How Alternative Are Private Markets?[1]

William N. Goetzmann[2] Elise Gourier[3] Ludovic Phalippou[4]

August 2019

PRELIMINARY AND INCOMPLETE

Abstract

We present a new and flexible methodology to build factors in illiquid markets, from an unbalanced panel of smoothed asset returns. We apply this methodology to a large and unique panel of private market funds. We build a set of eight private factors that capture nearly 60% of the variation in private market returns. Four of these factors command a risk premium above 3%, but half of the variation in their returns is explained by standard listed equity factors. Exposure to these factors can be gained by forming a portfolio with given fund characteristics that include region of investment, industry focus, investment strategy, and fund size.

[1]We thank Burgiss for supplying the data, and especially James Bachman for feedback and data guidance, and Wendy Hu for patiently executing all of our codes and sending us the anonymized output. We thank Pierre Collin-Dufresne, Francesca Cornelli, Thierry Foucault, Andras Fulop, René Garcia, Amit Goyal (discussant), Emmanuel Guerre, Michel Habib, Emmanuel Jurczenko (discussant), Kevin Kaiser, Tarun Ramadorai, Sergei Sarkissian (discussant), Berk Sensoy (discussant), seminar participants at Amsterdam U., Bank of Portugal, Bath U., ESSEC, INSEAD, HEC Paris, Liverpool U., London School of Economics, Paris Dauphine U., and Zurich U., participants to the HF&PE Paris-Dauphine Conference, the ESSEC Amundi Workshop, the 7th Luxembourg Asset Management Summit, the 2018 PERC Symposium, the 6th ESSEC Empirical Finance Workshop, the 2019 SFS Cavalcade and the Glion Private Capital Conference. We are also grateful to the Q-Group for the 2018 Jack Treynor Prize. The authors have consulted to various financial institutions. In particular, Ludovic Phalippou has been an advisor on private market research for the Blackrock Investment Institute. None of the authors have any conflicts of interest to report.

[2] Yale School of Management - International Center for Finance; National Bureau of Economic Research (NBER), 165 Whitney Ave., P.O. Box 208200, New Haven, CT 06520-8200, United States; tel: (+1)2034325950; Email: william.goetzmann@yale.edu.

[3] Corresponding author. ESSEC Business School and Centre for Economic Policy Research (CEPR), 3 avenue Bernard Hirsch, 95021 Cergy-Pontoise, France; tel: (+33)1 3443 3270; Email: elise.gourier@essec.edu.

[4] University of Oxford - Said Business School, Park End Street, Oxford, OX1 1HP, United Kingdom; tel: (+44) 1865 288 719; Email: ludovic.phalippou@sbs.ox.ac.uk.

1 Introduction

A fundamental question about private market funds is whether they access payoffs that are not spanned by factors available in public securities markets. A closely related question is whether there is an underlying factor structure latent in private markets that delivers premia or higher expected returns to funds exposed to these factors.

In the forty years since Ross introduced the APT, scholars and practitioners have sought to identify a parsimonious set of priced factors latent in the space of asset returns. Virtually all of this research has focused on the time-series of returns to publicly traded securities because the theory postulates that factors are identified by the covariance structure – something that is hard to measure for non-traded assets. The difficulty of extending this analysis to private markets stems from the irregularity of cash-flows, the non-linearity of the fee structure, valuations that are appraisal-based, limited fund lives, and the heterogeneity of investments within private market funds (see, e.g., Metrick and Yasuda (2009), Sorensen, Wang, and Yang (2014)).[1] The premise of such an investigation is that all priced factors can be accessed via publicly traded securities. If instead the market for factors is segmented between public and private markets, then private markets present opportunities for investors to access additional risk premia.

This paper presents a new methodology to estimate a parsimonious set of latent common factors in illiquid markets, and applies it to a large panel of private market funds. We rely on two steps to build these private factors. The first step infers fundamental fund returns from appraisal valuations that rely (in possibly complex ways) on prior period estimates, thereby inducing autocorrelation in returns. The second step identifies the latent factors that drive the variation in fundamental returns. These steps are however not independent as inference on the autocorrelation function depends on the factors that have been identified. Our methodology is a clustering approach that addresses these two steps jointly and builds private factors from auto-correlated data structured as

[1] Boyer, Nadauld, Vorkink, and Weisbach (2018) use transaction prices on the secondary market for fund stakes. Only Leveraged Buyout (LBO) funds between 2006 and 2017 can be analyzed due to the small number of transactions for other fund type and in early years. Even for that sub-set, the number of transactions is limited and Boyer, Nadauld, Vorkink, and Weisbach (2018) use a statistical model based on fund characteristics to construct an index of LBO returns, from which an alpha can be evaluated. Ang, Chen, Goetzmann, and Phalippou (2018) work with private equity fund cash flows, take fund classification as given, and work only with LBO, Venture Capital, and Real Estate funds. They use a Bayesian approach to infer the most likely loadings on pre-specified linear factor models and the corresponding alpha.

an unbalanced panel.

We use cash flows and quarterly appraised net asset values (NAVs) as inputs and unsmooth them statistically to estimate *fund-level* return time-series. The unsmoothing stage is motivated by the assumption that the appraisal process is non-Markov, whereas factors – at least those in public capital markets – are likely to follow a random walk. Our specific econometric innovation is a model for (smoothed) fund returns, which represents observed returns as a linear function of past observed returns and specifies autocorrelation as a linear combination of a set of several exogenous variables related to the fund (e.g., fund age) and the economy (e.g., recession times). We use methods from the machine learning literature to select the relevant covariates that enter the specification of the autocorrelation function. We invert this function to estimate a time-series of unsmoothed fund returns that are "best" estimates of the fund fundamental returns.

We use the time-varying group fixed effects model of Bonhomme and Manresa (2015) and represent the dynamics of unsmoothed returns as sum of a group effect (the factor), and a fund-specific component. We cluster these unsmoothed return series using a standard distance metric, and each cluster represents a dynamic, ex post buy-and-hold portfolio strategy of subsets of the private market fund universe. By design, returns of funds within each of these portfolios are driven by a common latent private factor, which is estimated. Such approach ensures maximal homogeneity of fund returns driven by each private factor. In the appendix, we formally show that our estimators for the clusters and factors are statistically consistent under a set of standard assumptions.

We apply our methodology to a comprehensive panel of nearly 5000 funds provided by Burgiss, a service firm catering to private market investors. We do not restrict the sample by fund type, but instead incorporate new asset types as they appear through time (e.g., private credit, infrastructure). We find that private fund returns are best explained by a set of eight private factors. Four of the eight private factors command sizable risk premia in the cross-section of private market fund returns.

Investing in these factors requires knowledge on the mapping of funds to clusters. We find that the investment strategies that are commonly used to classify funds are not sufficient criteria to determine which factor a given fund yields exposure to. In particular, US Buy-Out (BO) funds are spread out across all of the eight factors. The situation is similar, albeit less strong for Venture Capital (VC) funds. Within BO and to a certain extent VC, we cannot find characteristics to explain why funds are assigned to different clusters. More narrowly defined categories, however, are clearly associated with a specific factor.

We observe all European private funds (except those focusing on Venture Capital) clustering together, thereby forming a first factor. Non-small (i.e. largest three quartiles) Venture Capital funds, especially those focused on Early-stage IT companies, are forming another factor. US non-small Real Estate funds form a third factor. US non-small Distressed Debt funds form a fourth factor. Energy (Oil&Gas) funds form a fifth factor (size does not matter here). The factor comprising most funds gathers funds with a low-risk profile, which tend to be US-focused and small: General Debt, US Mezzanine, small US Distressed Debt, small US Real Estate, Timber, and US Infrastructure. The other two factors cannot be easily characterized. Our factors capture 57.2% of the total variance of private market returns.

Four of our private factors are relatively well spanned by a 5-factor model that includes the US market equity factor, the size factor [SMB] of Fama and French (2015), the alternative value factor [HMLd] of Asness and Frazzini (2013), the quality of earnings factor [QMJ] of Asness, Frazzini, and Pedersen (2018), and the low-beta factor [BAB] of Frazzini and Pedersen (2014). The (significant) loadings on the SMB, HMLd, BAB, and especially QMJ factors suggest that our private factors capture part of the variation in these factors.

As our private factors are only partly spanned by public factors, our results suggest that the quest for APT factors should not be limited to public capital markets. Focusing on public capital markets ignores the issue of endogenous choice about ownership and control of enterprises. Perhaps some assets perform better, or more true to their underlying factor exposures, when held by private capital. This paper shows that private markets provide exposures that public markets do not, thereby offering an additional source of factor risk premia. This may help to understand why institutional investors regard private markets as a source of diversification.

The remainder of the paper is organized as follows. Section 2 discusses the institutional background of private markets. Section 3 introduces the data and present key descriptive statistics. Section 4 presents our statistical model. Section 5 reports the main empirical results and section 6 concludes.

2 Background

The past twenty years witnessed an important shift in the asset allocation of institutional investors. Figure 1 shows how the $10 trillion in pension fund assets worldwide moved from a 60-40 US stock-bond asset allocation to an allocation that is nearly equally spread across five types of assets: US public equity, non-US public equity, US listed debt (fixed income), non-US listed debt (fixed income), and alternative investments. Within alternatives, private market funds are the main piece and have grown significantly over time to reach assets under management of $5 trillion in 2016, exceeding that of hedge funds ($3 trillion).

[Figure 1 about here.]

[Table 1 about here.]

Another important and clear stylized fact is that the allocation to private markets is strongly related to investors' size. For example, Table 1 shows the results published in the 2018 NACUBO-TIAA survey of US endowments. The relationship between the size of the endowment and the allocation to private markets is striking.[2]

Most of the investments in private markets are accessed via specialized funds, which we refer to as private funds. Private funds differ from public market investment vehicles in several ways. Investors commit ex-ante to providing a given amount of capital over a fixed period of time (e.g. five years). The quantity, amount and timing of capital calls are uncertain at inception. So is the type of investments, which do not necessarily match the announced intention at the time of fund raising. The announced objective of most funds is to exit all their investments by their tenth anniversary. In practice, most funds are active for about fifteen years.

Studying the risk exposures of private funds requires knowledge on the time series of their returns. The main challenge that we face is that these returns are not observable. Instead, what we observe

[2] This relationship between size and private market allocation is also observed for pension funds (non-tabulated).

is cash flows that LPs receive from and pay into each fund, and quarterly NAVs reported by each fund manager.

The rules that govern the estimation of NAVs are SFAS 157 ("Fair Value Measurements") for the US since 2008, and IFRS 13 ("Fair Value Measurement") for Europe since 2006 (Crain and Law (2016)). Yet, fair value accounting has been slowly implemented over time and was in partial use before these dates.

According to IFRS, fair value assessment is an "estimate of the price at which an orderly transaction to sell an asset or to transfer a liability would take place between market participants at the measurement date under current market conditions (i.e., to estimate an exit price)."Yet, NAVs ultimately result from subjective judgments about the appropriate valuation technique and input parameters for each portfolio company. Despite these accounting rules and although NAVs have no direct impact on investors' wealth, fund managers may purposely smooth NAVs with the aim of facilitating investor relationship management (e.g. avoid negative return news)[3], facilitating fund-raising (Barber and Yasuda (2017), Brown, Gredil, and Kaplan (2017)), or because they believe public market returns are excessively volatile.

Crain and Law (2016) provide evidence that NAVs are quite accurate overall, although sluggish. Nadauld, Sensoy, Vorkink, and Weisbach (2018) show that some secondary market transactions are executed at prices that differ significantly from NAVs. These doubts around NAV accuracy is an issue and we design a statistical approach to address it as precisely as possible.

3 Data

3.1 Datasets

We use a comprehensive dataset of private funds, which is collected and maintained by Burgiss. Our dataset is as of June 2018 and should be the largest and best-quality dataset available (Harris,

[3] Private funds are structured as closed-end funds. In open-ended funds, investors can buy and sell their fund stakes at a price equal to the NAV on a daily basis. In closed-ended funds, investors can only trade their stakes with one another at a price which may deviate from NAV. As a result, an inaccurate NAV in a closed-end fund does not create a wealth transfer between buying and selling investors (unlike for open-ended funds).

Jenkinson, and Kaplan (2014)). For each fund, the dataset includes timed cash flows (to and from investors net of fees), and the time series of quarterly Net Asset Values (NAVs). As cash flows are recorded daily, we make an assumption regarding the re-investment and financing policy for these intra-quarter cash flows, and construct fund-level return time-series at a quarterly frequency.[4]

In addition, we have access to a set of fund characteristics: the year in which the first capital call occurred (a.k.a. *vintage year*), the firm that operates it (from which we derive *firm experience*), and the amount of committed capital (a.k.a. *fund size*);[5] as well as indications on the fund's investment focus (industries, geography, asset class), which we detail in the next section.

We select the funds that started between 1984 and 2013, so that all funds have past their investment period. In addition, we select funds with a size of at least $10 million, with at least twenty pairs of consecutive quarters of return data available, and have been classified by Burgiss. Appendix E details the filters we apply and their impact on sample size.

As our dataset contains non-US funds, we use equity pricing models that are available both globally and domestically. The first model is made up of the five factors posted on AQR's website, and we term this the AQR model. The second model is Fama-French [FF] (Fama and French (2015)). Their five factors are RmRf, SMB, HML, RMW and CMA and are provided on Ken French's website.[6]

The first two factors in AQR's model are similar to those of FF (RmRf, SMB). The third AQR factor also captures the value premium but using a different methodology (HMLd).[7] The fourth factor is the Quality Minus Junk (QMJ) factor of (Asness, Frazzini, and Pedersen (2018). The fifth factor is the Betting Against Beta (BAB) factors of Frazzini and Pedersen (2014)).[8]

[4] We use a Modified Internal Rate of Return (MIRR). Intra quarter distributions (resp. investments) are brought to the end (resp. beginning) of the quarter using an 0.021% daily compounding rate. This daily rate coincides to an 8% annual rate, which is the usual return at which funds start to earn a performance based compensation. All statistics and tests have also been computed with an Internal Rate of Return. Results are similar. MIRR avoids the IRR re-investment assumption; but at this frequency it does not affect results, it just avoids a few outliers.

[5] As fund size is not comparable over time, we categorize funds into vintage-based size quartiles. Small funds are those falling into the first quartile of fund sizes in their vintage year. Mid, Large and Mega funds have a fund size that is, respectively, in the second, third and fourth quartile of their vintage year.

