact the Derivatives market and several risk measures should be updated. A methodology for option pricing for deriving the ratio for stock to hold when going short in a European call option called Delta was already introduced [1]. In this paper, the authors do include a rate term but assume it is constant, while they do not include the discounted factor in the delta calculation. As for the sensitivity of delta to the interest rate, literature on the Vera term, also expressed as the sensitivity of the vega to the interest rate, is not common and should become more relevant in the coming years.

This new economic environment has an impact on the diffusion process used to represent the path of an underlying. In the low-interest rate period, assuming a non-existent or very low drift seemed reasonable, which cannot be the case nowadays.
This paper tries to extend [7] and [8] that develop a methodology to compute the probability that a stock following a geometric Brownian motion breaches a certain level over $t \in [0, T]$. It also draws upon [9], which introduces a ratio between the probability to hit a barrier and the probability of still being below this barrier at maturity, by deriving the sensitivity to the interest rate that a stock following a drifted geometric Brownian motion breaches a Barrier B over $t \in [0, T]$.

The objective of this work is to develop a second-order Greek by taking the derivatives of the Delta by the interest rate and formalizing Vera, which is the sensitivity of the Vega to the interest rate. The second part of this research starts by explaining the methodology to compute the probability that a stock, following a geometric Brownian motion, breaches a barrier over $t \in [0, T]$ [7]. Since this probability can be thought of as the price of a One-Touch option [1][2] and therefore twice the price of a European Binary option [[2], [4]], a pricing application is included. Directly stemming from this methodology, we develop a new Greek which is the sensitivity of this probability (price) to a change in ten basis points in the interest rate. The last part is a comparative analysis of this probability against the one obtained through Monte Carlo Simulation.

[1] Price of a one-touch option before discounting
[2] In this research, we only consider One Touch options which pay at maturity

2 Formalization of the Impact of Interest Rates Movements on Delta and Vega

Prior to delving into the derivation of the Greeks, it is essential to review the underlying model [1]. In this paper, the authors introduce a closed-form solution to compute the prices of European-style options, based on the following assumptions:

1. **Constant Volatility**: The volatility of the underlying asset is constant over time.
2. **Continuous Trading and Log-Normal Distribution**: The underlying asset can be traded continuously, and its price S_t is assumed log-normally distributed, which implies the log returns of S being normally distributed.
3. **Short Selling**: One can always short-sell the underlying stock.
4. **No Transaction Costs or Taxes**: No transaction costs or taxes are incurred when trading the underlying asset or option.
5. **Perfectly Divisible Securities**: All securities are perfectly divisible.
6. **Risk-Free Borrowing and Lending**: Possibility of always borrowing and lending cash at the risk-free interest rate r, which is assumed to be constant.
7. **Constant Dividend Yield**: The stock pays a constant dividend yield q.

The authors are able to find a closed-form formula to price a call and a put option:

$$C = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2) \tag{1}$$

$$P = K e^{-rT} [N(d_2) - 1] - S_0 e^{-qT} [N(d_1) - 1] \tag{2}$$

where:

$C(P)$ is the price of the call (put) option,
S_0 is the current price of the underlying asset,
K is the strike price of the option,
r is the risk-free interest rate,
q is the dividend rate,
T is the time to expiration of the option,
N is the cumulative standard normal distribution function,

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}},$$

$$d_2 = d_1 - \sigma\sqrt{T}.$$

However, trading derivatives can be a source of risks that come from various factors. Thus, one needs to be aware of the sensitivity of the derivative's price to the various parameters that impact its value. From the closed-form formula of the derivatives computed equations (1) and (2), deriving the sensitivity with respect to the Spot, Implied Volatility, and Interest rate is trivial. The results are derived in [2]. Note that it is also common to compute the second derivative of the price with respect to the Spot, hence the Gamma (γ):

Table 1: Option Greeks: Common

	Call Option	**Put Option**
Delta	$e^{-qT} N(d_1)$	$e^{-qT}(N(d_1) - 1)$
Gamma	$\frac{e^{-qT}}{S_0\sigma\sqrt{T}} N'(d_1)$	
Vega	$S_0 e^{-qT} N'(d_1)\sqrt{T}$	
Theta	$\frac{-S_0 e^{-qT}\sigma}{2\sqrt{T}} N'(d_1) - rKe^{-rT} N(d_2) + qS_0 e^{-qT} N(d_1)$	$\frac{-S_0 e^{-qT}\sigma}{2\sqrt{T}} N'(d_1) + rKe^{-rT} N(-d_2) - qS_0 e^{-qT} N(-d_1)$

One can, as well, state that these risks are first-order risks (only gamma is a second-order risk). This means that these underlying risks are also sensitive to various parameters. In this study, we consider only the second-order risks arising from the interest rate on the Delta and Vega.

Our primary aim is to examine the impact of fluctuations in interest rates on the Delta and Vega, of both the Call and Put options. To achieve this, we compute the derivatives of Delta and Vega with respect to the interest rate. Detailed derivations of these calculations are provided in the Appendix A.1 and A.2.

Table 2: Options Greeks: Dera and Vera

	Computation	**Call Option**	**Put Option**
Dera	$\frac{\partial \delta}{\partial r}$	$N'(d_1) \cdot e^{-qT} \cdot \frac{\sqrt{T}}{\sigma}$	
Vera	$\frac{\partial V}{\partial r}$	$-d_1 \cdot N'(d_1) \frac{T}{\sigma} \cdot S_0 \cdot e^{-qT}$	

One can observe that both Dera and Vera exhibit identical behavior whether it is a call or a put option. Both products have a positive Vega that has to be the same due to the Black-Scholes assumption of constant volatility. Hence for any strike, time to maturity, or spot, Vega must be the same, and so the Vera. As for the Vega, the Gamma of a call and a put options are equal, and so is the derivation by the interest rate. The graphical representation, in Appendix B, further illustrates that Dera exerts a positive influence on the delta of our options, especially close to the strike price.

This effect can be intuitively explained by considering the impact of an increase in the interest rate, which subsequently raises the forward value. As a result, the probability of the option (δ) being in the money is heightened. While deep out-of-the-money options remain relatively unaffected by changes in the forward value, those closer to the strike price, show more significant changes in their probabilities of being in the money (δ). This is due to the fact that a higher forward value noticeably increases the likelihood of these options being in the money (δ). Graph illustrating Dera is provided in Appendix B.

Vera can be interpreted as the measure of how changes in the interest rate influence the sensitivity of the option price to implied volatility. We know that for deep out-the-money options, volatility has a small impact on price, which results in an insignificant effect if the interest rate is heightened by 10 bp. However, when the spot (more precisely the forward) approaches the strike, the interest rate has a positive impact on Vera. It becomes larger for out-the-money options as the forward moves closer to the strike. However, when the forward approaches the decreasing part of the Vera bell-shaped, we observe Vera becoming negative. As the Spot increases, for in-the-money calls, the less the option is sensitive to a change in volatility and thus, the lower interest rate has an effect on Vega. Graph illustrating Vera is provided in Figure C.

3 Construction of a New Greek: Probability of Barrier Breach and Comparison to Monte Carlo Simulation

Through our research, we aim to provide insights into how interest rate fluctuations impact the likelihood of an asset crossing a barrier over $t \in [0, T]$, offering valuable information for exotic traders.
First of all, we establish a diffusion process as a model for the stock price. Subsequently, we provide a concise introduction to the theory of stopping time and its application in this study. We then proceed to compute the probability of the asset crossing a predetermined barrier and analyze its sensitivity to changes in the interest rate. Finally, our results are compared with those obtained through Monte Carlo simulation.

In contrast to [3], which uses Monte Carlo simulation to evaluate options, this technique is used to estimate the theoretical percentage of paths that cross a specified Barrier within a given period T. This methodology is particularly valuable for traders dealing with exotic options such as autocall, barrier, and binary options.

3.1 Diffusion Process

In this part, we assume that the stock price follows a Geometric Brownian Motion, as in [1]. Geometric Brownian Motion can be represented by the stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{3}$$

where:

S_t is the stock price at time t,
μ is the drift,
σ is the volatility, the standard deviation of the stock's return,
dW_t is the Brownian motion.

However, the volatility is assumed to remain constant across the whole section.

3.2 Reflection principle to find the PDF of the maximum of a Brownian Motion

The cumulative distribution of the maximum of a drifted Brownian Motion on a given period can be found in closed form using the reflection principle. This Cumulative Distribution Function will be used to compute the sensitivity of stock prices, following a Geometric Brownian Motion, to the interest rate. Note that the next two sub-parts aim to clarify the maths, which can be found in [8] and [7].

3.2.1 Maximum of a Brownian Motion using the Reflection Principle

The maximum of a Brownian Motion can be represented as:

$$X_t := \max_{t\in[0,T]} W_t, \quad t \geq 0 \tag{4}$$

with:

$$W_0 = 0$$

which implies:

$$X_t := \max_{t\in[0,T]} W_t \geq 0$$

We aim to compute the Cumulative Probability Distribution of this maximum, using the Reflection Principle.

Let's define the first passage time as the earliest time at which W_t takes a particular level m. This first passage time to level m can be written as:

$$\tau_m = \min\{t \geq 0 \mid W_t = m\} \tag{5}$$

with:

$$m \geq W_0 := 0$$

Intuitively, we can deduce that the event X_0^T being at least equal to m is the same as the first passage time being before T:

$$\{X_0^T \geq m\} = \{\tau_m \leq T\}. \tag{6}$$

Thus, the probability that W_t is larger or equal to m during a predefined period is the same as the addition between the probability that W_t is higher than m given a predefined period and the probability that W_t is smaller or equal to m given the same period. Considering that the first passage of time happens before T, any path is thus a mirror of another one, which implies that the probability of each event is $\frac{1}{2}$:

$$P(W_T \geq m|\tau_m \leq T) = P(W_T > m|\tau_m \leq T) = P(W_T \leq m|\tau_m \leq T) = \frac{1}{2}. \tag{7}$$

Then, [7] shows that X_0^T has the same distribution as the absolute value $|W_T|$ of W_T:

$$P(\tau_m \leq T) = P(|W_T| \geq m). \tag{8}$$

Therefore, the Cumulative Distribution Function of X_0^T, assuming $a \geq 0$, $P(X_0^T \leq m)$ is:

$$P(X_0^T < m) = P(\tau_a > T) = P(|W_T| < a) = \frac{1}{\sqrt{2\pi T}} \int_0^a e^{-\frac{x^2}{2T}}\, dx \tag{9}$$

The proof can be found in [7], chapter 10, page 283.