[6] RmRf is the difference between the stock market portfolio return and the risk-free rate. SMB, HML, RMW are differences between the returns of small and big stocks, of high and low book-to-market stocks, and of robust and weak profitability stocks. CMA is the difference in the returns of stocks of conservative and aggressive firms in terms of investment policy.

[7] Asness and Frazzini (2013).

[8] For BAB and QMJ, securities are weighted by the ranked beta. For the global model, portfolios are formed separately for each country and countries are value-weighted by their total (lagged) market capitalization.

Note that some factor return time-series start later than others. As the global FF factors return series start in Q3-1993, our tests involving factor models use the one hundred quarters from Q3-1993 to Q2-2018. Other variables come from WRDS/CRSP, Bloomberg and FRED (Federal Reserve Economic Data).

3.2 Private Fund Classifications

Consultants and data providers categorize private funds using various criteria. We argue that these fund taxonomies basically combine four dimensions thought to capture risk-return profiles: (1) the seniority of the capital claim (e.g. junior debt vs. common equity), (2) the stage of development of the asset (e.g. young vs. mature companies), (3) the industry of investment focus (e.g. natural resources vs. real estate), and (4) the source of return (e.g. opportunistic vs. value-add).[9]

Recently, Burgiss drew on its extensive experience to manually classify each private fund in their dataset using a combination of the above criteria. Table 2 shows the distribution of funds across the Burgiss categories per region of investment focus.

There are three categorization tiers, which vary by their level of granularity. Tier 1 categorization distinguishes between Debt, Real Asset, Equity, and General. Tier 2 categorization further divides each of these groups into a total of twelve categories. Tier 3 is a categorization we designed by combining additional fund characteristics.

Tier-1 is dominated by private equity funds (3,346 funds). There are 1,015 real asset funds, and 190 private market generalists. Finally, there are only 402 private debt funds but this is the fastest growing category.

Tier-2 is dominated by Venture Capital (VC) funds. VC funds are well-known for having backed most of the largest companies in the world (Apple, Amazon, Google, Microsoft, Facebook, Uber, AirBnB). They invest in *equity* of *young* companies. Within VC, we observe 660 funds focusing on Information Technology (IT), 256 funds focusing on health care, 405 funds classified as generalists

[9] Note also that many combinations of the four dimensions are not used in practice, probably by lack of observations. For example, a fund investing in equity claims in a proptech company (young companies operating in the real estate sector with a strong technology orientation) is classified as either Venture Capital (VC) or Real Estate (RE). A debt claim on the same company would probably be classified as private debt.

and the rest have 'unknown' reported as industry. In addition, we observe 868 early stage VC funds, i.e. funds investing in companies that are mostly an idea (they may not have any revenues yet). Early stage VC funds are thought to have higher failure rates than other VC funds. Next, on the stage dimension, we count 152 late-stage funds (focus on companies that may start to generate a profit), and 394 generalist funds (which invest across different stages). Given the distribution of funds, we opt for three tier-3 categories within VC, each of about equal size: early stage funds focusing on IT, other early stage funds, and the remainder (other VC).

In dollar terms, Buy-Out (BO) funds are much larger than VC funds. BO funds invest in leveraged *equity* claims of *mature* companies. The industry of investment focus plays an important role in the classification of BO funds. Funds conducting BOs in infrastructure (e.g. airports, utilities), real estate (e.g. shopping malls, hotels), and natural resources (oil and gas, timber) are classified in separate tier-2 categories. These investments are separated because they are perceived as *real* assets. For the tier-3 classification, we thus opt to distinguish between BOs based on their industry. There are 110 funds focusing on consumer discretionary and staples, 92 funds focusing on industrials, 279 funds are industry generalists, and the remainder is a large category including other specialized funds and funds with unknown industry specialization.

The main Real Asset category is Real Estate (RE). Within Real Estate, funds are usually split into core, value-add and opportunistic. This classification reflects the level of idiosyncratic risk (from least to most). Core real estate funds are basically passive holders of income generating assets. Their returns are close to those of fixed income securities (e.g. renting out student accommodations). Core funds are usually publicly listed closed end fund vehicles, not private market funds. Hence, our sample includes only value-add and opportunistic real estate funds.

A classical example of a value-add real estate strategy consists in buying properties with a high vacancy rate at a substantial discount, and initiating physical and operational improvements to increase the occupancy. Once the property is fully occupied, it may be sold to a core real estate fund. A value-add strategy should not bear much business cycle risk and uses low leverage. Opportunistic funds are basically BO funds focusing on real estate. Compared to value-add strategies, the environment is usually more entrepreneurial, assets are held longer, and the level of leverage is higher.[10] For the real estate tier-3 categorization, we have 235 opportunistic funds, 256 value-add funds, 154 generalists, and the remainder do not have an entry.

Infrastructure (Infra) funds are also sometimes classified as core, value-add and opportunistic. In the Burgiss dataset, they are classified based on their industries of investment focus. As the category is small we only split Infra funds between industry generalists (34 funds) and specialized (58 funds).[11] Similarly, within Natural Resources (NatRes) funds, the main distinction is across industry of investment focus. 113 NatRes funds focus on Energy (oil and gas) and the rest are specialized in other industries (only five are generalists). Hence, we have two tier-3 categorizations: Energy, and NatRes Other.

For private debt funds, tier-2 classification distinguishes between Mezzanine, Distressed Debt and Generalist Debt funds. Mezzanine funds mainly hold junior debt claims. Distressed funds often hold debt securities of companies that are close to bankruptcy with the objective to obtain equity in a restructuring plan. These funds therefore hold a debt claim that is more akin to equity than Mezzanine. We do not design any tier-3 classification for Debt funds.

We also group funds into regions of investment focus. Two thirds of the funds invest in the United States. Yet, more than 80% of Debt, Natural Resources, and VC funds are US-focused.[12] For Western Europe, there is a distinction between UK-focused funds (137 funds), and the rest (which includes Scandinavia, but excludes Eastern Europe). In Asia, we have 121 China-focused funds. Most of these funds are VC. For the rest of the countries we simply distinguish between developed regions (Middle East plus Canada, Oceania) and the remainder.

[Table 2 about here.]

[10] For example, the acquisition of Hilton Hotels was jointly executed by an opportunistic real estate fund and by a BO fund (both run by Blackstone).

[11] There is no particular cluster in any industry.

[12] Infrastructure funds and LBO funds are more spread-out across geographical areas (about one quarter of the funds invest in Europe). Expansion capital funds (a.k.a. growth capital) are most common in emerging markets. Most funds operating in emerging markets are expansion capital funds because it is difficult to purchase controlling stakes, and use leverage in these geographies. These funds tend to be smaller because the underlying companies are small. Most of these funds are in USD and their returns should be sensitive to exchange rate fluctuations.

4 Model

4.1 Factor structure of returns

Estimating private factors requires reducing the dimensionality of the private fund universe to a small set of portfolios that summarize well the dispersion in returns and risk exposures. As a preliminary test of whether private market returns exhibit a factor structure, we run a Principal Component Analysis (PCA).[13] Following Jolliffe (2002)[14], we apply PCA directly on private market returns, ignoring autocorrelation. We find that five (seven and twelve, respectively) principle components explain more than 80% (85% and 90%, respectively) of the total variance in returns. The first component explains about half of this variation. These results indicate that although private market returns follow a clear factor structure, 10 to 20% of their variation cannot be easily spanned by a few common processes and is idiosyncratic.

Beyond the potential issues linked to applying PCA to unbalanced and autocorrelated data, several pitfalls make it suboptimal to use in the context of private market portfolios. First, the resulting factors describe the return variation of portfolios of funds that may involve short positions. Given the challenges linked to selling private market funds, we would prefer factors that can be invested in using long-only portfolios. Second, it is typically challenging to economically understand factors that result from PCA, and to pinpoint which assets should be bought to gain exposure to them. Third, PCA factors have often been shown to perform poorly out-of-sample, as the weights of the different assets in the portfolio may vary substantially with time. To address these issues, we build a set of factors for private market returns using a clustering approach. The main upside compared to a method based on PCA is that the resulting factors are easier to invest in using long-only portfolios of private market funds. As a downside and as a result of the additional constraints imposed on the factors, the variation in returns will be well described by a larger number of factors.

Our methodology clusters funds into portfolios by minimizing the distance between funds' fundamental returns within the same portfolio, thereby creating homogeneous long-only portfolios of

[13] Standard PCA methods are not applicable to unbalanced datasets. We use the Alternating Least Square implementation of PCA, described in Ilin and Raiko (2010), and treat returns before and after a fund's life as missing values.

[14] Jolliffe (2002), p.299, writes: "when the main objective of PCA is descriptive, not inferential, complications such as non-independence [in time] does not seriously affect this objective".

funds. The common return variation of each of these portfolios is one of our factors. With a balanced panel dataset and without appraisal-induced autocorrelation, our estimation would be similar to the correlation-based clustering methods that have been used in the context of individual stock returns (Brown, Goetzmann, and Grinblatt (1997), Ahn, Conrad, and Dittmar (2009)), mutual funds (Brown and Goetzmann (1997)), and hedge funds (Brown and Goetzmann (2003)).[15]

4.2 Unsmoothing returns

The first step is to filter from *observed* returns each fund's fundamental returns. As in Getmansky, Lo, and Makarov (2004), we assume that all the biases that arise from the NAV appraisal process distort the fundamental returns by introducing autocorrelation. Hence, the time-series of unsmoothed returns has no autocorrelation and is our best approximation of fundamental returns.[16]

Formally, let us consider N funds indexed by $i \in \{1, ..., N\}$. We denote by $R_{i,t}^{obs}$ the observed (smoothed) return process between times $t - \Delta$ and t, where $\Delta = 1$ quarter and $t = 1, ..., T$. The time series of unsmoothed returns is labeled $R_{i,t}^{unsm}$.

We model observed returns as a weighted average of past observed returns and unsmoothed returns:

$$R_{i,t}^{obs} = \sum_{l=1}^{L} \theta_{i,l,t} \cdot R_{i,t-l}^{obs} + \left(1 - \sum_{l=1}^{L} \theta_{i,l,t}\right) R_{i,t}^{unsm}, \tag{1}$$

The autocorrelation function is specified as a linear combination of D time- and fund-specific dummy variables.[17] Each dummy variable is collected in the vector $\mathbf{D}_{i,t} \in \mathbb{R}^D$. The coefficients form a matrix $\boldsymbol{\delta} \in \Theta^{D \times L}$. Denoting $\mathbf{X}_{it}$ the vector of lagged returns: $\mathbf{X}_{it} = (R_{i,t-1}, ..., R_{i,t-L})'$, we have:

$$\sum_{l=1}^{L} \theta_{i,l,t} \cdot R_{i,t-l}^{obs} = \mathbf{D}_{i,t}' \boldsymbol{\delta} \mathbf{X}_{it}. \tag{2}$$

[15] Appendix D provides additional discussion of the literature.

[16] Note that, in principle, even fundamental private market returns might be smooth due to, e.g., the inelasticity of the demand and supply of private capital to private funds.

[17] For example, to test whether fund managers smooth NAVs more during recessions, we would estimate:

$$\sum_{l=1}^{L} \theta_{i,l,t} \cdot R_{i,t-l}^{obs} = \delta \cdot 1_{recession} R_{i,t-1},$$

To select the covariates of the autocorrelation function, we use a penalized maximum likelihood applied to the panel of unsmoothed returns. This procedure computes the optimal loadings on covariates so that the unsmoothed returns are as close as possible to i.i.d. normally distributed random variables, and pushes the insignificant loadings to zero by penalizing their norm. We use both a Lasso penalization (Tibshirani (1996)) and an elastic net regularization, shown to often outperform the Lasso method (Zou and Hastie (2005)). We verify that the covariates that are selected by both methods are the same. The regularization parameter is chosen by cross-validation to be within one standard deviation of the optimal parameter, so as to obtain the smallest possible set of selected covariates.

4.3 Model for unsmoothed returns

In a second step, we use the model of Bonhomme and Manresa (2015) to endogenously cluster funds so that the trajectories of their unsmoothed returns are as similar as possible within a cluster.

Specifically, let us consider fund i in cluster $g(i)$, $g(i) \in \{1, ..., G\}$. Its unsmoothed return has three components: a fund-specific constant η_i, a mean zero time-varying component which is common to all the funds in a given group, $\alpha_{g(i),t} \in \mathcal{A}$ (or, for simplicity, $g(i) = g_i$) and labeled *group return*, and an idiosyncratic component $\epsilon_{i,t}$:

$$R_{i,t}^{unsm} = \eta_i + \alpha_{g_i,t} + \epsilon_{i,t}. \tag{3}$$

We estimate jointly the loadings on covariates, group returns and assignments of funds to a group by minimizing the sum of square idiosyncratic terms in both the cross-sectional and time series dimensions:

$$(\hat{\theta}, \hat{\alpha}, \hat{\gamma}) = \underset{(\Theta,\alpha,\gamma)}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}^2, \tag{4}$$

where the minimum is taken over all possible $\theta \in \Theta$, groupings $\gamma = \{g_1, \cdots, g_N\}$ of the N funds into G groups, and the group return trajectory $\alpha_{g_i,t}$.

The minimization problem (4) is solved by alternatively selecting the optimal grouping of funds conditionally on the trajectories of the group returns, and computing the loadings on covariates and group returns conditionally on the grouping of funds. Details on the algorithm and Monte Carlo simulations showing small-sample accuracy of the estimators are provided in Appendix B.

All funds within a group, therefore, have a common time-varying group fixed-effect, which is interpreted as a private factor driving the returns of these funds. This is equivalent to a latent factor model, with one factor per cluster, where all funds within one cluster have a beta of one on the corresponding factor.

4.4 Discussion of the model

The main benefits of our modeling approach are twofold. First, it allows for a flexible and rich modeling of the autocorrelation in observed fund returns. Our autocorrelation function is an extension of the classical $AR(1)$ process introduced to capture appraisal smoothing in real estate by Geltner (1991), Ross and Zisler (1991).[18]

Second, our model for fundamental returns belongs to the class of latent factor models, but does not suffer from the common disadvantage of having to specify the factors driving returns ex-ante, or of imposing a linear structure. Our approach reduces the dimensionality of a dataset endogenously and leaves each group return time series unspecified.[19]

An important issue is the determination of the number of groups. Theoretically, if the estimated number of groups exceeds the true number, the estimator of θ is still consistent. However, the estimated group effects suffer from a finite T bias of order $O(1/\sqrt{T})$. In contrast, if we work with a number of groups that is smaller than in the data generating process, and if the group returns are correlated with the covariates, the estimator of θ becomes inconsistent, due to an omitted-variable bias. The factor structure of the unsmoothed returns will not be well captured by the group returns, which may result in excess correlations across clusters and in residual autocorrelation. As a result, we not only use the Bayesian Information Criterion (BIC) to select the number of clusters (as suggested by Bonhomme and Manresa (2015)), but also use average correlation across clusters and average autocorrelation of factors.