Note, in this part we try to define the PDF of a Brownian motion.

$$M_0^T = \max_{t\in[0,T]} S_t \tag{10}$$

when S_t can be modeled as :

$$(S_t)_{t\in[0,T]} = S_0 e^{\sigma W_t} \tag{11}$$

which is found by [7], page 385, to be:

$$\varphi_{M_0^T}(y) = 1_{[S_0,\infty)}(y) \frac{1}{\sigma y} \sqrt{\frac{2}{\pi T}} \exp\left(-\frac{1}{2\sigma^2 T}(\log(y/S_0))^2\right), \quad y > 0, \tag{12}$$

3.3 From Brownian Motion to Drifted Brownian Motion

This subsection facilitates the transition from the Probability Density Function (PDF) of a stock price following a non-drifted Geometric Brownian Motion to a drifted Geometric Brownian Motion.

However, finding the PDF of a drifted Geometric Brownian Motion differs from the previous approach, as the property of symmetry in the non-drifted case no longer holds.

To begin, we define the maximum of a drifted Brownian Motion as:

$$\widehat{X_0^T} = \max_{t\in[0,T]} W_t = \max_{t\in[0,T]} (W_t + \mu t) \tag{13}$$

The second step is to compute the joint PDF of a Brownian motion $\widehat{W_T}$ and its maximum $\widehat{X_0^T}$, by using the Girsanov theorem. The computation and proof can be found in [7](Proposition 10.3), and in [8] (Theorem 7.2.1), and can be expressed as:

$$\varphi_{\widehat{M}(T),\widehat{W}(T)}(a,b) = 1_{a\geq \max(b,0)} \frac{2(2a-b)}{T\sqrt{2\pi T}} e^{\mu b - \mu^2 T/2 - (2a-b)^2/(2T)}, \quad a > Max(b,0) \tag{14}$$

By integrating the equation above, over the region (see Appendix D), we obtain the closed form of the Cumulative Distribution Function of the maximum of a drifted Brownian Motion over $t \in [0,T]$:

$$P(\widehat{X_0^T} \leq a) = N\left(\frac{a-\mu T}{\sqrt{T}}\right) - e^{2\mu a} N\left(\frac{-a-\mu T}{\sqrt{T}}\right), \quad a \geq 0 \tag{15}$$

Now that we have the Cumulative Distribution Function of the maximum of a drifted Brownian Motion, one should remember that the final objective is to find the Cumulative Distribution Function of the maximum over $t \in [0,T]$ for a stock following a Geometric Brownian Motion:

$$M_0^T = \max_{t\in[0,T]} S_t = S_0 \max_{t\in[0,T]} e^{\sigma W_t + (r - \frac{\sigma^2}{2})t} \tag{16}$$

Corollary 10.5 in [7] shows that this Cumulative Distribution Function is given by:

$$P(M_0^T \leq B) = N\left(-\frac{(r-\frac{\sigma^2}{2})T + \log\left(\frac{B}{S_0}\right)}{\sigma\sqrt{T}}\right) - \left(\frac{S_0}{B}\right)^{\frac{1-2r}{\sigma^2}} N\left(-\frac{(r-\frac{\sigma^2}{2})T - \log\left(\frac{B}{S_0}\right)}{\sigma\sqrt{T}}\right), \quad B \geq S_0 \tag{17}$$

Then, we can deduce:

$$P(M_0^T \geq B) = 1 - P(M_0^T \leq B - \epsilon) \tag{18}$$

Which implies,

$$P(M_0^T \geq B) = 1 - \left[N\left(-\frac{(r-\frac{\sigma^2}{2})T + \log\left(\frac{B-\epsilon}{S_0}\right)}{\sigma\sqrt{T}}\right) - \left(\frac{S_0}{B-\epsilon}\right)^{\frac{1-2r}{\sigma^2}} N\left(-\frac{(r-\frac{\sigma^2}{2})T - \log\left(\frac{B-\epsilon}{S_0}\right)}{\sigma\sqrt{T}}\right)\right], \quad B-\epsilon \geq S_0 \tag{19}$$

This Cumulative Distribution Function can then be used to compute the probability that an asset, following a drifted Geometric Brownian Motion, crosses a certain level, over a defined period. It implies that this Cumulative Distribution Function can also be used to compute the sensitivity of that probability to the drift term, and hence to the interest rate.

Note that this Cumulative Distribution Function may also be a methodology to price one-touch options, which means that we can use it to compute the price of a digital European option [[2], [4]]. As a path touches the barrier level B, we can admit that a reflected path that mirrors the one touching the barrier exists. Then the probability that the path ends above or below the level B is $\frac{1}{2}$, while for a one-touch option since the path already touches B, the probability of the path reaching B is 1. Then, we can deduce the price of an American Digital option (one touch) as twice the price of a digital European Option. See an example in Appendix [E, F].

3.4 Sensitivity of the Probability to Breach a Barrier over $t \in [0, T]$ to Interest Rate

From the Cumulative Distribution Function that an asset, following a Geometric Brownian Motion, crosses a Barrier B over a defined period, we can compute its sensitivity to the interest rate. Using the forward differentiation, we obtain:

$$f'(S_0, B, r, \sigma, T) \approx \frac{f(S_0, B, r+h, \sigma, T) - f(S_0, B, r, \sigma, T)}{h}, \tag{20}$$

where h is a small finite difference step (10bp) and $f(S_0, B, r, \sigma, T)$ the Cumulative Distribution Function computed previously. To simplify the following, we name the sensitivity of this probability ζ. To compute the sensitivity, we assumed $B = 120\%$, $r = 2\%$.
Note that this task is the equivalent of computing the sensitivity of the price of a one-touch option (without the discount factor) to the interest rate.

Figure G shows that ζ is a bell-shaped curve that is null when Spot is too far from the Barrier and null when at the Barrier and beyond. We notice that increasing the time period increases the amplitude of the bell-shaped, which means that a move in interest rate on a larger period has a larger impact on the probability of breaching the barrier. On the other hand, larger volatility makes the bell-shaped larger when far from the Barrier but smaller as the spot approaches the Barrier. Meaning that for a more volatile process, interest rates affect less (more) the probability of crossing when the spot is close (far) from the barrier. Intuitively, a volatile stock close to the barrier depends less on interest rate variation to reach the barrier than a less volatile stock.
Illustrating the parameter ζ through a practical example, we evaluate a one-touch option on a stock governed by a geometric Brownian motion. With 20% volatility, maturity $T = 1$, and initial value $S_0 = 100\%$, ζ is determined as 1.65%. This implies that a 10 basis points increase in the interest rate raises the probability of barrier breach by 1.65%. Considering the option's payoff as this probability times a discount factor, the resulting change in price amounts to 1.67% [3]

$$Price_{OneTouch} = P_{breach} * exp(-rT) \tag{21}$$

3.5 Comparison with the American Monte Carlo Simulation

This part is dedicated to the explanation of the American Monte Carlo Simulation used to find the theoretical ζ and to compare the results with the method we develop through this paper.

In this part, we create n number of paths, starting from a point S_0 and following a Geometric Brownian Motion. Then, we choose a Barrier B and compute the Percentage of Monte Carlo Paths crossing a Barrier B (Algorithm 1).

Algorithm 1 Percentage of Monte Carlo Paths that crossed a Barrier

Require: $S0$: Initial stock price
Require: r: Risk-free rate
Require: σ: Volatility of the stock
Require: T: Time to maturity (in years)
Require: M: Number of time steps
Require: I: Number of simulation paths
Require: B: Barrier Level
$dt \leftarrow \frac{T}{M}$
Define starting value of each path as S_0
for $sim \leftarrow 0$ **to** $I - 1$ **do**
  Compute $x :=$ Generate $M - 1$ random variable under the standard normal distribution
  for $t \leftarrow 1$ **to** M **do**
    $paths[t, sim] \leftarrow paths[t-1] \cdot \exp\left((r - \frac{1}{2}\sigma^2) \cdot dt + \sigma \cdot \sqrt{dt} \cdot x[t-1]\right)$
  end for
end for
Compute P:= the percentage of paths that crossed the level B
return P

The algorithm allows us to find the theoretical probability of a stock crossing a barrier over $t \in [0, T]$ for a spot price. The next step is then to use equation (20) by defining $f(x)$ as in Algorithm 1. Figure in Appendix H shows the theoretical probability of a stock crossing a barrier over $t \in [0, T]$ computed for a range of spot prices. Note that the Monte Carlo Simulation uses 1 000 000 simulation paths to compute the probability, with $B = 120\%$ and $r = 2\%$:

Figure in Appendix H exhibits the results of the comparison with the Monte Carlo method. The probabilistic method seems to be consistent with the Monte Carlo approach. To check the consistency of the estimations, we perform a Kolmogorov-Smirnov two-sided test ([5]) and conclude that the underlying distributions are identical since we cannot reject the null hypothesis (p-value is 0.99).

[3] The change from 1.65% to 1.67%, while pricing, arises from the discount factor in the pricing formula of the one-touch, that have an impact on its derivatives by the interest rate.

4 Conclusion

In this paper, we first focus on enhancing the risk analysis for European options by considering interest rate movements. More specifically, we compute the sensitivity of both the Vega and Delta to changes in the interest rate. Our findings reveal various intriguing patterns in their behavior.

Regarding Delta's sensitivity to interest rates, a minimal impact is observed when the Spot is distant from the Strike. However, when the Spot range is close to the Strike price, the interest rate significantly influences the probability of the option being "in the money."

On the other hand, Vega exhibited a different response to interest rates. When the option is "out of the money", interest rates positively impacted Vega. Conversely, for "in the money" options, interest rates exerted a negative influence on Vega.

In addition to the sensitivity analysis, a novel methodology is introduced to estimate the sensitivity of the probability that a stock, following a drifted Geometric Brownian Motion, breaches a Barrier within $t \in [0, T]$. This approach is based on [7] and [8] and explains the computation of the Cumulative Distribution Function of the maximum of a Geometric Brownian Motion that we use to derive the probabilities while employing finite differentiation to estimate the sensitivities related to the interest rate. Note that the computation of the Cumulative Distribution Function can also be useful in the pricing of digit options (American and European) as shown in Appendix F.

Examining the results shown in Figure G, we observe that as the length of the time period increased, the amplitude of the sensitivity also increased. Furthermore, heightened volatility resulted in a diminished probability for spots near the barrier and an augmented probability for spots far from the barrier.

Overall, this research contributes to a more comprehensive understanding of interest rate effects on European options and presents a methodology for estimating the sensitivity of the probability of breaching a Barrier within a specified time frame to the interest rate. These insights hold, we hope, valuable implications for risk management strategies, and for exotic options pricing, such as American and European Digital options. Limits of this method come from the diffusion process, assumed as a drifted Geometric Brownian Motion, with constant volatility. An interesting extension could incorporate a stochastic volatility model to diffuse the prices.

An application has been made on one-touch options, but an extension of this method analyzing the impact of interest rate on the price of a double barrier option would also be interesting. Lastly, the pricing of the one-touch and the computation of its sensitivity has been done by stating that the option pays at maturity while commonly, the one-touch option pays 2 days after the day of touch. It implies that computing its price requires computing the expected value of τ_m, defined in (5).

References

[1] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973. doi: 10.1086/260062.

[2] Mohamed Bouzoubaa and Adel Osseiran. *Exotic Options and Hybrids: A Guide to Structuring, Pricing, and Trading*. Wiley, Hoboken, New Jersey, 2006. ISBN 978-0470011621.

[3] Phelim P. Boyle. Options: A monte carlo approach. *Journal of Financial Economics*, 4(3):323–338, 1977.

[4] J. Gatheral. *The Volatility Surface: A Practitioner's Guide*. Wiley finance series. John Wiley & Sons, 2006. ISBN 9781119202073.

[5] Jr. Massey, Frank J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[6] Robert C. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1977. doi: 10.2307/3003143.

[7] Nicolas Privault. *Introduction to Stochastic Finance with Market Examples*. CRC Press, 2022.

[8] Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2004.

[9] Lujing Su and Marc O Rieger. How likely is it to hit a barrier? theoretical and empirical estimates. *Technical report, Working Paper No. 594, National Centre of Competence in Research, Financial Valuation and Risk Management*, 2009.

A Derivation of Second-order Greeks for European-style option

A.1 Derivation of Dera for Call European-style option

Starting with the Δ of a European Call option.

$$\Delta_c = e^{-qT} N(d1) \tag{22}$$

Deriving by r the call option, to observe its sensibility to the rate.

$$\frac{\partial \Delta_c}{\partial r} = e^{-qT} \cdot \frac{\partial N(d_1)}{\partial r} \tag{23}$$

$$= e^{-qT} N'(d_1) \cdot \frac{\partial (d_1)}{\partial r} \tag{24}$$

We can derive $\frac{\partial (d_1)}{\partial r}$:

$$\frac{d(\partial d_1)}{\partial r} = \frac{\partial \left[\frac{\ln\left[\frac{S}{K}\right] + \left(r + q + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}} \right]}{\partial r} \equiv \frac{\sqrt{T}}{\sigma} \tag{25}$$

Then, by replacement:

$$\frac{\partial \Delta_c}{\partial r} = e^{-qT} N'(d_1) \cdot \frac{\sqrt{T}}{\sigma} \tag{26}$$

To check the results, we can also derive the formula by deriving ρ by S:

$$\theta_c = K e^{-rT} + N\left(d_1 - \sigma\sqrt{T}\right)$$

$$\frac{\partial \theta_c}{\partial S} = K e^{-rT} T N'\left(d_1 - \sigma\sqrt{T}\right) \frac{\partial (d_1)}{\partial S} \tag{27}$$

We reduce $N'\left(d_1 - \sigma\sqrt{T}\right)$:

$$N'\left(d_1 - \sigma\sqrt{T}\right) = \frac{1}{\sqrt{2\pi}} e^{-\left(d_1 - \sigma\sqrt{T}\right)^2/2} = N'(d_1)\, e^{-\frac{\sigma^2 T}{2} + d_1 \sigma\sqrt{T}} = N'(d_1) \frac{S}{k} e^{(r-q)T} \tag{28}$$

Then, we replace $N'\left(d_1 - \sigma\sqrt{T}\right)$:

$$= K e^{-rT} T N'(d_1) \frac{S}{K} e^{(r-q)T} \cdot \frac{1}{S\sigma\sqrt{T}} \tag{29}$$

And obtain the same result as found earlier:

$$\frac{\partial \theta_c}{\partial S} = e^{-qT} N'(d_1) \frac{\sqrt{T}}{\sigma} \tag{30}$$

A.2 Derivation of Vera for Call European-style option

Starting with the Vega closed form:

$$\frac{\partial C}{\partial v} = Se^{-qT} N'(d_1)\sqrt{T} \tag{31}$$

$$\frac{\partial Vega}{\partial r} = Se^{-qT}\sqrt{T}\frac{\partial (N'(d_1))}{\partial r} \tag{32}$$

$$\frac{d(N'(d_1))}{\partial r} = N''(d_1)\cdot\frac{\partial (d_1)}{\partial r} = N''(d_1)\cdot\frac{\sqrt{T}}{\sigma} \tag{33}$$

$$Vera = -d_1 N'(d_1)\cdot\frac{T}{\sigma}Se^{-qT} \tag{34}$$

B Dera of an ATM call (put)

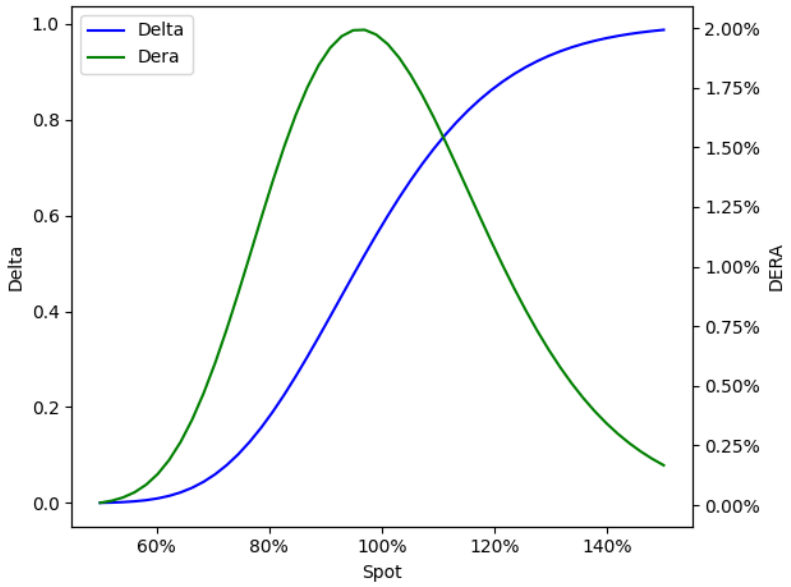


Figure 1: Effect of Strike price on the Dera of a European-style option. Here volatility of the underlying is set to 20%, dividend to 0%, time to maturity to 1 year and interest rate to 2%.

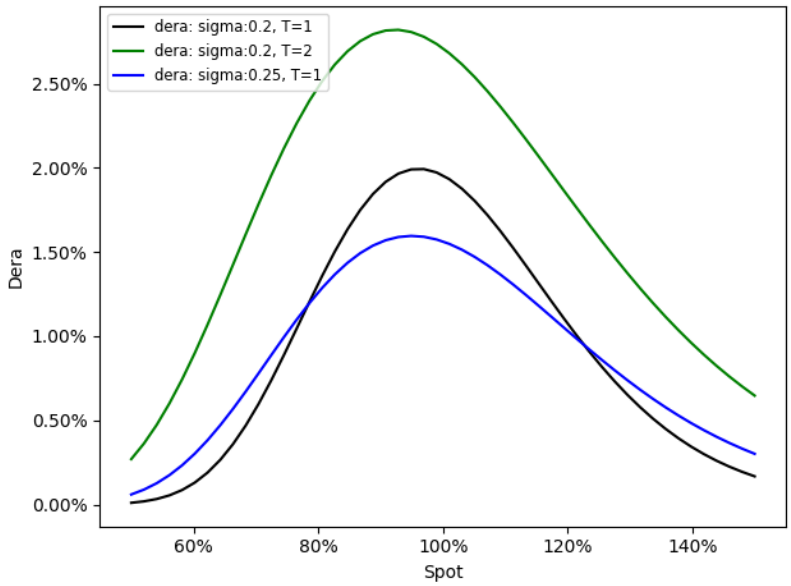


Figure 2: Effect of volatility and time to maturity on the Dera of an ATM European-style option. Here interest rate of the underlying is set to 2%, and the dividend to 0%. Higher volatility smooths Dera while higher time to maturity heightens it.