[18] Similar models are applied to hedge fund returns (Kat and Brooks (2002), Getmansky, Lo, and Makarov (2004)), art market returns (Campbell (2008)) and collectible stamp returns (Dimson and Spaenjers (2011)).

[19] The assumption that each fund belongs to a single group may be relaxed with a mixture model.

5 Empirical Results

5.1 Estimation of the Model

We start with the selection of covariates entering the autocorrelation function. We cross the first four lags of the returns with a set of dummy variables that capture fund characteristics (age, size, investment objective), and market conditions (sign of contemporaneous returns, recession, pre-defined private market cycles). The complete list of variables is provided in Appendix C.

[Table 3 about here.]

The selected covariates are displayed in Panel A of Table 3. The first two lags are retained. The first lagged return is larger in times of recession and when the return is exceptionally low (less than -15%).[20] These findings are in line with managers voluntarily smoothing NAVs in order to slowly reveal bad information. The second-order autocorrelation, however, is more sensitive to prior abnormally large (rather than low) returns.[21]

Fund characteristics are unrelated to the first-order autocorrelation, except for experience and US-focused. The economic magnitude of these two cross-effect is small, however. In addition, and most interestingly, we note that fund categorization is not significant. There is no more smoothing of returns in VC than in BO or RE. Also, size buckets, age buckets, most geographical areas, and time-periods are not significant.

Using the selected covariates for the specification of the autocorrelation, we estimate the model for different number of clusters. Note that the LASSO approach was simply used to select the covariates, and that the loadings are estimated at this stage, without penalization. While the point estimates of these loadings can be estimated using simple panel regression, the computation of their standard errors is conditional on the clusters funds are assigned to. The coefficients shown in Panel A result from the joint estimation of the autocorrelation coefficients, cluster assignments, and group return trajectories.

[20] -15% is about two standard deviations away from the mean and about 5% of the returns are below this threshold.

[21] This indicates that when returns are good, observed returns are not smoothed; they are correctly reported. The period next, however, depends on what happened before. If returns were low (at t-2) then following the high return (at t-1), the observed return (at t) tends to be lower than the correct return. Similarly, if returns were high (at t-2) then following the high return (at t-1), the observed return (at t) tends to be higher than the correct return.

Panel B of Table 3 shows indicators of goodness-of-fit for different number of clusters. The BIC statistic is minimal (indicating optimum) at fourteen clusters but the difference between the BIC at eight and fourteen clusters is small. By construction, the MSE, i.e. the objective function evaluated at the parameters optimum, decreases with the number of clusters. However, we observe a slight decay in the MSE gain around eight clusters.

Turning to the time-series of factor returns that correspond to each cluster, we also observe an optimum at eight clusters. The average correlation between the private factor returns is minimized at eight clusters and is quite low overall.[22] As mentioned above, in small samples and with a misspecified number of clusters, there can be residual autocorrelation in the factor returns. We observe that the autocorrelation is lowest at eight clusters. In view of these tests, we opt to work with eight clusters, hence eight private factors.

5.2 Ex ante characteristics of the eight clusters

By design, each fund in our database is allocated to a cluster whose latent common return variation is described by one of our private factors. Investors can therefore trade a given factor by investing in the funds that are allocated to the corresponding cluster. In this section we provide guidance on how to pick funds to gain exposure to a specific private factor.

5.2.1 A simple descriptive approach

Table 4 shows the fraction of funds in a given category assigned to each of the eight clusters. To measure the heterogeneity of these funds across clusters, we compute a Herfindhal index. For example, if all the funds within the category were in the same cluster, the Herfindhal index would be one, whereas if they were equally spread across clusters, the Herfindhal index would be 0.125.

[Table 4 about here.]

We begin with the funds classified as Generalists in their tier-1 classification. In line with their classification, they are spread across clusters. We observe no difference across Generalists in terms

[22] The increase of the average correlation beyond eight clusters indicates that funds with similar return trajectories start to be categorized in different clusters.

of size, geographical focus or experience (non-tabulated). After conditioning upon these variables, Generalists are still spread across clusters.

For funds classified as Debt in their tier-1 classification, we observe that 45% are assigned to Cluster 7 and 19% are assigned to Cluster 4. Going down to tier-2 classification, we note that it is Mezzanine and General Debt funds that are in Cluster 7. Mezzanine funds are more common in Europe than in the US, and we observe a split along that dimension (not along others). 50% of European Mezzanine is in Cluster 1 whereas 64% of US Mezzanine is in Cluster 7.

Distressed Debt funds are separated from other Debt funds: 40% are in Cluster 4. Most Distressed Debt funds are US-focused, but we can analyze whether there is a size or experience differential, and the former is strongly significant. 50% of the small Distressed debt funds are in Cluster 7 alongside Mezzanine and General Debt funds. All of the other size quartile of Distressed Debt funds are in Cluster 4. Hence, we split Distressed Debt between small and non-small funds.[23] These results make economic sense. The payoff of Distressed Debt funds should be more idiosyncratic. These funds are more like arbitrageurs / hedge funds. In fact, many hedge funds pursue Distressed Debt strategies. In addition, within Distressed Debt funds, the large ones (e.g. Apollo) are those who are most active in the loan-to-own strategy, which results in funds holding significant equity positions rather than debt positions. This would explain while the small Distressed Debt funds are more debt-like.

We thus see a clear split of Debt funds into three groups: European Mezzanine (Cluster 1), non-small Distressed debt (Cluster 4), and all the rest in Cluster 7 (small Distressed Debt, US Mezzanine, Generalist Debt).

For funds classified as Real Asset funds in their tier-1 classification, we observe a large heterogeneity. Their Herfindhal index is as low as that of Generalist funds. We nonetheless observe a slight concentration in Clusters 3 and 7.

The largest tier-2 category of Real Assets is composed of Real Estate funds, and they too are widely spread with Clusters 3 and 7, together receiving 62% of the funds. There is again a clear

[23] We do not observe any differences across firm experience for Distressed Debt funds.

split by region with European Real Estate going to Cluster 1 (with European Mezzanine).[24] US Real Estate remains split between Clusters 3 and 7. We do observe a difference across size quartiles between the US Real Estate funds in Cluster 3 versus Cluster 7. As with Distressed Debt, it is the small quartile that is distinct and we therefore show results with small versus non-small US Real Estate. Moreover, as for Distressed Debt funds, Cluster 7 is where the small funds go. Non-small US Real Estate funds are in Cluster 3. Perhaps surprisingly, we observe no difference explained by the tier-3 classification (i.e. Value-add, Opportunistic), or by firm experience.

Natural Resources funds are strongly concentrated in Clusters 5 and 7 (together gathering two thirds of Natural Resources funds). This split is captured by the tier-3 classification: Energy funds are in Cluster 5 whereas Timber funds are in Cluster 7.[25] We do not observe any further split according to size: All Energy funds irrespective of their size equally concentrate in Cluster 5, so do Timber funds. Nearly all Natural Resources funds are US-focused.

Infrastructure appears to be very heterogeneous (low Herfindhal index) with some concentration in Clusters 1 and 7. There is again a clear split by region with European Infrastructure going with European Real Estate and European Mezzanine in Cluster 1. US Infrastructure is mostly in Cluster 7. We do not observe any difference among US Infrastructure funds as a function of size, tier-3 classification, and experience, but the data only contain 45 US Infrastructure funds.[26].

From the results above, we conclude that it is not pertinent to consider Natural Resources and Infrastructure as asset classes, albeit for two different reasons. For Natural Resources, we clearly see that Energy funds are separated from Timber funds, which makes intuitive sense. These two industries have different business cycles, and different idiosyncratic and tail risk profiles. For Infrastructure, geography seems more important. European Infrastructure goes with other European funds and US infrastructure goes with Debt funds and small US Real Estate funds.

The largest (tier-1) type of funds are Private Equity funds, and are those most spread out across clusters. The Herfindhal index is as low as 0.14 for Equity funds. It is not surprising that pooling together VC and BO makes little sense. Yet, the Herfindhal index hardly increases when we separate

[24] This effect is particularly strong for Value-add European Real Estate where 87% of the funds are in Cluster 1, and weakest for Generalist European Real Estate, which may invest more outside Europe.

[25] There are only 35 non-energy non-timber real asset funds; they do not concentrate in any particular cluster.

[26] Generalist Real Assets funds are split between Clusters 5 et 8, but there are only 19 funds in our dataset.

out VC and BO funds. There is, therefore, considerable heterogeneity within VC funds and even more so within BO funds.

Nonetheless, we observe some orthogonality between VC and BO within Clusters 1 and 2. There are only 7% of VC funds in Cluster 1 but 18% of the BO funds are assigned to it. Conversely, 28% of VC funds are in Cluster 2, but only 5% of BO funds are assigned to it. 22 and 23% of VC and BO funds, respectively, are in Cluster 7 – together with Debt and Timber funds–, a cluster that spans across fund categories and gathers funds with lower (idiosyncratic) risk profiles. Expansion funds and Generalist Equity funds are split across clusters.

Regional focus is a distinguishing characteristic for BO funds but less for VC funds. 52% of European BO funds are in Cluster 1 against 28% of European VC. As nearly one third of the Asia-focused funds are VC funds, it is interesting to search for possible distinctive features of these funds, but we do not observe any particular concentration of Asia-focused funds.

US BO funds are spread across clusters. We do not observe any size or experience tilt. Nor do we observe a differentiation based on their industry of specialization (tier-3 classification). We are therefore not able to characterize ex-ante the private factor that US BO funds provide exposure to.

Half of US and Asia VC funds are allocated to Clusters 2 or 7. European VCs are spread out across many clusters. Size plays a role for VC, with a split operating again at the first quartile. Cluster 2 receives 34% of non-small VC funds. We do not observe much difference according to the tier-3 categorization, except for Early-stage IT VC funds being slightly over-represented in Cluster 2. If we isolate Early-stage IT non-small VC funds then 39% of these are in Cluster 2. Hence Cluster 2 is close to being a pure-play VC Cluster, but the effect is not strong and many VC funds are assigned to other clusters.

These results show that the assignment of funds to clusters does not follow the usual asset class classifications. Hence, these usual asset class classifications are not sufficient to determine which private factor a fund provides exposure to. They do not parsimoniously capture the different return dynamics, hence diversification benefits, of the different private market funds. Nonetheless, we can characterize most of the clusters from ex-ante observable characteristics, thereby providing guidance on how to pick funds so as to be exposed to a given private factor. Cluster 1 contains European private market funds excluding Venture Capital (Mezzanine, Infrastructure, Real Estate, Buyout). These are therefore the funds an investor should pick in order to track the risk exposures of Factor 1. Cluster 2 contains non-small Venture Capital funds. Cluster 3 is composed of non-small US Real Estate funds. Cluster 4 contains non-small US Distressed Debt. Cluster 5 is composed of Energy funds (oil and gas). Cluster 7 contains many funds which they tend to be US, small, and low-risk. Clusters 6 and 8 are not well defined.

5.2.2 A Multinomial Logit approach

In this sub-section, we run a similar analysis as that of the previous sub-section but using a statistical approach. We estimate the log relative probability that a fund belongs to a given factor (i.e., the probability that it belongs to this given factor versus to another factor), on a set of dummy variables that cover the tier-2 fund classifications, geographical focus, firm experience, and size quartile. These multinomial logit regressions are run iteratively, eliminating after each step the regressors that are not significant at the 5% confidence level, until convergence. We next cross all regressors that remain with one another and run a regression with the selected variables and their cross-effects.[27]

Table 5 lists the variables with positive loadings that are significant at the 5% level test when using the tier-2 categorization of funds. Next to each variable name we report the corresponding coefficient. Variables are ordered from most economically significant to least. Overall, results in this Table confirms what we described informally in the previous sub-section.

The most important characteristic to identify the first cluster is the fund geographical focus. Funds investing in Western Europe (including the UK) have a relative probability of being in cluster 1 that is over 6 times larger than funds with other geographical focuses. Funds classified as BO, generalist equity, RE or infrastructure have a relative probability to be assigned to cluster 1 that is 2.5 to 3.5 times larger than other categories of funds (respectively).

Cluster 2 is dominated by VC funds. Being a VC fund multiplies the relative probability of being in cluster 2 by more than 10. Running the multinomial logit regression for tier-3 categories reveals

[27] For example, if the dummies corresponding to BO and mega fund have been selected, we build a new dummy that is 1 for all mega BO funds.

that for Early IT funds, this number increases to nearly 30. It is in line with the intuition that early-state VCs have risk-return profiles that are distinct from late-stage VCs. While the former have a high failure rate and few but strong home runs, the latter may be closer to the return profile generated by other private market funds.

Cluster 3 is characterized by RE funds, especially US RE. Cluster 4 is characterized by Distressed Debt, and mega funds. Cluster 5 is characterized by Natural Resources funds, and the effect is stronger if we distinguish between Energy and Timber funds (non-tabulated). These funds have a relative likelihood of being in cluster 5 that is around 10 times higher than funds with other characteristics. Cluster 6 was not easily identified in the previous sub-section. here it appears that there is an Asian tilt. Cluster 7 is characterized by Mezzanine funds and US BO funds. We had difficulties to split US BO funds in the previous sub-section. The statistical approach allocates them to Cluster 7. Cluster 8 also has a European tilt, as Cluster 1, although much weaker.

These results confirm the separation between funds investing in Europe, the US and Asia. This implies that we need to distinguish across regions for private markets, perhaps even more so than by usual asset class labels. In addition, the common distinction between mega BO and mid-market BO, for example, does not seem to be supported by the data. We should nonetheless stress that at present the results from these multinomial regressions are sensitive to the selection process for the explanatory variables and we plan to use more advanced methods (LASSO type) to run this analysis.

[Table 5 about here.]