C Vera of an ATM call (put)

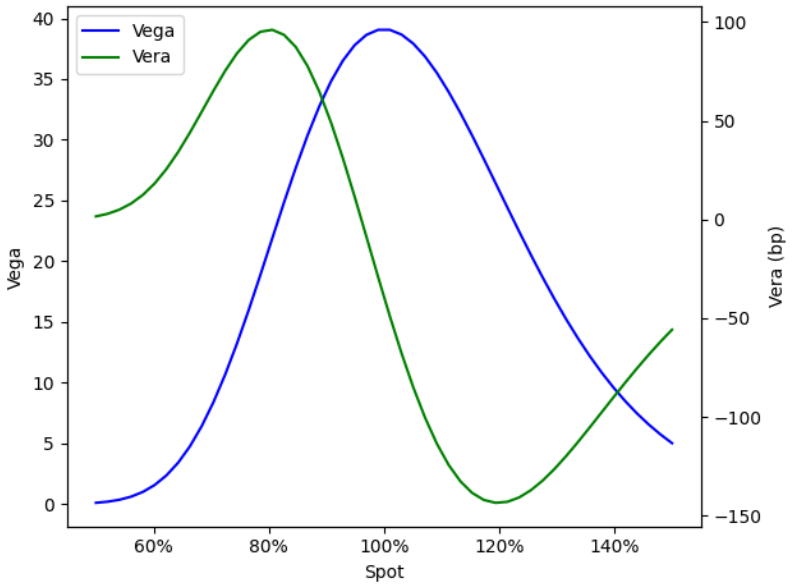


Figure 3: Effect of Strike price on the Vera of a European-style option. Here volatility of the underlying is set to 20%, dividend to 0%, time to maturity to 1 year and interest rate to 2%.

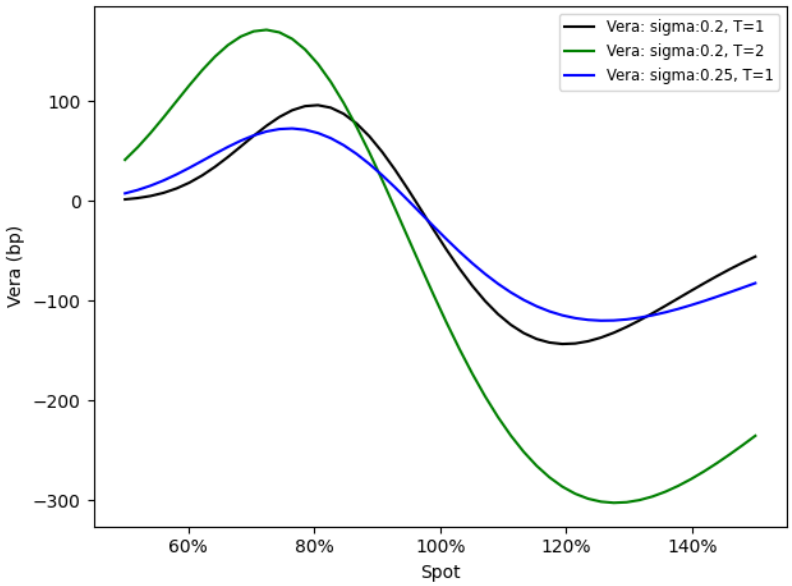


Figure 4: Effect of volatility and time to maturity on the Vera of an ATM European-style option. Here interest rate of the underlying is set to 2%, and the dividend to 0%. Higher volatility smooths Vera while higher time to maturity has the opposite effect.

D Range of $\widehat{M}(T)$,$\widehat{W}(T)$

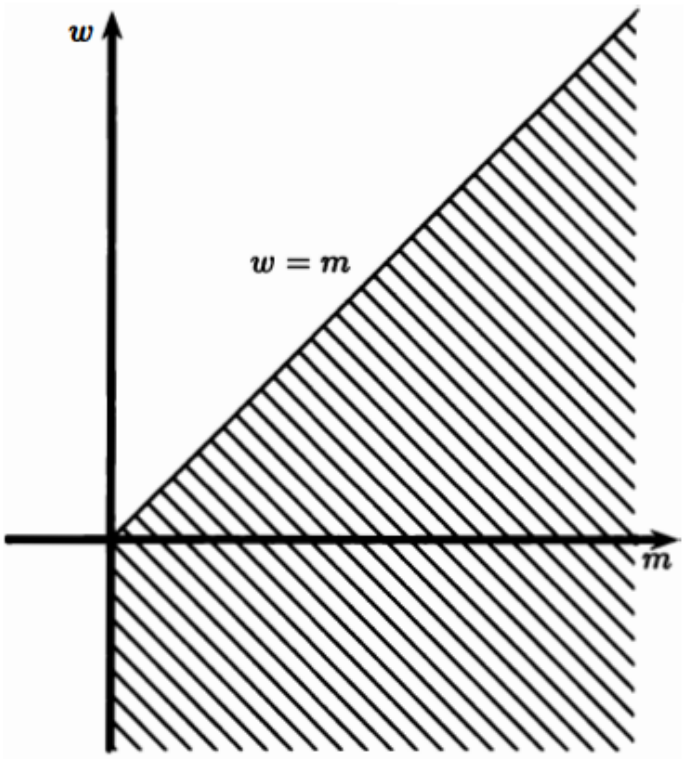


Figure 5: Since $W(0) = 0$, it follows that $\widehat{M}(T) \geq 0$. Additionally, $\widehat{W}(T) \leq \widehat{M}(T)$. As a result, the pair of random variables $(\widehat{M}(t), \widehat{W}(T))$ only takes values in the set $\{(m, w) \mid w \leq m, m \geq 0\}$.

E European to One Touch Digital Price Ratio

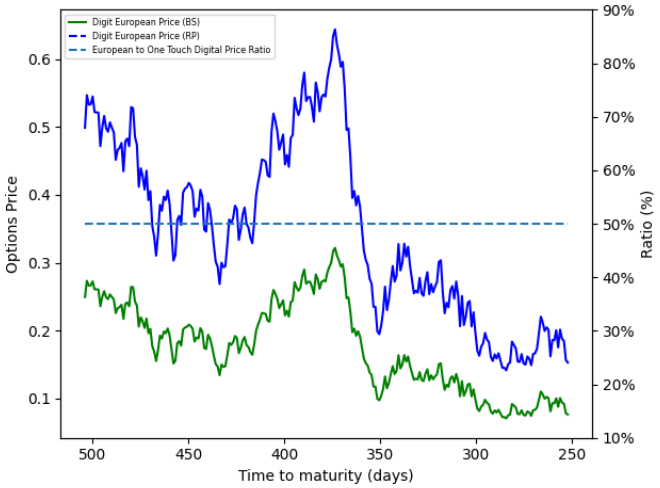


Figure 6: Price of an American digital (One Touch option) and of a European Digital. Here the underlying follows a Geometric Brownian Motion and starts at w0ith $\sigma = 2\%$, the barrier is set to , while the interest rate to , and dividend to . Time to maturity starts at 2 years and ends 1 year before the term. The European to One Touch Digital Price Ratio is approximately (a bit less) across the whole period, meaning that The One Touch option is twice the price of European Digital. Its range varies between and . However, this shows that our methodology can be used to price American Digital and European Digital by simply computing the closed-form formula (19), which can be seen as the price of a One Touch option before discounting and taking its half. Note that the volatility chosen for the underlying process impacts more the price of the one-touch option, thus changing this parameter leads to a slight deviation from this equality.

F European Digital Price from the American Digital vs. Black-Scholes

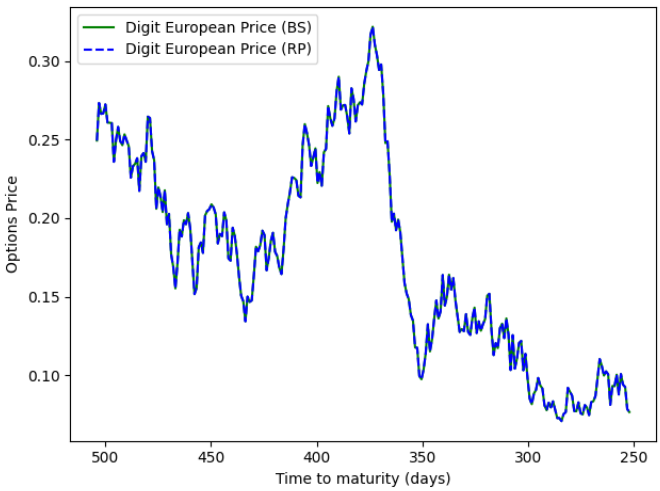


Figure 7: Price of a European Digital computed from an American Digital (RP) compared to the price of a European Digital from the Black-Scholes closed formula (BS), assuming constant volatility across strikes. Here the underlying follows a Geometric Brownian Motion and starts at $S_0 = 100\%$ with $\sigma = 20\%$, the barrier B is set to 120%, while the interest rate to 2%, and dividend to 0%. Time to maturity starts at 2 years and ends 1 year before the term. In order to determine the degree of similarity between the two methods, we measure the average squared difference (MSE) between the corresponding data points and find that the MSE is 4^{-19}. From this MSE score, we can admit that the methods of pricing give similar price. As mentioned in Figure 6 the volatility chosen for the underlying process impacts more the price of the one-touch option, thus changing this parameter leads to a slight deviation in this equality.

G Effects of different volatilities and maturities on the sensitivity of probability that a stock cross the barrier B over $t \in [0, T]$

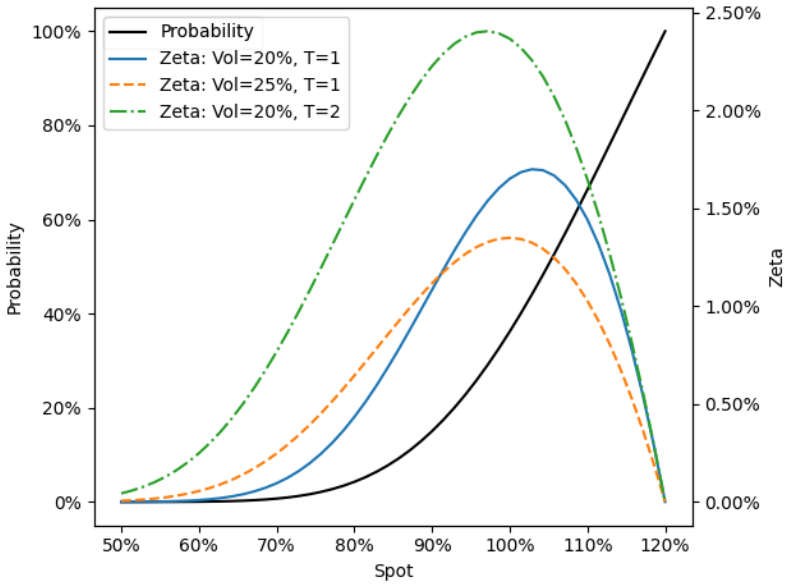


Figure 8: The sensitivity of probability of a stock surpassing the barrier B over the interval $t \in [0, T]$ to interest rate is larger as T increases. However, higher volatility makes the bell-shaped higher when the spot is far from the Barrier but lower as the spot approaches the Barrier. Here volatility of the underlying is set to 20%, dividend to 0%, time to maturity to 1 year, the Barrier to 120%, and the interest rate to 2%.

H Comparison between the computation of ζ using Monte Carlo and the CDF of a Geometric Brownian Motion

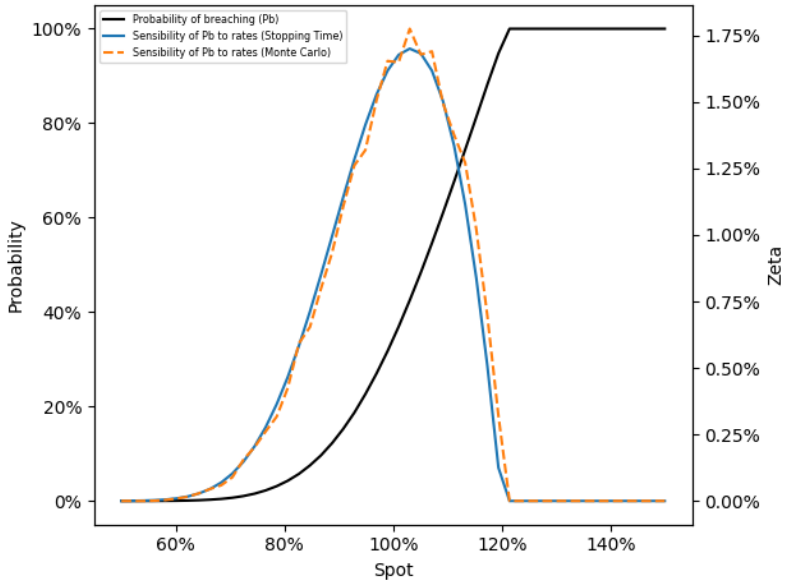


Figure 9: The sensitivity of probability that a stock crosses barrier B over $t \in [0, T]$ using the CDF of a Geometric Brownian Motion is consistent with the results found using Monte Carlo. Here volatility of the underlying is set to 20%, dividend to 0%, time to maturity to 1 year, the Barrier to 120%, the interest rate to 2%, the step time for the American Monte Carlo to 1 day, and the number of simulated path to 1 000 000.

Single Factor Modeling

Building a Financial Stress Index (FSI)

Contents

Introduction and motivation

Detecting stress is an essential part of the asset allocation process. The aim of this study is to identify financial tensions that can be considered as a more or less strong and persistent disturbance to the normal functioning of markets. Identifying markets under stress is not as straightforward as it might seem. Financial stress is by definition an unobservable variable. From an empirical point of view, stress is generally accompanied by an increase in the historical volatility of risky assets, a rise in the value of safe-haven assets or an increase in the price of credit default swaps (CDS), a reduction in valuation multiples, an increase in short- and medium-term refinancing costs and a widening of spreads. The impact of this stress can therefore be seen globally and simultaneously on a wide range of financial assets.

The intuition is based on the idea that each exogenous variable reacts in a synchronized and specific way to a market stress situation. Aggregating this information would, in principle, enable us to construct a synthetic indicator of market stress. To achieve this, we opt for a one-factor model on high-frequency data (daily), which will be estimated iteratively on a large number of market variables, to be presented in the section devoted to the presentation of the 1.

The benefits of this approach are manifold. It allows a large number of variables to be taken into account without compromising their respective interpretations, enabling the origin of risk to be identified according to geographical zone or asset class. This approach also makes it easy to integrate data with different histories without modifying the estimation methodology. This methodology can therefore integrate time series with heterogeneous histories. Finally, it should be noted that the model is based exclusively on daily public market data available on Bloomberg or Refinitiv.

In the remainder, we spent the first part to the data description and the second part to the presentation of the model. The next sections will be devoted to the description of the implementation process of this indicator, and we will end with the presentation of the results and focus on event studies.

This work is inspired by multiple research articles regarding stress indicators, as cited in the references. It draws notable inspiration from the article "The OFR Financial Stress Index" by Monin (2017).

1 Data

Data selection and pre-processing are essential to the construction of our stress indicator. As our indicator will be estimated using a single-factor model (SFM), the common factor linking all exogenous variables will be interpreted as a proxy for financial stress. Consequently, the interpretation of the dataset must be performed at level. In other words: "When the level of the variable is high, it denotes a stressful situation, while low levels reflect the absence of stress". To achieve this interpretation of variables, it may be necessary to apply transformations and rotations, which will be explained in detail below.

1.1 Volatility Metrics

First, volatility measures are natural candidates for capturing a stressful environment. The following indicators are considered:

Equity Market	**Bond Market**	**Fx Market**
VIX Index	SMOVE EU1M	Nikkei Implied Volatility
V2X Index	SMOVE EU3M	EURUSD Implied Volatility
Brent Volatility (realized)	MOVE 1M	JPYUSD Implied Volatility
Emerging Markets Implied Volatility	MOVE 3M	
Put Call Ratio	SMOVE US1M	
	SMOVE US3M	

These metrics capture the volatility of various asset classes and geographical regions. Volatility can be either realized or implied, depending on the market. For instance, Brent crude oil volatility is realized

as it is primarily traded through delta one future-type products. The MOVE and SMOVE indices reflect bond market volatility trough numerous tenors. The Put Call ratio, although not strictly a measure of volatility, directly indicates investors' risk appetite and is suitable for inclusion among the volatility indicators. Lastly, volatility on options for major currencies serves as a natural metric, reflecting stress differentials between geographical regions. For obvious reasons several volatility data exhibit high correlations. For instance, the VIX and V2X are highly correlated, as are the MOVE and SMOVE for different maturities. To maintain a balanced approach, we decide to introduce new measures based on both on spreads and slopes approaches. The following transformations are applied:

1. V2X is replaced by the VIX-V2X spread.
2. SMOVE EU 3M is replaced by the SMOVE EU 1M - 3M spread.
3. The MOVE 3M is replaced by the MOVE 1M - 3M spread.
4. The US 1M and 3M SMOVE are removed from the series due to their high correlation with the 1M and 3M MOVE, as they do not provide additional information.

This set of variables does not require further transformations as an increase in any of them can already be seen as a raise in stress.

1.2 Safe Assets

The second group of variables is "safe assets". The selected variables are:

Safe Assets
Bund YTM (generic 10Y)
T-Note YTM (generic 10Y)
USDCHF
USDJPY
DXY Index
Gold (generic)

US and German long term interest rates are proxies of risk free asset both in the US and Euro Zone (EZ). The DXY index, which return the value of the US dollar against a basket of international currencies, is also used as a relevant indicator of stress as the USDJPY and USDCHF pairs. Gold price is a considered as a safe haven as well by investors. **However, this set of variables cannot directly be use as such. Their levels must be compared to their recent dynamic. The following transformations are then applied:**

- DXY, USDCHF, USDJPY, and Gold are transformed into average log returns over a moving window of 250 observations (1 year). Their current levels are now compared to their respective moving average by calculating their relative deviation.
- Bund and T-Note rates are adjusted for their moving average over a window of 250 observations (1 year), allowing the rates to be analyzed in terms of absolute deviations from their moving average.

Moreover, German and US rates, as well as USDCHF and USDJPY, are multiplied by -1. An increase in these rotated variables translates a raise in market rates.

1.3 Funding

The third category capturing stress is "Funding," which reflects short and medium-term funding conditions. The selected variables in this category are:

The aforementioned variables are used to observe the prevailing conditions of short-term funding for banks, and other markets participants. Liquidity can be frozen when financial markets came under pressures which is reflected in liquidity costs.

Funding metrics
TED Spread (T-Bills 3M - LIBOR USD 3M)
Basis Spread EU 1M VS 3M 1Y
Basis Spread US 1M VS 3M 1Y
TONA SOFR SWAP 2Y
€STR SOFR SWAP 2Y

1.4 Valuation

The fourth segment analyzes the level and dynamics of equity market valuation, which can be seen as a proxy of risk appetite. The metrics used in this category are:

- Price to Book for S&P500 Index
- Price to Book for MSCI Europe Index
- Price to Book for Nikkei 225 Index
- Price to Book for MSCI Emerging Markets Index

We have selected the price to book indicators to capture equity valuation. Adding other equity valuation such as P/E or earnings yield would provide almost the same information than the Price to Book. The price to book indicators are transformed using the same approach as described before :

- Price to book are adjusted for their 1 year moving average and seen in relative deviations. This transformation has the advantage to cut stochastic trends from time series.
- Price to book are multiplied by -1. In such a way that an increase in this indicator translates into an increase in stress (i.e., a fall in price to book, a fall in equity valuation).

1.5 Credit and CDS

The last category focuses on credit and CDS-related risks. The selected variables are:

OAS indices	CDS indices
Bloomberg Credit US IG OAS Index	IBOX US HG 5Y
Bloomberg Credit US HY OAS Index	IBOX US HY 5Y
Bloomberg Credit EU IG OAS Index	IBOX EM 5Y
Bloomberg Credit EU HY OAS Index	ITRX JPN IG 5Y
Bloomberg EM GCC Credit & High Yield Benchmark	ITRX EU IG 5Y
Bloomberg EM Local Currency Government Bonds Universal Index	ITRX EU HY 5Y

For credit index data, credit spreads or prices are used for the emerging markets. Credit indexes are only adjusted from their moving average over 1 year. Price indexes are transformed into average log returns over a moving window of 1 year. CDS indexes expressed as prices are multiplied by -1 and transformed into average log returns over a moving window of 250 observations (1 year). CDS indexes expressed as spreads are adjusted for their moving average over 250 observations (1 year).