5.3 Private Factors

5.3.1 Description of Private Factors

Table 6 reports descriptive statistics of our eight private factors. The time series of factors are displayed in Figure 2. By design, our factors have mean zero and are not traded. However, by identifying the characteristics of funds that belong to each cluster, we can identify which ones provide exposure to each factor. Around 10% of the funds in our database provide exposure to each factor, except for Factor 7 (Debt-funds dominated), which results from the cluster of a quarter of our database. In terms of market capitalization (total fund size), the portfolio of funds that form this factor represents 21% of the total capitalization. The split of other factors according to capitalization is similar. Factors 1, 4 and 7 together weigh nearly 60% of the total capitalization. The clusters that were not well defined (6 and 8) are the smallest ones. Factor 7 is also the factor with the lowest annualized volatility: 9% against up to 26% for Factor 2 (VC-dominated). Factor 3 (RE-dominated) is the next least volatile one (15%) while other factors have similar volatility.[28] The autocorrelations of our factors are on average rather small (9%), only one of our factor (Factor 2) exhibits a significant remaining AR(1) term.

[Figure 2 about here.]

[Table 6 about here.]

Our model can be interpreted as a one-factor model where all funds within a clusters are exposed to the same factor, with a beta of one. As the PCA uncovered earlier that one factor can only, at best, capture half of the variation in returns, we test how much of the total variance our factors explain, when taken all together. We therefore regress unsmoothed returns on the eight private factors. We find that they explain 57.2% of the total variance.

5.3.2 Private Factors & Standard Equity Asset Pricing Models

We next measure the extent to which private factors are spanned by public factors. There are many public factors proposed in the literature. As we have funds that are operating both in the US and in Europe, we use the two factor models that have both a domestic version and an international version.[29] The first model contains the five factors of Fama-French (FF). The second model is taken from the AQR website; it contains the market factor and the size factor of FF, plus the alternative value factor of Asness and Frazzini (2013), the Quality Minus Junk (QMJ) factor of Asness, Frazzini, and Pedersen (2018) and the Betting Against Beta (BAB) factor of Frazzini and

[28] By construction, the average returns of our private factors are zero.
[29] In non-tabulated results, we use the liquidity factor of Pastor and Stambaugh (2003) but did not find any significant loadings on it.

Pedersen (2014). Results for the domestic AQR model are reported in Table 11.[30] Importantly, as our private factors are not trade-able assets, the coefficients in this table cannot be interpreted as standards alphas and betas. By construction, private factors have a zero-mean, hence the constant, although included, is always near zero and thus not reported. The loadings on the factors could, at best, be interpreted as proxies for betas.

The Rm-Rf (public equity risk premium) is highly correlated with our Factors 1 and 4. The correlation is significant (but less strong) with Factors 2, 6 and 8. Hence the factors dominated by Debt, RE and Energy funds have low correlation with the public equity premium. We might expect a higher correlation for Factor 2 (VC-dominated) and we checked for simple non-linearity but we find significant concavity with respect to the equity risk premium, rather than convexity.

SMB is strongly negatively correlated with Factor 3, indicating that the RE funds selected in the corresponding cluster co-move with the size premium, performing better when large caps outperform small caps. HML is strongest with Factor 2. This is reassuring: Factor 2 is dominated by VC funds and perform better when growth stocks outperform value stocks. Only Factor 4 has a positive coefficient on HML, which may be explained by the strong overlap between distressed and value stocks.

QMJ is negatively correlated with all the factors and (weakly) significantly related to Factors 4 and 5 (i.e. Large Distress and Energy funds). BAB is positively correlated with four factors and negatively correlated with four others and the magnitude are large. This public market factor is clearly relevant for private market funds. This result extends the evidence in Frazzini and Pedersen (2014) that BO funds target publicly-traded low beta companies and are exposed to the BAB factor. Of particular interest, Frazzini, Kabille, and Pedersen (2018) show that the QMJ and BAB factors explain most of the alpha of Warren Buffet, who has always had significant exposure to private markets.

Factor 4 (large Distressed Debt) is most positively related to BAB. Factors 1, 3 and 5 are also positively and significantly related. Unsurprisingly, the VC-dominated factor (Factor 2) is significantly negatively related to BAB.

[30] Results are similar with other models and are thus shown in Appendix F.

When we use the Fama-French model (see Appendix), we find that RMW and CMA are both significantly negatively correlated with Factors 4 and 8, except for Factor 4 and RMW, for which it is positive.

Overall, these public factors explain a significant share of the time-series variation of private factors, but not all of it and not for all private factors. About half of the variance in Factors 1, 2, 4 and 8 is explained by standard equity asset pricing models. These private factors are therefore public equity-like. The other factors, which are dominated by funds commonly classified as Debt and Real Assets are not well spanned by these standard asset pricing models.

[Table 7 about here.]

In non-tabulated results, we also analyze whether lagged and squared public equity factors provide a better fit to the returns.[31] These additional factors are usually not significant. One exception is that most factor returns are increasing and concave in BAB. The other exception is for lagged return in Factor 2. They are all (but QMJ) significant. In particular Factor 2 has a large positive loading on lagged public equity premium and a large negative one on lagged SMB, HML and BAB.[32]

5.3.3 Private Factors & Asset/Macro Indices/Variables

The private factors that have low R-square on standard asset pricing models are dominated by funds that are classified as non-equity (Real Assets, Debt), which may explain the low explanatory power. We now look at how the private factors load on macroeconomic variables and on publicly traded assets that are closest to those held by the non-equity private market funds.

[31] Harvey and Siddique (2000) specify the marginal rate of substitution to be quadratic in the market return and derive an asset pricing model where expected returns on an asset are a function of expected market returns and square returns. The aim is to price securities with nonlinear payoffs. VC returns have features similar to those of out-of-the-money equity call options; whereas LBO returns may have features similar to those of at-the-money equity call options. Similarly, we could argue that Debt funds may have returns resembling those of equity put options. In addition, Glosten and Jagannathan (1994) show that a value can be assigned to managers' skill and generates a nonlinear payoff. Non-linear fees may also induce non-linearities.

[32] A substantial literature including Fung and Hsieh (1997), Ackermann, McEnally, and Ravenscraft (1999), Getmansky, Lo, and Makarov (2004), Agarwal and Naik (2005), study the non-linearities in the risk and returns of hedge funds. In non-tabulated results, we test whether any of the five factors of Fung and Hsieh (2001) are related to our private factors. These five factors are obtained by constructing lookback straddles on five different option markets: Bond, Commodity, Currency, Short Term Interest Rate, and Stock Index respectively. As hedge funds are usually grouped with private funds as Alternative Assets, it is interesting to study the extent to which our factors load on the same risk factors as those shown to capture Hedge Fund return dynamics. Only the short-term interest rate option has a negative and statistically significant coefficient. The fact that many private market funds, especially BO funds, use substantial leverage and regularly refinance their underlying investments may explain this result.

The correlations between private market factors and a set of macroeconomic factors are shown in Table 8. Of potential interest to institutional investors is whether any of our private factors offer a hedge against inflation. For example, most pension funds require an inflation hedge because they have nominal liabilities. We find that Factor 5, which is the factor dominated by Natural Resources funds (especially, Energy funds), offers a strong hedge against inflation, while the other factors do not.

Industrial production growth has a significant and large positive effect for all factors. This confirms that private funds are pro-cyclical and generate returns that are correlated to business cycles. Loadings on credit spreads are overall negative. The negative sign is intuitive: when credit spreads increase, risky credit becomes more expensive, risk premia increase, and as a result existing investments decrease in value, especially those financed with debt.

Factor 1 does not load much on credit spread, which is surprising given that it is dominated by BO funds, hence funds that are mainly financed by leverage. However, the credit spread captures the spread between BAA and AAA rated debt, whereas BO funds are usually more exposed to High Yield debt (a.k.a. junk bonds). Consistent with this intuition, we find that the returns of Factor 1 are indeed positively correlated with T.Rowe high yield bond returns. They are therefore lower when interest rates on high yield debt increase. In line with these results, we find no relation between factor returns and long-term government bond returns.

The VIX index, which is considered a proxy for the degree of uncertainty in the economy is not related to private factor returns. By design, most private fund returns should resemble those of options on equity. VC returns should be similar to those of out-of-the-money equity call options; whereas BO returns have features similar to those of at-the-money equity call options. Similarly, Debt funds may have returns resembling those of an equity put option. However, when we use returns on Put options and Collar options, we find little correlation, except for a positive relation between Collar returns and the returns of Factor 1.

[Table 8 about here.]

Panel B of Table 8 shows several strikingly high correlations between private factors and stock indices that cover similar industries to the private funds. We collected stock indices and mutual funds that invest in similar types of assets as private equity funds. The FTSE REITs is an index of publicly listed closed end funds investing in Real Estate, very much like Real Estate Private Equity funds. Dimensional Fund Advisors flagship mutual funds invest in companies similar to those targeted by LBO funds. The LPX indices are based on publicly listed companies that operate in the venture capital, mezzanine or generally the private equity sector (LPX VC, LPX Mezz and LPX 50 indices, respectively). We also use some widely used indices on Commodity and Natural Resources (whose returns may relate to those of Natural Resources funds) and an Infrastructure index.

5.4 Private factor risk premia

Our factors are the latent drivers of positive-weighted portfolios of funds, which, by construction, capture the dispersion in private fund returns. We test whether these factors are priced in the cross-section of funds by running Fama-MacBeth regressions. Following Avramov and Chordia (2006), Gagliardini, Ossola, and Scaillet (2016) and Ang, Liu, and Schwarz (2018), we do not form portfolios but exploit the full information present in our dataset by running these regressions at the fund level. Table 9 reports the prices of risk of each of the eight factors as well as their Newey-West t-statistics.

The first factor, dominated by Europe-focused funds –mostly BO, RE and Infrastructure– , and the sixth factor, characterized by Natural Resources, are found to be significantly priced in the cross-section of private funds. Quarterly risk premia for these two factors are respectively of 4.5% and 3.9%. Factors 2 and 8 also exhibit sizable risk premia, although these are not significant. The other private factors have a negligible risk premium. If we form portfolios based on Burgiss primary classification (Debt, Equity, Real asset or Generalist) instead, none of the risk premia are significant.

[Table 9 about here.]

6 Conclusion

Private funds are primarily an ownership structure with a broad range of financing terms and industry exposures (e.g. energy exploration, cryptocurrencies, direct lending) seeking to profit from the purchase, active management and resale of companies. Understanding the risk exposures and diversification benefits of private funds is of interest not only for academics, but also for practitioners. Most institutional investors use some form of the Markowitz mean-variance framework for asset allocation decisions. The implementation requires as inputs expected returns, variances and a correlation matrix for a set of predefined asset classes. Our approach can be used to provide guidance on how to optimally categorize funds into distinct asset classes, and sheds light on the extent to which private funds complement exposures to public markets.

We show that it is feasible to use cash flows and appraised values to group PE funds into meaningful categories that diversify portfolios and add exposure to risk premia. These groups differ from industry classifications and thus may add information relevant to investor choice. Our methodology lets us study the exposure of these groups to known traded factors and test whether private fund returns are spanned by them. We find evidence that private market funds are not entirely spanned by publicly available factors. We also provide a richer set of insights about the varied exposure of fund type to traded factors. We conclude that private funds offer something more that exposure to traded factors. Depending on the classification of the fund and the specification of our model, they offer exposure to a set of compensated, systematic latent factors. A question to be considered is whether these premia are due to alpha or smart beta. To the extent they are systematic within groups as opposed to orthogonal to returns, the smart private beta hypothesis seems more likely.

References

Ackermann, C., R. McEnally, and D. Ravenscraft, 1999, "The Performance of Hedge Funds: Risk, Return and Incentives," *Journal of Finance*, 54, 833–74.

Agarwal, V., and N. Y. Naik, 2005, "Hedge Funds," *Foundations and Trends in Finance*, 1(2).

Ahn, D.-H., J. Conrad, and R. F. Dittmar, 2009, "Basis Assets," *Review of Financial Studies*.

Ang, A., B. Chen, W. N. Goetzmann, and L. Phalippou, 2018, "Estimating Private Equity Returns from Limited Partner Cash Flows," Forthcoming in the Journal of Finance.

Ang, A., J. Liu, and K. Schwarz, 2018, "Using Stocks or Portfolios in Tests of Factor Models," Jacobs Levy Equity Management Center for Quantitative Financial Research Working Paper.

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

Asness, C., 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*, pp. 1–79.

Avramov, D., and T. Chordia, 2006, "Asset pricing models and financial market anomalies," *Review of Financial Studies*, 19(3), 1000–1040.

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

Bauwens, L., and J. V. K. Rombouts, 2007, "Bayesian clustering of many GARCH models," *Econometric Reviews*, 26, 365–386.

Bollen, N. P. B., and B. A. Sensoy, 2016, "How Much for a Haircut? Illiquidity, Secondary Markets, and the Value of Private Equity," Working Paper.

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

Boyer, B. H., T. Nadauld, K. Vorkink, and M. S. Weisbach, 2018, "Private Equity Indices Based on Secondary Market Transactions," Fisher College of Business Working Paper No. 2018-03-021. Charles A. Dice Center Working Paper No. 2018-21.

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

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

——— , 2003, "Hedge funds with style," *The Journal of Portfolio Management*, 29(2), 101–112.

Brown, S. J., W. N. Goetzmann, and M. Grinblatt, 1997, "Positive Portfolio Factors," Yale School of Management Working Paper No. F-57.

Buss, A., R. Uppal, and G. Vilkov, 2016, "Asset Prices in General Equilibrium with Recursive Utility and Illiquidity Induced by Transactions Costs," INSEAD Working Paper No. 2014/01/FIN.

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

Cochrane, J. H., 2005, "The risk and return of venture capital," *Journal of Financial Economics*, 75, 3–52.

Constantinides, G., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 94(4), 842–62.

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

Di Bartolomeo, D., and E. Witkowski, 1997, "Mutual fund misclassification: Evidence based on style analysis," *Financial Analysts Journal*, 53(5), 32–43.

Dimmock, S. G., N. Wang, and J. Yang, 2018, "The Endowment Model and Modern Portfolio Theory," Working Paper.

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

Dor, A., and R. Jagannathan, 2002, "Understanding mutual fund and hedge fund styles using return based style analysis," National Bureau of Economic Research. (No. w9111).

Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

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

Frazzini, A., D. Kabille, and L. H. Pedersen, 2018, "Buffett's Alpha," *Financial Analysts Journal*, 74(4).

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

Frühwirth-Schnatter, S., 2006, *Finite Mixture and Markov Switching Models*. New York: Springer.

Frühwirth-Schnatter, S., and S. Kaufmann, 2008, "Model-based clustering of multiple time series," *Journal of Business & Economic Statistics*, pp. 78–89.

Fung, W., and D. A. Hsieh, 1997, "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds," *Review of Financial Studies*, 10(2).

Fung, W., and D. A. Hsieh, 2001, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers," *Review of Financial Studies*, 14(2).

Gagliardini, P., E. Ossola, and O. Scaillet, 2016, "Time?Varying Risk Premium in Large Cross?Sectional Equity Data Sets," *Econometrica*, 84(3), 985–146.

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

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

Glosten, L. R., and R. Jagannathan, 1994, "A Contingent Claim Approach to Performance Evaluation," *Journal of Empirical Finance*, 1, 133–160.

Hansen, P., N. Mladenovic, and J. Perez, 2010, "Variable neighbourhood search: methods and applications," *Annals of Operations Research*, 175, 367–407.

Harris, R. S., T. Jenkinson, and S. N. Kaplan, 2014, "Private Equity Performance: What Do We Know?," *Journal of Finance*.

Harvey, C. R., and A. Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55(3).

Ilin, A., and T. Raiko, 2010, "Practical Approaches to Principal Component Analysis in the Presence of Missing Values," *Journal of Machine Learning Research*, 11, 1957–2000.

Jolliffe, I., 2002, *Principle Component Analysis*. Springer-Verlag, 2 edn.

Juárez, M. A., and M. F. J. Steel, 2010, "Model-based clustering of non-Gaussian panel data based on skew-t distributions," *Journal of Business & Economic Statistics*, 28, 52–66.

Kaplan, S., and A. Schoar, 2005, "Private equity performance: returns, persistence and capital flows," *Journal of Finance*, 60(4), 1791–1823.

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.

Korteweg, A., and S. Nagel, 2016, "Risk-Adjusting the Returns to Venture Capital," *Journal of Finance*, 71(3), 1437–1470.

Korteweg, A., and M. Sorensen, 2010, "Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies," *Review of Financial Studies*, 23, 3738–3772.

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

Ljungqvist, A., and M. Richardson, 2003, "The cash flow, return and risk characteristics of private equity," Working paper, NYU Stern.

Longstaff, F. A., 2009, "Portfolio Claustrophobia: Asset Pricing in Markets with Illiquid Assets," *American Economic Review*, 99(4), 1119–44.

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.

McNicholas, P. D., and T. B. Murphy, 2010, "Model-based clustering of longitudinal data," *The Canadian Journal of Statistics*, 38, 153–168.

Metrick, A., and A. Yasuda, 2009, "The Economics of Private Equity Funds," *Review of Financial Studies*, 23, 2303–2341.

Nadauld, T. D., B. A. Sensoy, K. P. Vorkink, and M. S. Weisbach, 2018, "The Liquidity Cost of Private Equity Investments: Evidence from Secondary Market Transactions," Forthcoming in the Journal of Financial Economics.

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

Phalippou, L., 2014, "Performance of Buyout Funds Revisited?*," *Review of Finance*, 18(1), 189–218.

Phalippou, L., and O. Gottschalg, 2009, "The Performance of Private Equity Funds," *Review of Financial Studies*, 22, 1747–1776.

Robinson, D. T., and B. A. Sensoy, 2016, "Cyclicality, Performance Measurement, and Cash Flow Liquidity in Private Equity," *Journal of Financial Economics*, 122(3), 521–543.

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.

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

Sorensen, M., N. Wang, and J. Yang, 2014, "Valuing private equity," *Review of Financial Studies*, 27(7), 1977–2021.

Stafford, E., 2017, "Replicating Private Equity with Value Investing, Homemade Leverage, and Hold-to-Maturity Accounting," Working Paper.

Steinley, D., 2006, “K-means Clustering: A Half-Century Synthesis,” *British Journal of Mathematical and Statistical Psychology*, 59, 1–34.

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

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

A Model assumptions and asymptotic properties

Assumptions 1.a to 1.h are similar to Bonhomme and Manresa (2015) whereas Assumptions i and j are specific to our model. These assumptions are required for consistency. The superscript ^{0} is used to denote the true parameters.

Assumption A.1. *There exists $M > 0$ such that:*

a Θ *and* $\mathcal{A}$ *are compact subsets of* $\mathbb{R}$.

b $\mathbb{E}(\|\mathbf{X}_{it}\|^2) \leq M$, *where* $\|\cdot\|$ *denotes the Euclidean norm.*

c $\mathbb{E}(\epsilon_{it}) = 0$ *and* $\mathbb{E}[\epsilon_{it}^4] \leq M$.

d $\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(\epsilon_{it}\epsilon_{is}\mathbf{X}_{it}\mathbf{X}_{is})\right| \leq M$.

e $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\epsilon_{it}\epsilon_{jt})\right| \leq M$.

f $\left|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}Cov(\epsilon_{it}\epsilon_{jt},\epsilon_{is}\epsilon_{js})\right| \leq M$.

g *For* $d \in \{1,...,D\}$, *let* $\overline{X}^d_{g\wedge\tilde{g},t}$ *denote the mean of* $\mathbf{X}_{it}$ *over* i,t *such that* $\mathbf{D}_{it} = 1$, *in the intersection of groups* $g_i^0 = g$, *and* $g_i = \tilde{g}$. *Let* $\hat{\rho}^d$ *be the minimum eigenvalue of the following matrix, where the infimum is taken over all possible groupings* $\gamma = \{g_1,...,g_N\}$:

$$\inf_{\gamma\in\Gamma_G}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{it}\left(\mathbf{X}_{it}-\overline{X}^d_{g_i^0\wedge g_i,t}\right)\left(\mathbf{X}_{it}-\overline{X}_{g_i^0\wedge g_i,t}\right)'.$$

Then $plim_{N,T\to\infty}\hat{\rho}^d = \rho^d > 0$.

h $\exists\, b$ *such that* $\forall i,t$, $|\sum_{l=1}^{L}\theta_{i,l,t}| \leq b < 1$.

i $\forall i$, $rank\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it}\mathbf{X}_{it}'\right) = L$.

Assumption 1.a requires the parameter spaces to be compact. This implies stationarity of the factor returns. Similarly, we rule out non-stationary covariates and errors in Assumptions 1.b and 1.c. Weak dependence conditions on errors, covariates and group returns are required in Assumptions 1.d and 1.f. Endogeneous covariates would be ruled out, but lagged returns interacted with dummies satisfy these conditions. Assumption 1.e restricts the amount of cross-dependence between error terms. In our representation, the dependence structure of returns should be captured in the group effects, leaving little dependence between error terms, if any. Assumption 1.g requires that the dummies times the lagged returns exhibit sufficient variation over time and across individuals to identify the components of $\boldsymbol{\delta}$. Similarly, Assumption 1.i requires that the lagged returns exhibit sufficient variation over time, for each fund, to identify the fund-specific autocorrelation component. Finally, Assumption 1.h requires the absolute value of the total autocorrelation, for each fund, to be smaller than 1. This condition is necessary to be able to recover residual errors from observed returns, conditionally on the parameters.

Theorem A.1. *Consistency. Under Assumption 1, as T and N tend to infinity, $\hat{\boldsymbol{\delta}} \to^p \boldsymbol{\delta}^0$, for all i, $\hat{\boldsymbol{\theta}}_i \to^p \boldsymbol{\theta}_i^0$ and $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(\hat{\alpha}_{\hat{g}_i,t} - \alpha^0_{g_i^0,t}\right) \to^p 0$.*

Proof. The proof is a variation of the one by Bonhomme and Manresa (2015). We have $\mathbb{E}[R^{obs}_{i,t}] = \eta_i$, so without loss of generality we can work with the demeaned observed returns and assume $\eta_i = 0$. Let us define the function $\hat{Q}(\theta, \eta, \alpha, \gamma)$ as follows:

$$\hat{Q}(\theta, \eta, \alpha, \gamma) = \frac{1}{NT}\sum_{i=1}^N \left(\sum_{t=1}^T \frac{R^{obs}_{i,t} - \sum_{l=1}^L \theta_{i,l,t} R^{obs}_{i,t-l}}{1 - \sum_{l=1}^L \theta_{i,l,t}} - \alpha_{g_i,t}\right)^2.$$

Errors of the true model are given by

$$\epsilon_{i,t} = \frac{R^{obs}_{i,t} - \sum_{l=1}^L \theta^0_{i,l,t} R^{obs}_{i,t-l}}{1 - \sum_{l=1}^L \theta^0_{i,l,t}} - \alpha^0_{g_i,t}.$$

Hence, observed returns can be rewritten as a function of the error terms as follows:

$$R^{obs}_{i,t} = \sum_{l=1}^L \theta^0_{i,l,t} R^{obs}_{i,t-l} + (\epsilon_{i,t} + \alpha^0_{g_i,t})(1 - \sum_{l=1}^L \theta^0_{i,l,t}).$$

This yields the following formulation for the function $\hat{Q}$:

$$\hat{Q}(\theta, \eta, \alpha, \gamma) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(\frac{\sum_{l=1}^L (\theta^0_{i,l,t} - \theta_{i,l,t}) R^{obs}_{i,t-l}}{1 - \sum_{l=1}^L \theta_{i,l,t}} + \epsilon_{i,t} + \alpha^0_{g_i,t} - \alpha_{g_i,t}\right)^2.$$

Define the function $\tilde{Q}(\theta, \eta, \alpha, \gamma)$ as follows:

$$\tilde{Q}(\theta, \eta, \alpha, \gamma) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{i,t}^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{\sum_{l=1}^{L}(\theta_{i,l,t}^0 - \theta_{i,l,t})R_{i,t-l}^{obs}}{1-\sum_{l=1}^{L}\theta_{i,l,t}} + \alpha_{g_i,t}^0 - \alpha_{g_i,t}\right\}^2.$$

The difference between $\hat{Q}$ and $\tilde{Q}$ is given by:

$$\begin{aligned}\hat{Q}(\theta, \eta, \alpha, \gamma) - \tilde{Q}(\theta, \eta, \alpha, \gamma) &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{i,t}\left\{\frac{\sum_{l=1}^{L}(\theta_{i,l,t}^0 - \theta_{i,l,t})R_{i,t-l}^{obs}}{1-\sum_{l=1}^{L}\theta_{i,l,t}} + \alpha_{g_i,t}^0 - \alpha_{g_i,t}\right\} \\ &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{i,t}\left\{\frac{[\mathbf{D}_{i,t}'(\boldsymbol{\delta}^0 - \boldsymbol{\delta})]\mathbf{X}_{i,t}}{1-\sum_{l=1}^{L}\theta_{i,l,t}} + \alpha_{g_i,t}^0 - \alpha_{g_i,t}\right\}\end{aligned}$$

Lemma A.1. *Under the set of assumptions 1, the following uniform convergence result holds:*

$$plim_{N,T\to\infty}\sup_{(\theta,\eta,\alpha,\gamma)}|\hat{Q}(\theta, \eta, \alpha, \gamma) - \tilde{Q}(\theta, \eta, \alpha, \gamma)| = 0.$$

- Convergence of the first term $\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{i,t}\frac{[\mathbf{D}_{i,t}'(\boldsymbol{\delta}^0-\boldsymbol{\delta})]\mathbf{X}_{i,t}}{1-\sum_{l=1}^{L}\theta_{i,l,t}}$

  By Assumption 1.a and because $\mathbf{D}$ contains zeros and ones, $||\frac{(\boldsymbol{\delta}^0-\boldsymbol{\delta})'\mathbf{D}_{i,t}}{1-\sum_{l=1}^{L}\theta_{i,l,t}}||$ is bounded. Apply Cauchy-Schwarz inequality:

$$\begin{aligned}\mathbb{E}\left(\left\|\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}\mathbf{X}_{it}\right\|^2\right) &= \frac{4}{N^2T^2}\mathbb{E}\left(\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}\mathbf{X}_{it}\right\|^2\right) \\ &\leq 4\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\mathbf{X}_{it}\right\|^2\right).\end{aligned}$$

  By Assumption 1.d, there exists a constant M such that:

$$\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(\epsilon_{it}\epsilon_{is}\mathbf{X}_{it}'\mathbf{X}_{is}\right)\right| = \left|\mathbb{E}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\epsilon_{it}\epsilon_{is}\mathbf{X}_{it}'\mathbf{X}_{is}\right)\right| \leq M.$$

Therefore,

$$\left| \mathbb{E} \left(\frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{t=1}^{T} \epsilon_{it} \mathbf{X}_{it} \right\|^2 \right) \right| \leq M.$$

Hence,

$$\mathbb{E} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{it} \mathbf{X}_{it} \right\|^2 \right) \leq \frac{M}{T}.$$

Therefore the first term converges towards zero as N and T go to infinity.