After the initial preprocessing specific to each variable, a final normalization step is applied to all the variables. This involves standardizing all the previously transformed variables based on their own characteristics (Z-scores).

1.6 Geographical and risk repartition

The variables used in this analysis are classified based on a risk approach and they are obviously associated with geographical regions. We can therefore split the stress origine in accordance with risk and geographical groups. Prior to delving into model estimation and results, it is essential to acknowledge the presence of a selection bias in the variable proportions chosen. The charts (1a, 1b, 2a, 2b, appendix A) depict the distribution of the various classes.

We observe an over-concentration of variables within certain categories. The US and European data are obviously more prominently represented than other regions. This is firstly due to the abundance of indicators available and secondly to their systemic feature. Note also, variables related to market volatility are over-represented. This is explained by the fact that these variables are calculated for a large panel of assets over different geographies. Secondly, they are the most closely monitored indicators to assess the level of stress in global markets. However, it is essential to keep in mind the characteristics of our dataset during the interpretation phase.

2 Modeling

In this section, we introduce the model used to derive the synthetic indicator from an extensive database.

2.1 Single Factor Model (SFM)

The synthetic indicator is based on the simple idea that different financial assets or indices, etc. tend to **"re-correlate" with each other during periods of financial stress**. Thus, we employ a Single Factor Model (SFM) to capture the underlying common factor that links all variables in our dataset.

$$X_{i,s} = \omega_i f_s + e_{i,s} \tag{1}$$

$X_{i,s} \in R$: The i's indicator observed at date s;
$f_s \in R$: The common factor at date s;
$\omega_i \in R$: The loading of indicator i to the common factor f_s

In contrast to classical Principal Component Analysis (PCA), it is important to note that in our approach, the weight vector used in the Single Factor Model (SFM) is **re-estimated every time the index is computed**. This re-estimation takes into account all available information up to time t and does not consider any information from subsequent dates. It is worth mentioning that past values are not revised or recalculated in this process.

Since the index is computed daily using available information, it is crucial that the data undergoes appropriate transformations presented in the previous section. This ensures consistent interpretation of variations and maintains stability in key statistical measures such as mean and standard deviations. By achieving this stability, we can provide a uniform daily interpretation of the risk indicator without being affected by any potential edge effects. These edge effects, if not addressed, could lead to abrupt movements in the indicator purely due to technical reasons.

According to equation 1, we can rewrite the model in a vector form as :

$$X_s = \omega f_s + e_s \tag{2}$$

$X_s \in R^n$: The row vector of the n's indicator at date s;
$f_s \in R$: The common factor at date s;
$\omega \in R^n$: The indicators' loadings to the common factor f;
n = Number of indicators.

In essence, this approach involves an iterative process of conducting principal component analysis on a preprocessed dataset. The objective is to maintain the manageability and stability of the model as the considered timespan expands. The model is solved using the least squares method, following the formulation of an unobservable one-factor model with unobservable loadings. This formulation necessitates to apply an iterative algorithm, such as the Gibbs algorithm. Notably, the adoption of the least squares estimator for the common factor is supported by several empirical studies conducted by Bernanke, Boivin, and Eliasz (2005), Boivin and Giannone (2006), Stock and Watson (2006), and Eickmeier and Zielger (2008). These studies provide valuable insights that motivate the use of the least squares estimator in this context.

NB: The index s indicates that we consider an observation at date $1 < s < t$, while the index t indicates that the variable is taken from $s = 0$ to $s = t$, representing the complete time-series which is extended by one observation each day as the model is re-estimated.

2.2 The modeling framework

The model is estimated using least squares and is solved through an iterative algorithm that shares similarities with the Gibbs algorithm in a Bayesian framework. We have :

$$\min_{\omega_i f_s} \sum_{i,s} (X_{i,s} - \omega_i f_s)^2$$

$$\iff \min_{\omega_i f} \sum_{i=1}^{n} \|X_i - \omega_i \mathbf{f}\|_2^2 \tag{3}$$

$$\iff \min_{\omega_i f} \sum_{i=1}^{n} \|X_i' - \mathbf{f}'\omega_i\|_2^2 \tag{4}$$

By applying the first-order condition, we take the derivative of the loss function and use the properties of positivity and convexity of the square function.

$$\hat{\omega}_{ols} = (\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\mathbf{X} \text{ based on equation 3}$$

and

$$\hat{\mathbf{f}}'_{ols}|\omega = (\omega'\omega)^{-1}\omega'\mathbf{X}' \text{ based on equation 4}$$

Based on this approach, we can employ an iterative algorithm similar to Gibbs' algorithm. We start with an initial point $\omega_{(0)}$, which is initialized using the first principal component of the dataset[1]. We then iterate the Gibbs algorithm:

$$\mathbf{f}'_{(n),t} = (\omega'_{(n)}\omega_{(n)})^{-1}\omega'_{(n)}\mathbf{X}' \text{ and } \omega_{(n+1)} = (\mathbf{f}'_{(n),t}\mathbf{f}_{(n),t})^{-1}\mathbf{f}'_{(n),t}\mathbf{X}$$

For $n = 1$ to 200 (arbitrary number of iteration to stabilize the solution). Since $\hat{\mathbf{f}}'_t = (\hat{\omega}'\hat{\omega})^{-1}\hat{\omega}'\mathbf{X}' = \frac{\hat{\omega}'}{\|\hat{\omega}\|_2^2}\mathbf{X}' = \sum_{i=1}^{n} \frac{\hat{\omega}_i}{\|\hat{\omega}\|_2^2} X_i$ where X_i is the i'th column of $\mathbf{X}$ representing the time series of the indicator i.

Consequently, we can readily determine the marginal contribution of each variable to the common factor at any given date t. Additionally, we can decompose the common factor based on specific indicators or indicator categories, such as risk class or geographical class. This provides us with a deeper understanding of the underlying factors driving the synthetic indicator.

NB: During the estimation process, a constraint is imposed on the loadings of the indicators, requiring them to have a unit $\mathcal{L}_2$ norm. Consequently, this manipulation can be viewed as a mere translation of the index, having no significant impact on the model. It is important to note that the SFM model allows for multiple solutions that differ only by rotation. Therefore, the constraint of unit norm on the loadings is not restrictive in this model.

[1] We also perform this several times with different starting points to make sure of the stability of the answer (equal-weight vector, random vector).

3 Implementation

As mentioned in Section 2, the model is re-estimated on a daily basis using the available data up to that date. This iterative process allows for the calculation of the indicator's long history by estimating the weight vector ω_s at each date. With these weight vectors, it becomes possible to compute an estimator of the common factor $\mathbf{f}_s$ using the formula $\hat{\mathbf{f}}'_{ols}|\omega = (\omega'\omega)^{-1}\omega'\mathbf{X}'$ presented in Section 2.2. This approach ensures that the indicator is continuously updated and reflects the changing dynamics of the underlying data.

The code is divided into 3 parts:

- The "*Toolbox.py*" module: This module includes various functions that are useful for data processing, error management, data visualization, and model estimation. It serves as a toolbox of functions that can be utilized throughout the project.
- The "*Update.py*" module: This module handles the preprocessing of data and performs the model estimation at a specific date. It incorporates frequent consistency checks using the functions from the "*Toolbox.py*" module to ensure the accuracy and reliability of the results.
- The "*Main.py*" module: This module is responsible for retrieving the data and executing the "*Update.py*" file for each date. It saves the results, including interactive graphs in HTML format, to be used for further analysis or presentation.

By dividing the code into these three modules, the project achieves a modular and organized structure, facilitating code reusability, maintenance, and readability. A diagram explaining the implementation of the procedure for estimating the daily stress indicator is attached (see diagram 7, appendix C).

4 The Financial Stress Index

4.1 Global Results

After calculating the entire history using the methods described in 2 and the algorithm presented in 3, we obtain the FSI indicator with its three levels of decomposition: indicators, risk class, and geographical class (charts 3, 4, 5, appendix B).

The numerical level of stress is set arbitrarily as the variables are centered and scaled, and the weights are normalized with respect to the $\mathcal{L}_2$ norm. The stress level should therefore easily be interpreted . For example, the index exceeded 20 during the 2008 crisis, ranged between 4 and 7 during the European debt crisis episode and reached 15 during the COVID crisis. It must therefore be interpreted in relation to other events, but also in relation to its recent past (daily, weekly, monthly, etc.).

For each date, we can break down the level of stress provided by the FSI into the individual components, geographical zone and asset class (fund, credit, valuation, etc.). We first observe the most significant contributors are those associated with volatility. This result is consistent as volatility and implied volatility measures are indicative of market sentiment. We then identify various contributors depending on the period. Credit and CDS are predominating, followed by valuation measures (proxied by price to book). From a geographical standpoint, we can see that the main contributors are first the United States followed by Europe. Depsite the biais in the data construction, this outcomes shows how central are these markets in measuring financial stress. We also observe a category entitled "Cross-continental", dedicated to variables related to currencies and commodities that cannot be naturally be attributed to a single zone.

4.2 Event study : SVB and Credit Suisse failures

We note that the FSI exhibits the expected behavior during past stress events. For example, it locates the stress in March 2023 particularly in the USA and volatility variables, coinciding with the difficulties

faced by SVB. We can also mention the stress in 2011 related to the Eurozone debt crisis, for which the FSI localizes a significant portion of the stress in Europe and credit/CDS variables, once again aligning with the rational expectations associated with these stress events.

Let's take a look at recent bank failures. In March 2023, a series of regional bank failures in the USA sent shockwaves and panic through the financial markets. These bankruptcies naturally triggered support measures from the US Treasury, the Federal Reserve and the FDIC. In addition, the Saudi National Bank's decision to withdraw its commitment to recapitalize Credit Suisse triggered a sharp fall in the latter's share price, again leading to the implementation of a rescue plan. The speed and concerted nature of the measures taken helped to contain the fire. Economic repercussions were virtually non-existent.

The Financial Stress Index (FSI) quickly detected this increase in stress, attributing it mainly to volatility variables (see graph 6b, appendix B), both implied and realized. It is worth highlighting the temporal precision of the FSI, which initially pinpointed a significant stress event in the USA through the collapse of SVB, Signature Bank, and SilverGate Bank. Just a few days later, a similar phenomenon occurred, this time originating in Europe with the Credit Suisse's events. (see graph 6a, appendix B). What's more, our indicator also shows us how long it will be before market stress shows any signs of abating. This final characteristic of the indicator potentially makes it highly valuable as a signal to remit market sensitivity in the portfolio during and shortly after a market stress situation.

5 Conclusion

In conclusion, the index presented proved to be a valuable tool for assessing overall stress on a daily basis. Throughout this analysis, we have drawn key lessons from its effectiveness in flagging key stress events and identifying key contributors. **The FSI has no predictive purpose, but rather provides valuable information on the dynamics of stress**. By successfully identifying sources of stress, such as the SVB bankruptcy and the euro-zone crisis, it has proved its ability to capture and dissect crucial events affecting the financial markets. What's more, the index allows us to tell when stress begins to subside, which is crucial information for allocators. Further developments can be considered, the performance of the index can be improved. An in-depth analysis of the FSI's "decay factor" could provide useful information on the duration of the effects of stress. Exploring the application of this methodology to detect phases of complacency, euphoria or bubbles could pave the way for further researchs in risk assessment or market forecasting. In addition, refining the distribution of variables within the various categories and sub-categories will improve the index's accuracy and relevance. In summary, the index is a powerful tool for monitoring and understanding financial stress and, if further refined, offers great potential to help market players and researchers make informed decisions and mitigate risk in the dynamic landscape of quantitative finance.

References

[1] Monin (2017). The OFR Financial Stress Index

[2] Bai and Ng (2002). Determining the Number of Factors in Approximate Factor Models

[3] Bernanke, Boivin, and Eliasz (2005). Measuring the Effects of Monetary Policy Factor-Augmented Vector Autoregressive (FAVAR) Approach

[4] Boivin, and Giannoni (2006). DSGE Models in a Data-Rich Environment.

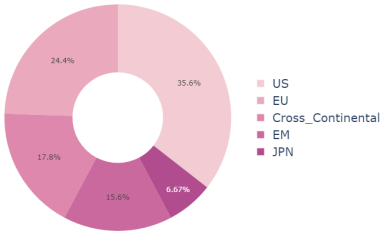
[5] Eickmeier, and Ziegler (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach.

[6] St Louis Federal Reserve (2022). The St. Louis Fed's Financial Stress Index, version 4.

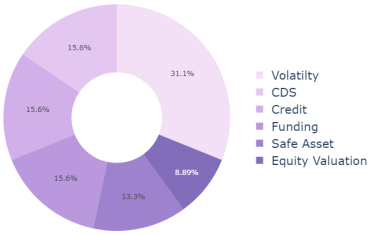
[7] Cleveland Federal Reserve (2012). The Cleveland Financial Stress Index.

[8] Kansas City Federal Reserve (2010). The Kansas City Financial Stress Index.

Appendix A : Geographical and Risk repartitions

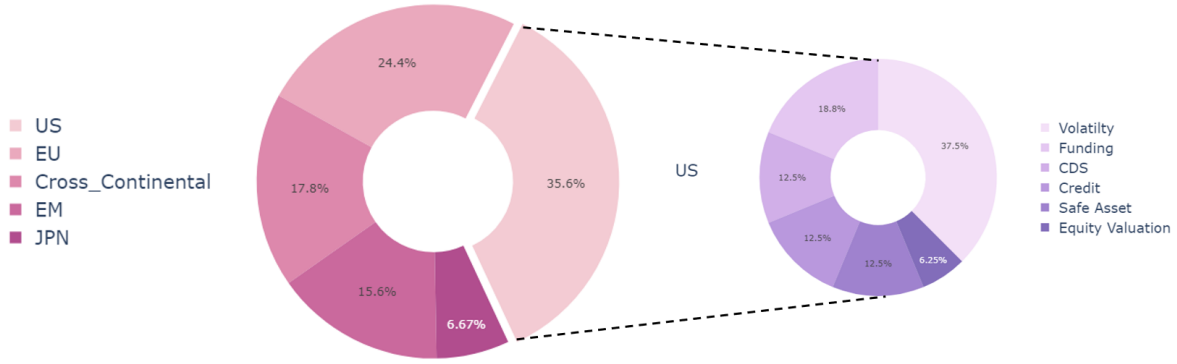


(a) Geographical decomposition of the dataset.

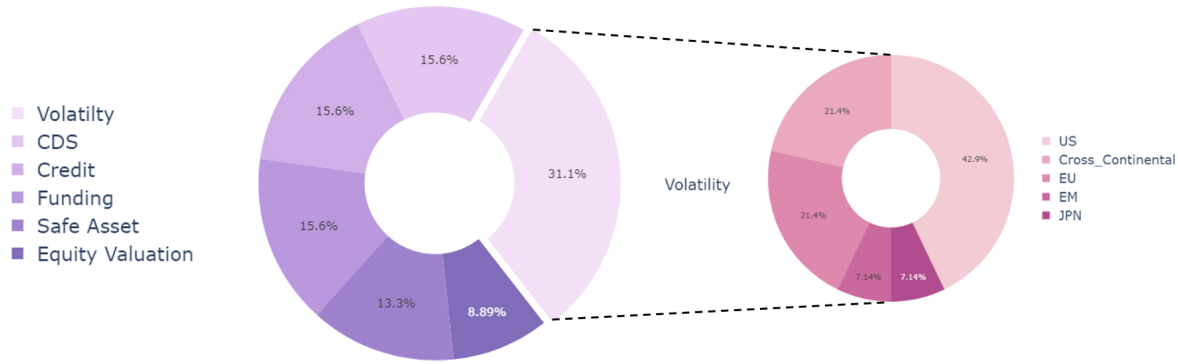


(b) Risk class decomposition of the dataset.

Figure 1: Distribution of variables across categories



(a) Focus on US Breakdown



(b) Focus on volatility Breakdown

Figure 2: Focus on in-class breakdown

Appendix B : Results

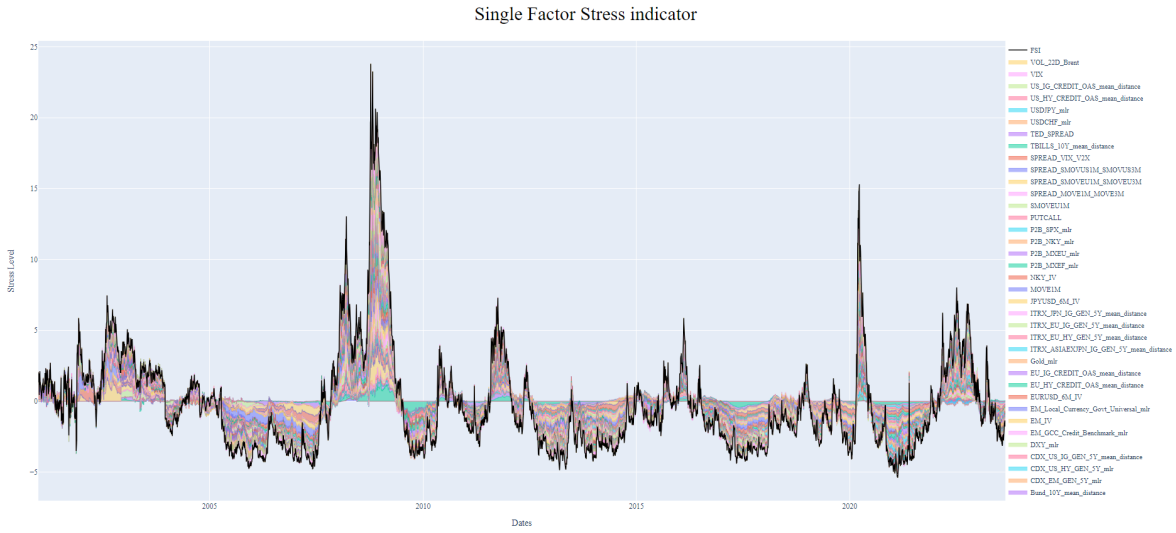


Figure 3: FSI with indicators decomposition.

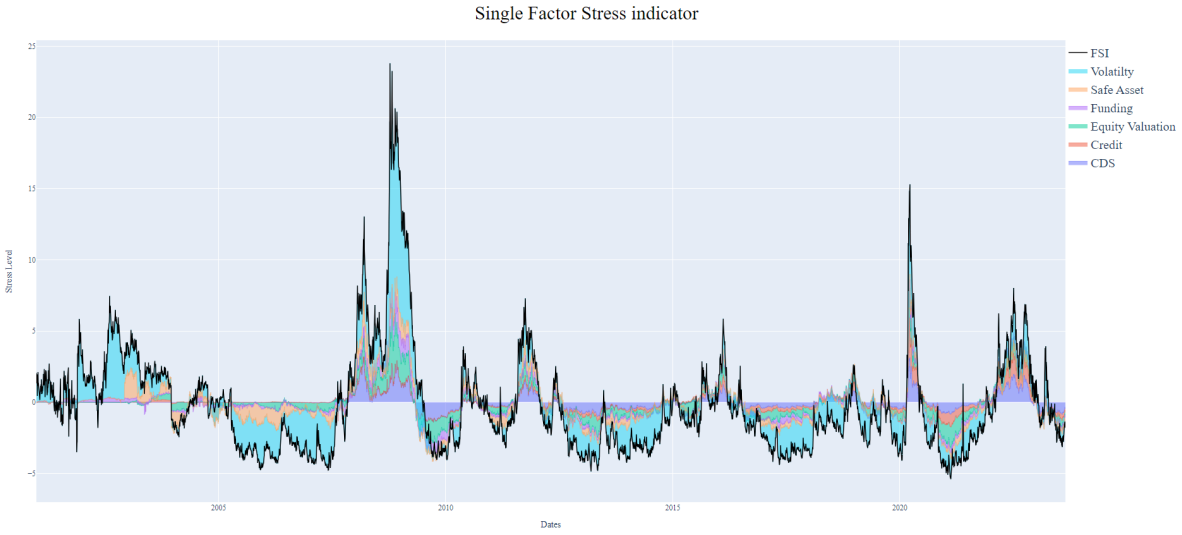


Figure 4: FSI with risk class decomposition.