- Convergence of the second term $\frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}(\alpha^0_{g_i,t} - \alpha_{g_i,t})$

We have

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{it} \alpha_{g_i,t} = \sum_{g=1}^{G} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \alpha_{g,t} \right] = \sum_{g=1}^{G} \left[\frac{1}{T} \sum_{t=1}^{T} \alpha_{g,t} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \right) \right]$$

By Cauchy-Schwarz inequality,

$$\begin{aligned} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{it} \alpha_{g_i,t} \right]^2 &= \left[\sum_{g=1}^{G} \left[\frac{1}{T} \sum_{t=1}^{T} \alpha_{g,t} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \right) \right] \right]^2 \\ &\leq G \sum_{g=1}^{G} \left[\frac{1}{T} \sum_{t=1}^{T} \alpha_{g,t} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \right) \right]^2. \end{aligned}$$

Furthermore,

$$\left(\frac{1}{T} \sum_{t=1}^{T} \alpha_{g,t} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \right) \right)^2 \leq \frac{1}{T} \sum_{t=1}^{T} \alpha^2_{g,t} \times \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{g_i=g\}} \epsilon_{it} \right)^2$$

By Assumption 1.a, $\frac{1}{T} \sum_{t=1}^{T} \alpha^2_{g,t}$ is uniformly bounded. The second term of the product is such that:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{g_i=g\}}\epsilon_{it}\right)^2 &= \frac{1}{N^2T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}_{\{g_i=g\}}\mathbf{1}_{\{g_j=g\}}\epsilon_{it}\epsilon_{jt} \\
&\leq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}\right|.
\end{aligned}
$$

$$
= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}\right| = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\epsilon_{it}\epsilon_{jt}] + (\epsilon_{it}\epsilon_{jt} - \mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|
$$

Using the triangular inequality,

$$
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}\right| \leq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\epsilon_{it}\epsilon_{jt}]\right| + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt} - \mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|
$$

Assumption e gives that the first term is lower than $\frac{M}{N}$. We therefore have:

$$
\begin{aligned}
\mathbb{E}\left(\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}\alpha_{g_i,t}\right]^2\right) &\leq G\mathbb{E}\left(\sum_{g=1}^{G}\left[\frac{1}{T}\sum_{t=1}^{T}\alpha_{g,t}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{g_i=g\}}\epsilon_{it}\right)\right]^2\right) \\
&\leq G\sum_{g=1}^{G}\mathbb{E}\left(\underbrace{\frac{1}{T}\sum_{t=1}^{T}\alpha_{g,t}^2}_{\leq M} \times \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{g_i=g\}}\epsilon_{it}\right)^2}_{\leq \frac{M}{N} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt} - \mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|}\right) \\
&\leq G^2\mathbb{E}\left(M \times \left[\frac{M}{N} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt} - \mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right]\right) \\
&\leq \frac{M}{N} + \mathbb{E}\left(\frac{G^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt} - \mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right)
\end{aligned}
$$

The second part of the right hand-side can be rewritten as

$$\mathbb{E}\left(\frac{G^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right)=\frac{G^2}{N^2}\sqrt{\left(\mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right)\right)^2}$$
$$\leq\frac{G^2}{N^2}\sqrt{\mathbb{E}\left(\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right)^2\right)}$$

Furthermore,

$$\frac{G^2}{N^2}\sqrt{\mathbb{E}\left(\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right|\right)^2\right)}\leq\frac{G^2}{N^2}\sqrt{\mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])\right)^2\right)}$$
$$=\frac{G^2}{N^2T}\sqrt{\mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])(\epsilon_{is}\epsilon_{js}-\mathbb{E}[\epsilon_{is}\epsilon_{js}])\right)}$$
$$=\frac{G^2}{N^2T}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left((\epsilon_{it}\epsilon_{jt}-\mathbb{E}[\epsilon_{it}\epsilon_{jt}])(\epsilon_{is}\epsilon_{js}-\mathbb{E}[\epsilon_{is}\epsilon_{js}])\right)}$$
$$=\frac{G^2}{N^2T}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}Cov\left(\epsilon_{it}\epsilon_{jt},\epsilon_{is}\epsilon_{js}\right)}$$
$$\leq\frac{G^2}{N^2T}\sqrt{MN^2T}$$
$$\leq\frac{\tilde{M}}{N\sqrt{T}}$$

where the second to last line uses Assumption f. This concludes the proof of Lemma A.1.

Lemma A.2 shows that $\tilde{Q}$ is uniquely minimized at true values

Lemma A.2. *For all* (θ,α,γ),

$$\tilde{Q}(\theta,\eta,\alpha,\gamma)-\tilde{Q}(\theta^0,\eta^0,\alpha^0,\gamma^0)\geq\hat{\rho}\|\theta-\theta^0\|^2.$$

where $plim_{N,T\to\infty}\hat{\rho}=\rho$.

$$\tilde{Q}(\theta,\eta,\alpha,\gamma)-\tilde{Q}(\theta^0,\eta^0,\alpha^0,\gamma^0)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{\sum_{l=1}^{L}(\theta^0_{i,l,t}-\theta_{i,l,t})R^{obs}_{i,t-l}}{1-\sum_{l=1}^{L}\theta_{i,l,t}}+\alpha^0_{g^0_i,t}-\alpha_{g_i,t}\right\}^2.$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{[\mathbf{D}'_{i,t}(\boldsymbol{\delta}^0-\boldsymbol{\delta})]\mathbf{X}_{i,t}}{1-\sum_{l=1}^{L}\theta_{i,l,t}}+\alpha^0_{g^0_i,t}-\alpha_{g_i,t}\right\}^2.$$

By assumption a, for all i,t, $\frac{1}{1-\sum_{l=1}^{L}\theta_{i,l,t}}\geq M$.

For $d\in\{1,...,D\}$, define $\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t}$ as the average of $\mathbf{X}_{i,t}$, for i and t such that $\mathbf{D}_{i,t}=1$, over $g^0_i\wedge g_i$. Let us denote by $\boldsymbol{\delta}_d$ the d^{th} row of $\boldsymbol{\delta}$.

$$\begin{aligned}\tilde{Q}(\theta,\eta,\alpha,\gamma)-\tilde{Q}(\theta^0,\eta^0,\alpha^0,\gamma^0)\geq\ &\frac{1}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{M[(\boldsymbol{\delta}^0_d-\boldsymbol{\delta}_d)](\mathbf{X}_{i,t}-\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t})+M[(\boldsymbol{\delta}^0_d-\boldsymbol{\delta}_d)]\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t}\\&+\alpha^0_{g^0_i,t}-\alpha_{g_i,t}\Big\}^2.\\=\ &\frac{1}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{M[(\boldsymbol{\delta}^0_d-\boldsymbol{\delta}_d)](\mathbf{X}_{i,t}-\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t})\Big\}^2+\\&\frac{1}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{\alpha^0_{g^0_i,t}-\alpha_{g_i,t}+M[(\boldsymbol{\delta}^0_d-\boldsymbol{\delta}_d)]\bar{\mathbf{X}}_{g^0_i\wedge g_i,t}\Big\}^2\end{aligned}$$

as

$$\frac{2}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{(\mathbf{X}_{i,t}-\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t})\Big\}\Big\{\alpha^0_{g^0_i,t}-\alpha_{g_i,t}\Big\}=0,$$

$$\frac{2}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{(\mathbf{X}_{i,t}-\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t})\Big\}\bar{\mathbf{X}}^d_{g^0_i\wedge g_i,t}=0.$$

Finally,

$$\tilde{Q}(\theta,\eta,\alpha,\gamma)-\tilde{Q}(\theta^0,\eta^0,\alpha^0,\gamma^0)\geq \frac{1}{NT}\sum_{d=1}^{D}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{M[(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d)](\mathbf{X}_{i,t}-\bar{\mathbf{X}}^d_{g_i^0\wedge g_i,t})\right\}^2.$$

Let us define, for each d and grouping γ, the matrix $\Sigma(\gamma)$:

$$\Sigma(\gamma)^d=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{X}_{it}-\bar{\mathbf{X}}^d_{g_i^0\wedge g_i,t}\right)\left(\mathbf{X}_{it}-\bar{\mathbf{X}}^d_{g_i^0\wedge g_i,t}\right)'.$$

Then we have

$$\begin{aligned}\tilde{Q}(\theta,\eta,\alpha,\gamma)-\tilde{Q}(\theta^0,\eta^0,\alpha^0,\gamma^0)&\geq M^2\sum_{d=1}^{D}(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d)'\Sigma(\gamma)^d(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d).\\ &\geq M^2\sum_{d=1}^{D}\min_{\gamma}(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d)'\Sigma(\gamma)^d(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d).\\ &\geq M^2\sum_{d=1}^{D}\min_{\gamma}\hat{\rho}^d(\gamma)\|(\boldsymbol{\delta}_d^0-\boldsymbol{\delta}_d)\|^2 \qquad \geq M^2\|(\boldsymbol{\delta}^0-\boldsymbol{\delta})\|^2\sum_{d=1}^{D}\min_{\gamma}\hat{\rho}^d(\gamma)\end{aligned}$$

The rest of the proof follows as in Bonhomme and Manresa (2015).

□

B Algorithm

We estimate model (1)-(3) using starting values for the groups each fund belongs to. We test two sets of initial values for the initial grouping. The first one assigns a group to each fund randomly. The second assigns it depending on the asset class of the fund. Initial trajectories of the group returns are calculated by minimizing the sum of square errors over time and over members of each group. Let us denote the initial value of θ, g_i and α_{g_i} by $\theta^{(0)}$, $g_i^{(0)}$ and $\alpha_{g_i}^{(0)}$. The first step consists to update all estimates conditionally on the initial values of θ and α. Set $k=0$.

Assignment step: Given $\theta^{(k)}, \alpha_{g_i}^{(k)}$, find for each fund the group that makes its sum of square errors minimal:

$$g_i^{(k+1)} = \underset{g\in\{1,\ldots,G}{\arg\min} \sum_{t=1}^{T} \left(R_{i,t} - \theta^{(k)}.X_{i,t} - \alpha_{g,t}^{(k)}\right)^2.$$

Update step: Update the estimates of $\alpha_{g_i,t}$ based on the new estimate for g_i, by minimizing the total sum of squared errors across funds:

$$\left(\theta^{(k+1)}, \alpha^{(k+1)}\right) = \underset{(\alpha)\in\Theta\times\mathcal{A}^{GT}}{\arg\min} \sum_{i=1} \sum_{t=1}^{N} \left(R_{i,t} - \theta X_{i,t} - \alpha_{g_i^{(k+1)},t}\right)^2.$$

As shown in the literature, simply iterating these two steps often result in a local minimum, which depends on the starting point. We use the variable neighborhood search proposed by Hansen, Mladenovic, and Perez (2010) to increase the probability of finding a global optimum.

Neighborhood jump: Randomly select n funds and put them in n randomly selected groups. A new grouping is obtained. Going through the update step yields estimates of θ and α for that grouping, and iterating an assignment and update step results in a new grouping. If the resulting objective function increases when using $g_i^{(k+1)}$, then iterate using the new grouping.

The number of groups is chosen by minimizing the following Bayesian Information Criterion (BIC):

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

where $\hat{\sigma}^2$ is an estimate of the variance of ϵ_{it} using a large number of groups.

Convergence to a global optimum: The k-means algorithm which is used to allocate funds to private market factors, has been shown to suffer from the risk of ending in a local minimum, usually close to the initial values chosen. As a consequence, the estimated groupings would depend on the starting values and the resulting trajectories of group effects would not be accurate. In order to check that it is not the case in our analysis, we try two different initial groupings of funds. The first one randomly assigns a group to each fund, while the second one starts from the Burgiss categories. We find that the initialization procedure only has a very marginal impact on the estimated groups and trajectories. In some cases, pre-assigning funds to groups according to the Burgiss categories allows achieving a lower mean square error of the residuals, however, it is not always the case. Using a Variable Neighborhood Search therefore proves to be efficient in finding a global optimum.

Simulations: We study the suitability of our algorithm using Monte Carlo simulations. We create three groups of 500 funds each. Each fund has a time-series of 40 quarterly returns following the dynamics of model (1)-(3). Our set of covariates contains lagged returns. The autocorrelation coefficient is set equal to 0.20 during the first half of the time period and 0.10 during the second half. We perform three case studies. In the first case, group returns are generated by three independent standard Brownian motions with volatility parameters 10%, 20% and 30% (per quarter). In the second case, the second group return is correlated with the first one, and the third one with the second one (but not with the first one). Correlation coefficients are set to 0.5 and volatilities to 0.3. In the third case, the three group returns are driven by a single standard Brownian motion with volatility 20%. Their volatilities are respectively 10%, 20% and 30%. Residuals follow a normal distribution centered about 0, with volatility ranging from 0.05 to 0.3. We expect group returns to be more difficult to retrieve correctly in the second case than in the first case and in the third case.

It appears to be seems difficult to find the true number of groups in the simulated data. In the easiest case (case two), the BIC criterion reaches its optimal for three groups in 88% of the simulations. When group returns are perfectly correlated, this number drops to 55%. However, the model parameters (thetas) as well as the trajectories of the group returns are well estimated, even when the estimated number of groups is larger than the true number. The estimates of the autocorrelation coefficients always have a bias smaller than 0.003. The MSEs of group returns are small: about 0.02 when the volatility of errors is 10% and increasing to 0.15 when the volatility of errors reaches 0.4. The error made on estimated group effects increases when the group effects are more correlated, in line with intuition.

C Specification of the autocorrelation function

Two types of dummies were considered in the specification of the autocorrelation function:

1. Dummies linked to characteristics of the funds: (i) dummy indicating whether the fund is

in its investment period, (ii) dummies indicating if the vintage year was before 2000, (iii) dummy indicating whether the vintage year was after 2007, (iv) dummy indicating whether the observed return was extremely positive (> 15%), (v) dummy indicating whether the observed return was extremely negative, (vi) one dummy per level 2 geographical area of funds, (vii) one dummy per level 2 asset class of funds, (viii) dummy indicating whether the fund is the first managed by its fund manager and dummy indicating whether it is the third or more, (ix) dummies indicating whether the fund is a small, large or mega fund.

2. Economic dummies: (i) dummy for NBER recession periods, (ii) dummy indicating when the market (S&P 500) had a return below one standard deviation of its mean, (iii) dummy indicating whether the market had a return above one standard deviation of its mean, (iii) dummy indicating the last quarter of the year (to capture end of year effects).

These dummies are crossed with lagged (observed) returns.

D Related Literature

Our paper builds on three strands of literature. The first strand of literature relates to the study of private equity funds as investment vehicles. A set of papers compares fund cash flows to what listed equity (usually proxied by the S&P 500 index) would have generated (Kaplan and Schoar (2005), Ljungqvist and Richardson (2003), Phalippou and Gottschalg (2009), Harris, Jenkinson, and Kaplan (2014), Robinson and Sensoy (2016)).[33] Given the type of companies private equity funds target, some studies have challenged PE funds as simply a different vehicle for factor exposures that could be obtained more directly (but with limited capacity in terms of total amount of investible capital) via the public securities markets (Phalippou (2014), Stafford (2017)).

Another set of papers implicitly or explicitly specify models for the data generating process of fund returns and search for model parameters that are most consistent with the dynamics of the cash flows (Cochrane (2005), Korteweg and Sorensen (2010), Korteweg and Nagel (2016), Ang, Chen, Goetzmann, and Phalippou (2018)). The drawback of such approaches is that they require a large

[33] As no risk adjustments are made, the debate has concentrated on which benchmark is most appropriate. See **??** for a review of the literature on private equity fund performance.

sub-set of funds to have the same risk exposures and these sub-sets need to be ex-ante defined. For example, an assumption would be that all large buyout funds have the same loadings on some pre-defined risk factors.

In this paper, the grouping of funds by risk profile is endogenous. What allows us to do this is the use of quarterly NAVs to compute a time-series of individual fund returns. As a result, the risk can be assessed at the fund level rather than per sub-set of funds. As discussed in the introduction, we have access to precisely reported NAVs for each fund. These NAVs, however, are not necessarily accurately marked-to-market; an issue we discuss at length below.

The second strand of literature relates to portfolio choice with an illiquid financial asset such as a private market fund. Sorensen, Wang, and Yang (2014) use a dynamic portfolio choice model to value the cost of illiquidity and management compensation in private equity. Longstaff (2009), Ang, Papanikolaou, and Westerfield (2014), among others, develop a parsimonious model of portfolio choice with a single illiquid asset, which is freely tradable at certain points in time but no trade is permitted at other times. Recently, Dimmock, Wang, and Yang (2018) have one alternative asset which becomes fully liquid at maturity (e.g., when a private equity fund is dissolved) but can be liquidated prior to maturity by paying a proportional cost (e.g., selling a private equity fund at a discount on the secondary market). Bollen and Sensoy (2016) incorporate real options insight and develop a model for valuing illiquid private equity when secondary markets exist.

Our paper is closest to the framework used by papers modeling illiquidity due to transaction costs (from Constantinides (1986) to Buss, Uppal, and Vilkov (2016)). In these models, the illiquid asset is always tradable but at a cost. Given the development of secondary markets for private market funds, this approach has both the benefit of being tractable, accommodate multiple private market fund portfolios, and being realistic.

The third strand of literature relates to basis assets (a.k.a. style analysis); and, more broadly, asset pricing factors. This extensive literature aims at grouping assets into portfolios so as to maximize return homogeneity within portfolios and return heterogeneity across portfolios. Our paper is the first to undertake such an exercise for Alternative Assets.

Sharpe (1992) introduces a quantitative methodology for mutual fund style analysis. He uses monthly returns to explain mutual fund performance by funds' exposures to a set of passive benchmarks. Mutual funds are then characterized by an ex-post positive weight portfolio of benchmarks.[34] The main advantage of this type of method is that it provides an intuitive way to explain the dispersion in the returns of a large cross-section of assets using a much smaller number of factors. However, it usually comes at the cost of having to decide ex-ante on a relevant set of factors (e.g. size, book-to-market ratio).

An alternative approach is to statistically identify clusters using the distance between certain individual asset return characteristics. Brown and Goetzmann (1997) choose to minimize the difference in observed mean returns within a group, whereas Ahn, Conrad, and Dittmar (2009) minimize the correlation between the time series of stock returns.

A related approach assumes that each time-series is drawn from two or more data generating processes. The objective is then to find the distribution that is closest to each time-series. These clustering kernels are typically based on dynamic regression models (Frühwirth-Schnatter and Kaufmann (2008), Juárez and Steel (2010)), multivariate normal distributions (McNicholas and Murphy (2010)), or GARCH models (Bauwens and Rombouts (2007)). These models need to restrict the relationship between unobserved heterogeneity and observed covariates, and require that the covariates be independent from the group effects. In addition, they are based on strict assumptions about the distribution of the error term. As an alternative to describe cross-sectional relationships in panel data, model-based methods using finite mixture models have been extended from static to dynamic setups with some success. See, for a review, Frühwirth-Schnatter (2006).

In comparison to these models, the model of Bonhomme and Manresa (2015), upon which we build, allows to leave the structure of the group effects unspecified. In addition, it allows for more flexible time and spatial correlation in the error term than in alternative methodologies.

Most of the literature on time series clustering extracts some specific features from the time series and applies a static clustering scheme to them, see, e.g., Liao (2005). Such methods suffer from the risk that a relevant feature be forgotten from the analysis.

[34] Di Bartolomeo and Witkowski (1997) use Sharpe's procedure to argue that many mutual funds are misclassified. Dor and Jagannathan (2002) show that the Sharpe methodology depends crucially on the selection of benchmarks. These benchmarks continue to evolve as asset pricing research identifies factors that best explain cross-sectional differences in fund returns (Fama and French (1993, 2015)).

The features of private equity returns, however, restrict the use of these methods. First, funds returns exhibit non-trivial autocorrelation, which if not handled properly may result in spurious cross-correlations. Second, each fund lives over a specific period of time, hence pairwise correlations would ignore non-overlapping observations. This issue would be amplified by missing values in our dataset. To overcome these challenges, we use a state-of-the-art econometric technique to model unobserved heterogeneity in panel data.

Statistical methods of classification aim to partition N variables into K disjoint and non-empty subsets so that within-group-object similarity and between-group-object dissimilarity are maximized. The most popular algorithm is called k-means and was developed by Sebestyen (1962), MacQueen (1967). Various variations to the basic k-means algorithm have been proposed: partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods; see Steinley (2006) for a review. These methods have been widely used for statistical analysis in many fields, including pattern recognition, image analysis, bio-informatics and computer graphics.

These methods were initially designed to cluster static data points, and are not trivial to extend when adding a time-series dimension and dealing with panels of data. One may assume that each time series is an entity that belongs to one cluster, and that all time series within the same cluster have the same data generating process. The main issue is then to assign time series to their corresponding cluster. Most methods developed in the literature have attempted to reduce the problem to a static one by extracting specific features from the time series of data, see Liao (2005). The standard algorithms are then applied to these features and, once the mapping from time series to clusters is established, the data generating process of each group can be estimated. The downside of this method is that if some relevant features are omitted in the classification step, time series will be wrongly assigned which will in turn bias the model parameter estimation.

These doubts around NAV accuracy explain why the private equity literature has focused on performance measures that are based on the Net Present Value of fund cash flows, with the discount rate being the benchmark index, e.g. the S&P 500 (Kaplan and Schoar (2005), Robinson and Sensoy (2016), Korteweg and Nagel (2016)), and has stayed away from time-series analysis and portfolio optimization exercises (see Sorensen, Wang, and Yang (2014) for an exception).

E Data cleaning

The original dataset contains 4499 funds. We eliminate funds with vintage year prior to 1984 and after 2012 (69 funds), funds with nothing paid-out (2 funds), funds with size smaller than 10 million USD (12 funds) and funds that Burgiss categorized as *Unknown* or *Not Elsewhere Classified* (92 funds). This first selection leaves a total of 4324 funds in the database, with average size slightly smaller than the original dataset (677 million USD and 661 million USD, respectively). Next, we eliminate NAV that are less than 10% of the fund size and NAV after 15 years of fund life. Quarterly returns that are larger than the 99% percentile or smaller than the 1% percentile are removed, and we only keep funds for which at least 12 pairs of consecutive returns could be calculated. These filters eliminate 162 funds. If a return is below -50% or above 100% we consider it as missing.

Table 10 summarizes the impact of each filter on the number of funds, returns and cash flows that remain in the database. Selected funds have similar characteristics to those of the original sample (non-tabulated).

[Table 10 about here.]

F Spanning of private factors by public factors

[Table 11 about here.]

Table 1. Asset Allocation of US Endowments

This table presents the portfolio allocation of US endowments across broad asset classes, as of June 30, 2018. Other investments (e.g. cash) bring the total to 100%. Source: NACUBO-TIAA survey.

Total Endowment Size	Domestic Equities	Fixed Income	Non-US Equities	Hedge Funds	Private Markets
	%	%	%	%	%
Over $1 billion	13	7	19	19	32
$501 million to $1 billion	22	10	22	18	19
$251 million to $500 million	24	12	22	18	19
$101 million to $250 million	31	15	22	12	11
$51 million to $100 million	34	19	22	10	10
$25 million to $50 million	39	22	18	8	6
Under $25 million	45	24	15	6	4
Dollar-weighted average	16	8	20	18	28
Equal-weighted average	31	16	21	13	12

Table 2. Distribution of funds across categories and regions of investment focus

This table presents the distribution of funds across the three tier classifications defined by Burgiss. For tier-3 classification, however, we design it using a combination of tier-3 characteristics provided by Burgiss. Distribution is also broken down by geographical areas of investment focus. Western Europe separates into the UK and the rest of Western Europe (WE-exUK). Asia separates into China and the rest of Asia (*Asia-ex China*). Other developed countries (*Other dvlp*) include Canada and Middle East countries. RoW refers to the rest of the world.

Categorization			USA	Western Europe		Asia		Other regions		
Tier 1	*Tier 2*	*Tier 3*	USA	UK	WE-exUK	China	Asia-exChina	Other dvlp	RoW	All regions
Debt	Mezzanine	Mezzanine	155	4	12	1	5	5	10	192
	Distressed	Distressed	80	1	8	1	6	4	51	151
	Gen Debt	Gen Debt	38	3	3	0	0	0	15	59
Real Assets	Real Estate	RE Opport	116	7	20	11	36	2	43	235
		RE v-add	221	4	11	0	12	0	8	256
		Gen RE	113	7	14	1	3	2	14	154
		RE Other	43	2	5	0	1	4	8	63
	Nat. Resources	Energy	87	0	1	0	0	19	2	109
		Non-e NatRes	49	0	1	1	2	9	25	87
	Infrastructure	Infra Spe	34	5	7	0	4	0	8	58
		Gen Infra	11	1	2	0	4	1	15	34
	Generalist RA	Gen RA	13	0	0	0	1	4	1	19
Equity	Venture Capital	Early IT	397	2	11	13	6	10	6	445
		Early Other	336	9	23	13	13	9	20	423
		Gen VC	92	1	5	12	11	6	2	129
		VC other	390	4	16	19	16	11	31	487
	Buyout	BO consumer	67	8	19	1	4	5	6	110
		BO industrial	66	4	16	0	1	3	2	92
		BO other spe	524	52	181	14	48	36	65	920
		Gen BO	189	13	26	2	7	12	30	279
	Expansion	Expansion	51	0	2	17	15	1	10	96
	Gen Equity	Gen Equity	226	10	24	15	26	10	54	365
Generalist	Generalist	Generalist	124	0	10	0	13	3	40	190
		All types	3422	137	417	121	234	156	466	4953

Table 3. Model selection

Panel A shows the covariates selected in the specification of the autocorrelation function by the elastic net procedure. Panel B shows measures of model goodness of fit for an increasing number of clusters. The second column shows the Bayesian Information Criterion (BIC). The third column shows the Mean Squared Errors (MSE), i.e. the variable we minimize (objective function). Next, we report statistics based on the private factor attached to each cluster. The fourth column shows the average pairwise correlation between the time-series of each factor. The fifth and sixth column show the average autocorrelation of the factor return time-series.

Panel A. Selected Covariates of the Autocorrelation Function

	Coefficient	t-stat
$R^{obs}_{i,t-1}$	0.012	2.59
$R^{obs}_{i,t-1} \cdot \mathbf{1}(R^{obs}_{i,t-1} < -15\%)$	0.124	5.17
$R^{obs}_{i,t-1} \cdot \mathbf{1}(\text{Recession})$	0.112	10.31
$R^{obs}_{i,t-1} \cdot \mathbf{1}(\text{Experienced firm})$	0.043	8.22
$R^{obs}_{i,t-2}$	0.019	2.94
$R^{obs}_{i,t-2} \cdot \mathbf{1}(R^{obs}_{i,t-1} > 15\%)$	0.105	4.95
$R^{obs}_{i,t-2} \cdot \mathbf{1}(USA)$	0.040	4.81

Panel B. Selection of the number of clusters

			Corresponding factor returns – Average		
	BIC	MSE	Pairwise correlation	AR(1)	AR(2)
4 clusters	1.720	1.265	29%	0.26	0.13
6 clusters	1.684	1.223	25%	0.18	0.10
8 clusters	1.653	1.196	**24%**	**0.14**	**0.09**
10 clusters	1.643	1.171	26%	0.14	0.10
12 clusters	1.624	1.156	29%	0.14	0.10
14 clusters	1.608	1.134	29%	0.15	0.09
16 clusters	1.621	1.132	29%	0.13	0.09

Table 4. Mapping of funds to clusters

This table reports the percentage of a given type of funds allocated to a given cluster. Fund types are formed by combining the three tier level fund classifications of Burgiss with geographical focus, size quartile (vintage year adjusted) and firm experience. The last column reports the Herfindhal index for each type of funds across the eight clusters. Cells are shaded when the fraction is more than one third, and when the Herfindhal index is above 0.25.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	HERF
Generalist		7	3	11	14	10	7	31	17	0.18
Debt		7	5	6	19	4	3	45	11	0.26
Gen Debt		10	3	8	17	0	3	49	8	0.30
Mezzanine		7	7	7	4	5	3	59	8	0.37
Mezzanine	US	1	7	9	3	5	3	64	8	0.43
	WE	50	0	0	6	6	0	25	13	0.34
Distressed		5	3	4	40	4	4	25	15	0.25
Distressed	Small	6	0	6	0	11	11	50	17	0.31
	Non-Small	5	4	4	45	3	3	21	15	*0.28*
Real Assets		12	3	25	6	13	5	28	7	0.19
Real Estate		12	3	33	5	6	4	29	6	0.22
Real Estate	WE	61	3	13	7	0	4	4	7	0.41
	US	5	3	41	2	5	4	35	5	0.30
Real Estate US	Small	4	5	26	3	5	6	43	7	0
	Non-Small	6	2	44	2	5	3	33	5	0
Nat. Resources		6	2	8	9	38	4	28	7	0.24
Nat. Resources	Energy	4	3	4	2	61	3	17	7	0.41
	Timber	12	0	12	10	0	8	58	2	0.38
Infrastructure		20	8	7	9	8	13	28	9	0.17
Infrastructure	WE	67	0	7	0	0	13	7	7	0.48
	US	2	13	7	7	9	13	42	7	0.24
Generalist RA		11	0	0	5	42	5	5	32	0.30
Equity		14	16	10	11	7	10	21	12	0.14
Venture Capital		7	28	10	7	6	9	22	11	0.17
Buyout		21	5	11	14	7	11	20	12	0.15
Expansion		7	6	8	17	8	15	20	19	0.15
Gen Equity		14	11	8	15	7	9	21	15	0.14
Buyout	WE	52	3	6	15	7	7	6	5	0.32
	US	10	6	13	10	7	13	27	14	0.15
Venture Capital	WE	28	17	8	8	4	6	15	13	0.17
	US	6	30	10	6	5	9	23	11	0.18
	Asia	10	21	14	12	3	12	23	6	0.16
Venture Capital	Small	8	18	13	7	8	10	25	11	0.15
	Non-Small	6	34	9	6	4	9	21	11	0.20
Venture Capital	Early IT	6	33	10	6	4	8	24	9	0.20
	Early Other	8	26	13	5	8	8	22	10	0.16
	Gen VC	10	23	9	7	6	11	22	12	0.15
	VC other	6	25	9	8	4	12	22	13	0.16
VC Early IT	Small	8	23	11	5	6	8	29	8	0.18
	Non-Small	5	39	10	6	3	7	21	9	0.23

Table 5. Characterization of the eight estimated clusters of funds

This table presents the characteristics which have a positive and significant loading in the multinomial logit regression, with their exponential loading, from the highest to the lowest. If a characteristic has an exponential loading of n, it means that funds with this characteristic have a relative probability being in the given cluster that is n times larger than funds which do not have this characteristic.