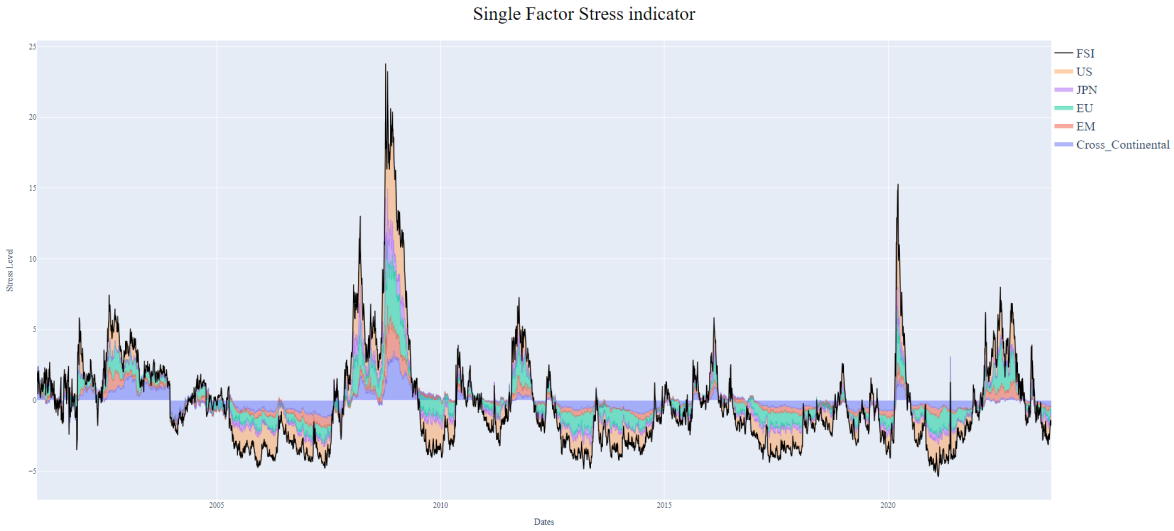
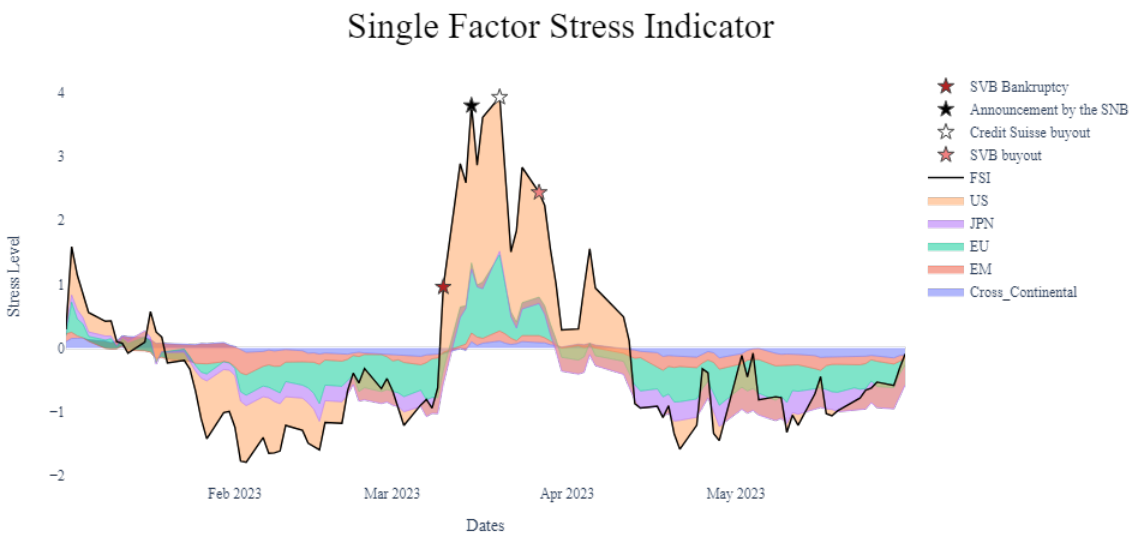
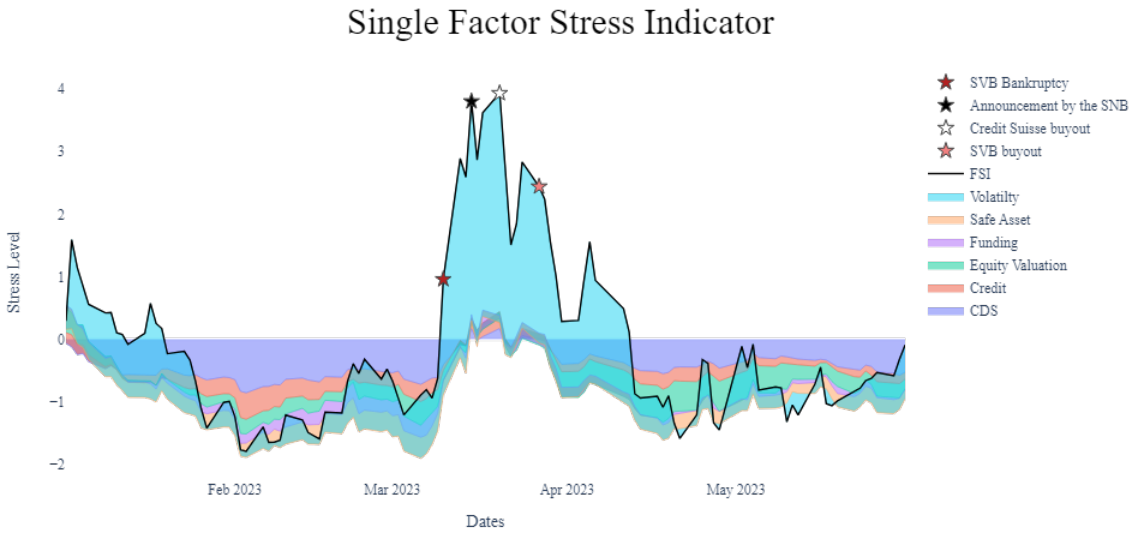


Figure 5: FSI with geographical class decomposition.



(a) Focus on March 2023 events: A geographical perspective



(b) Focus on March 2023 events: A Risk drivers perspective

Figure 6: Focus on SVB and Credit Suisse events

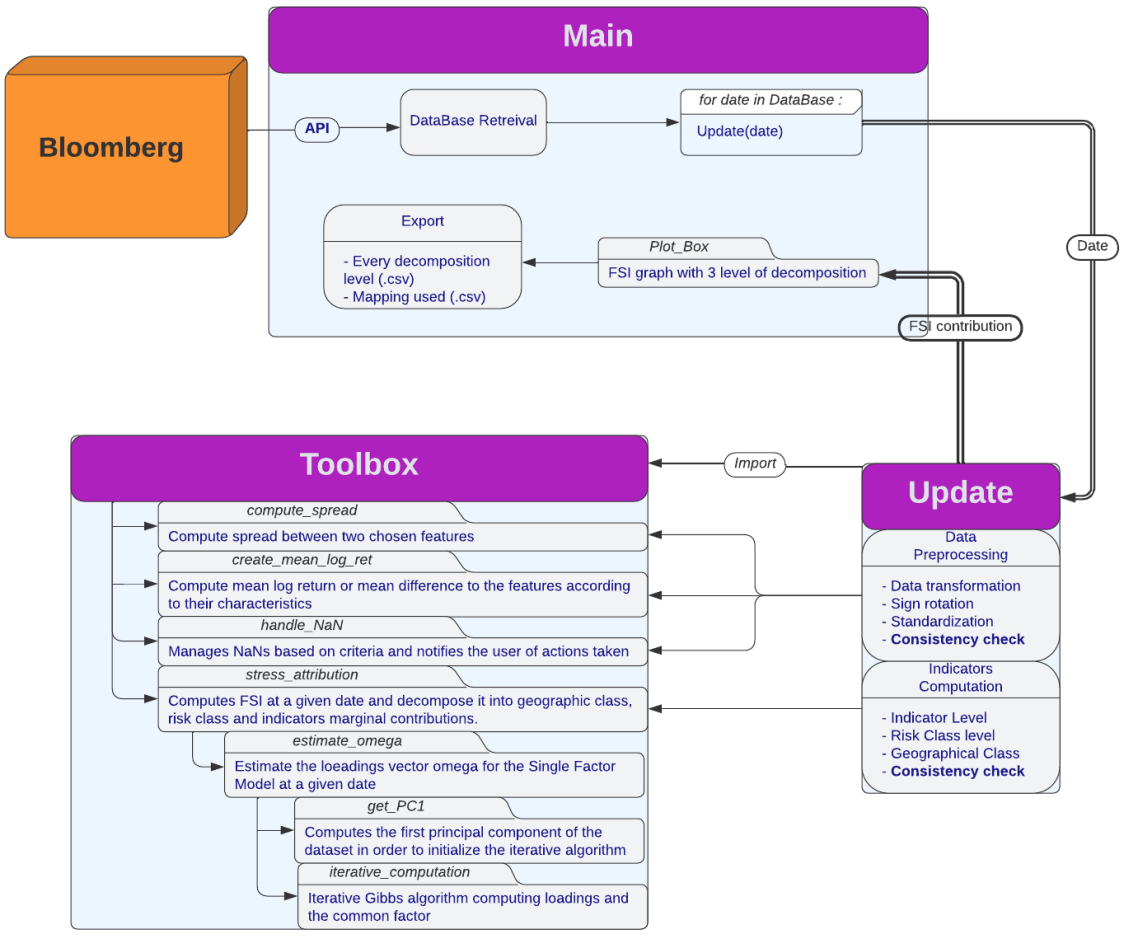


Figure 7: Implementation architecture