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Europe	6.11	VC	10.07	US RE	3.00	US Distressed	2.69
BO	3.56	Gen equity	2.66	RE	1.93	Distressed	2.05
Infra	3.10					Mega	1.77
Gen equity	2.86						
RE	2.51						

Cluster 5		Cluster 6		Cluster 7		Cluster 8	
RE generalist	10.49	Asia Mega	3.67	Mezzanine	2.94	Europe Mezzanine	6.17
NatRes	8.76	Infra	2.75	US BO	1.75	Europe VC	4.22
		Asia	1.58	US	1.70	Europe RE	3.93

Table 6. Summary statistics of the eight private factors

Panel A reports, for each private factor, the percentage of funds that provide exposure to the factor (% $Nfund$), the market capitalization of these funds (% *Capitalization*), the volatility of the factor time series, the total mean square error (MSE) of funds that form the factor and the first-lag autocorrelation of the factor time series. Panel B displays the correlation matrix of the eight private factors.

Panel A. Summary statistics of the eight private factors

	Funds exposed to factor		Factor trajectories		
Factors	% $Nfund$	% *Capitalization*	Volatility	MSE	AR(1)
1	12%	17%	22%	1.27	0.13
2	12%	6%	26%	1.73	0.48
3	13%	10%	15%	1.07	0.13
4	11%	18%	22%	1.21	0.28
5	8%	7%	18%	1.46	0.14
6	8%	6%	21%	1.40	-0.04
7	25%	21%	9%	0.76	0.26
8	11%	13%	24%	1.21	-0.17
	100%	100%			

Panel B. Correlation matrix of the eight private factors

Correlation	1	2	3	4	5	6	7	8
1	1	0.35	0.41	0.50	0.28	0.27	0.23	0.38
2	0.35	1	0.23	0.28	0.08	0.42	0.49	0.39
3	0.41	0.23	1	0.24	0.11	0.13	0.05	0.15
4	0.50	0.28	0.24	1	0.29	0.19	0.08	0.31
5	0.28	0.08	0.11	0.29	1	0.18	0.06	0.21
6	0.27	0.42	0.13	0.19	0.18	1	0.10	0.33
7	0.23	0.49	0.05	0.08	0.06	0.10	1	0.15
8	0.38	0.39	0.15	0.31	0.21	0.33	0.15	1

Table 7. Private factors vs. public factors

This table reports the results from regressions of the excess return for the eight private factors on the Domestic AQR model, which refers to the Fama-French 3-factor model augmented with the Quality Minus Junk factor of Asness, Frazzini, and Pedersen (2018) and the Betting Against Beta factors of Frazzini and Pedersen (2014).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Panel D. Domestic AQR								
Rm-Rf	0.616	0.288	0.168	0.510	0.113	0.256	0.032	0.284
	6.127	*2.439*	*1.990*	*6.935*	*1.133*	*2.334*	*0.660*	*3.397*
SMB	-0.264	0.161	-0.309	-0.179	-0.070	0.193	-0.124	0.238
	-1.751	*0.910*	*-2.453*	*-1.627*	*-0.466*	*1.173*	*-1.705*	*1.902*
HMLd	-0.161	-0.705	-0.127	0.161	-0.009	-0.250	-0.144	-0.088
	-1.773	*-6.603*	*-1.672*	*2.418*	*-0.101*	*-2.519*	*-3.275*	*-1.163*
QMJ	-0.015	-0.385	-0.114	-0.345	-0.357	-0.029	-0.107	-0.230
	-0.086	*-1.893*	*-0.787*	*-2.725*	*-2.074*	*-0.152*	*-1.283*	*-1.599*
BAB	0.328	-0.234	0.238	0.498	0.228	-0.137	-0.037	-0.013
	3.304	*-2.011*	*2.868*	*6.866*	*2.313*	*-1.268*	*-0.772*	*-0.156*
R^2	48%	53%	19%	72%	21%	25%	18%	47%

Table 8. Betas of private factors on other factors

This table reports the loadings of our eight private factors on macroeconomic variables, alternative indices related to private markets and alternative stock indices.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
LPX Europe	0.23 ***	0.39 ***	0.08 **	0.38 ***	0.12 ***	0.17 ***	0.05 ***	0.24 ***
LPX VC	0.14 ***	0.33 ***	0.03	0.18 ***	0.05 *	0.11 ***	0.04 ***	0.15 ***
FTSE REITs	0.18 ***	0.14 *	0.10 ***	0.39 ***	0.12 **	0.11 *	0.02	0.20 ***
T.Rowe High Yield	0.55 ***	0.40 **	0.15	1.15 ***	0.50 ***	0.31 **	0.03	0.57 ***
MSCI NatRes	0.38 ***	0.13 **	0.10 *	0.54 ***	0.43 ***	0.16 **	0.01	0.24 ***
SPG Infra	0.63 ***	0.28 ***	0.15 *	0.80 ***	0.39 ***	0.23 **	0.03	0.34 ***
BB Commodity	0.19 ***	0.07	0.05	0.27 ***	0.35 ***	0.12 *	0.00	0.15 *
LPX BO	0.18 ***	0.27 ***	0.08 **	0.39 ***	0.11 **	0.12 **	0.03	0.21 ***
DFA value	0.28 ***	0.26 ***	0.07 ***	0.40 ***	0.17 ***	0.19 ***	0.01	0.36 ***
CA PE	0.97 ***	1.15 ***	0.39 ***	0.74 ***	0.43 ***	0.67 ***	0.23 ***	0.80 ***
CA VC	0.19 ***	0.65 ***	0.07 *	0.11 **	0.04	0.20 ***	0.13 ***	0.19 ***
Inflation	0.20	0.16	0.44	1.95 ***	2.58 ***	-0.02	-0.23	-0.14
Indus prod growth	1.80 ***	1.65 ***	1.25 ***	1.00 **	0.88 **	1.37 ***	0.60 ***	1.16 **
Credit spread	-4.35 ***	-5.39 ***	-4.89 ***	-2.05	-2.58 **	-2.82 *	-1.89 ***	-3.65 **

Table 9. Private factor premia

This table reports the annualized risk premia and Newey-West t-statistics of our private factors, calculated using Fama MacBeth two-pass regressions.

	Risk premia	*t-stat*
Factor 1	4.5%	*1.99*
Factor 2	3.4%	*1.15*
Factor 3	0.2%	*0.14*
Factor 4	1.9%	*0.79*
Factor 5	-1.0%	*-0.58*
Factor 6	3.9%	*2.01*
Factor 7	0.6%	*0.70*
Factor 8	3.0%	*1.51*

Table 10. Filters applied to the data

This table reports the impact of the different filters applied to the data to the number of funds used in the analysis (# funds), returns (# returns) and cash flows (# cash flows). Filters on NAV are the following: NAV that are less than 10% of the fund size are removed, so are NAV after 15 years of fund life. Limits applied to returns are the 1% and 99% percentiles of the return distribution, equal to, respectively, -0.40 and 0.69. Funds considered "with enough returns" are funds that have at least 12 pairs of consecutive returns.

	# funds	# returns	# cash flows
Initial database	5457		248'545
Filters on fund characteristics			
1984 $\geq$ Vintage $\geq$ 2013	5385		246'261
Sum dividends > 0	5381		246'216
Size > 10m	5356		245'429
Asset class $\neq$ unknown	5297		243'116
Sub asset class $\neq$ unknown	5280		242'260
Asset class $\neq$ not elsewhere	5211		239'300
Sub asset class $\neq$ not elsewhere	5185		238'668
Total returns		217'712	
Filters on NAV and returns			
Filters on NAV		177'331	235'745
Returns within limits		175'217	
Filtered dataset[35]	**4953**	**173'700**	**228'970**

Table 11. Private factors vs. public factors

This table reports the results from regressions of the excess return for the four private factors on common asset pricing models. We examine the following models: Global and Domestic Fama-French 5-factor models, and Global AQR model, which refers to the Fama-French 3-factor models augmented with the Quality Minus Junk factor of Asness, Frazzini, and Pedersen (2018) and the Betting Against Beta factors of Frazzini and Pedersen (2014).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Panel A. Global Fama-French								
Rm-Rf	0.541	0.242	0.139	0.590	0.231	0.163	0.007	0.242
	5.743	*2.269*	*1.804*	*8.210*	*2.609*	*1.699*	*0.159*	*3.547*
SMB	0.013	0.071	-0.262	0.234	0.169	-0.067	-0.127	-0.055
	0.071	*0.329*	*-1.689*	*1.619*	*0.946*	*-0.349*	*-1.479*	*-0.401*
HML	0.021	-0.591	0.250	0.564	0.261	-0.021	-0.062	0.300
	0.108	*-2.712*	*1.590*	*3.845*	*1.443*	*-0.106*	*-0.713*	*2.160*
RMW	0.251	-0.443	0.096	0.566	0.148	-0.333	-0.033	-0.604
	0.998	*-1.553*	*0.468*	*2.951*	*0.626*	*-1.299*	*-0.294*	*-3.322*
CMA	-0.163	-0.346	-0.229	-0.732	-0.306	-0.351	-0.148	-0.580
	-0.669	*-1.252*	*-1.151*	*-3.934*	*-1.331*	*-1.411*	*-1.339*	*-3.289*
R^2	40%	49%	11%	64%	18%	24%	17%	53%
Panel B. Domestic Fama-French								
Rm-Rf	0.497	0.234	0.173	0.570	0.267	0.246	0.033	0.343
	5.058	*2.224*	*2.337*	*6.506*	*3.008*	*2.627*	*0.782*	*5.354*
SMB	-0.098	0.093	-0.218	0.056	0.161	0.132	-0.080	0.181
	-0.671	*0.594*	*-1.979*	*0.429*	*1.220*	*0.948*	*-1.259*	*1.901*
HML	-0.066	-0.329	0.243	0.261	0.024	-0.087	-0.075	0.167
	-0.426	*-1.985*	*2.083*	*1.892*	*0.172*	*-0.591*	*-1.122*	*1.656*
RMW	0.075	-0.637	-0.017	0.076	0.215	-0.066	-0.085	-0.235
	0.440	*-3.482*	*-0.132*	*0.501*	*1.391*	*-0.405*	*-1.153*	*-2.108*
CMA	-0.082	-0.032	-0.156	-0.177	-0.127	-0.230	-0.010	-0.355
	-0.370	*-0.136*	*-0.937*	*-0.892*	*-0.632*	*-1.088*	*-0.100*	*-2.459*
R^2	31%	48%	13%	44%	13%	24%	13%	57%
Panel C. Global AQR								
Rm-Rf	0.361	0.392	0.099	0.404	0.057	0.256	0.053	0.399
	2.963	*2.905*	*1.070*	*4.628*	*0.509*	*2.148*	*0.960*	*4.538*
SMB	-0.313	0.388	-0.444	-0.041	-0.071	0.177	-0.141	0.203
	-1.336	*1.495*	*-2.502*	*-0.245*	*-0.332*	*0.770*	*-1.339*	*1.197*
HMLd	-0.181	-0.518	-0.125	0.113	-0.091	-0.218	-0.094	-0.071
	-2.065	*-5.327*	*-1.874*	*1.790*	*-1.130*	*-2.535*	*-2.391*	*-1.125*
QMJ	-0.495	-0.255	-0.341	-0.516	-0.551	-0.141	-0.076	-0.189
	-1.730	*-0.804*	*-1.575*	*-2.512*	*-2.108*	*-0.501*	*-0.587*	*-0.916*
BAB	0.213	-0.197	0.216	0.384	0.151	-0.094	-0.030	0.032
	2.567	*-2.140*	*3.433*	*6.436*	*1.986*	*-1.152*	*-0.802*	*0.526*
R^2	37%	49%	20%	67%	20%	27%	14%	52%

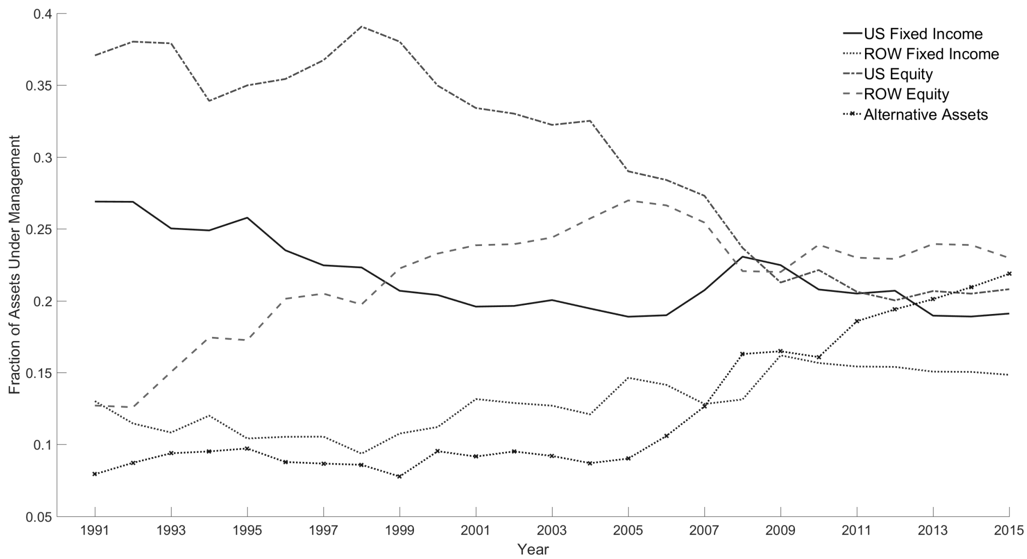

Figure 1. Asset allocation of pension funds. *Source: CEM Benchmarking*

Figure 2. Eight private factors

These graphs represent the time series of each of the eight private factors that have been estimated using our methodology. Each fund's unsmoothed returns are described by the model $R_{i,t}^{unsm} = \eta_i + \alpha_{g_i,t} + \epsilon_{i,t}$. The private factors are the latent group returns $\alpha_{g_i,t}$ of each cluster of funds g_i. Trajectories of factors are compared to the Cambridge Associates Private Equity (CA PE) index returns.