Q-Group Application for the Jack Treynor Prize, 2019

Title: **How Alternative Are Private Markets**?

Applicants: *William N. Goetzmann (Yale), Elise Gourier (Queen Mary), Ludovic Phalippou (Oxford)*

To the Committee: Thank-you for the invitation to submit a paper for the Jack Treynor Prize. We believe that our current project on private equity is well-aligned with the intent of the prize. It is a quantitative approach to breaking down the returns to private capital investment in a way that will help institutional asset allocation decisions. It also fits with the Q-Group's mission of exploring quantitative methods to advance the investment process. T

he paper adapts econometric tools used for "messy" panel data to address some of the problems of working with private capital data -- for example valuation smoothing, non-linear returns and other challenges. It part of an on-going collaboration by the authors to deepen our understanding of the nature of private capital investment. We are excited about the results in the working paper, and intend to expand upon them over the next several months. We believe the final paper will be of broad interest to Q-group members.

What follows is a brief executive summary and a longer description based on extracts from the working paper with informal commentary to Q-group committee members already familiar with some of the methods and issues.

Executive Summary

Question: Does private capital provide returns not spanned by traded factors. If so, what are they?

Background: Analysis of private capital performance to date (including that by the current authors) relies on using factors comprised of publicly traded assets. This paper develops new methods of analyzing private capital returns.

Approach: We extract latent variables from complex, non-linear unbalanced, smoothed private capital panel return data from Burgiss. We begin by estimating the space of private capital returns, rather than beginning from the space of publicly traded assets. We use advanced econometric methods related to clustering to reduce private capital returns to a set of positive-weight factor portfolios comprised of a time-varying set of funds. Then we examine the characteristics of these as well as their covariates.

Relevance: Private capital is a booming area of the institutional investor space. Asset allocation requires understanding of covariance, expected return, non-linearities and other characteristics of private capital. Back-tested Sharpe ratios suggest that knowing ex ante fund characteristics can help asset allocation/diversification decisions.

Results: Reducing dimensionality of private capital returns to four "basis" portfolios results in groupings that do not perfectly match stylistic classifications. Instead, both style and size matter to differences in performance.

- The resulting factors are dominated by (1) large buyout funds, (2) Large VC funds, (3) Real Estate and (4) Large Distressed. Other categories such as infrastructure, and smaller funds are distributed among these big four. We perform the analysis with seven bases as well.

- For three of four basis portfolios, we reject the hypothesis that the returns of are spanned by traded asset returns – the real estate related asset being the exception.

- We find that the same three have an historic positive risk premium, with the first basis being statistically significant.

- We use back-tests to show that that allocations that include these basis portfolios improved the efficient frontier, and that big funds are more useful to diversification than small funds.

- We find a dynamic allocation strategy across basis portfolios would have done even better.

Overview (from working paper introduction)

The recent growth in institutional interest in private market funds [PMFs] has motivated a need to understand their risk, return and covariance relationships to other investments. One challenge to any top-down asset allocation approach, however, is that PMFs are primarily an ownership structure with a broad range of industry exposures, compensation schemes and financing terms. PMFs are not passive investment vehicles, but instead pursue strategies intended to profit from the purchase, active management and resale of companies. This activity is by its nature opportunistic.

The returns of some of the assets held by PMFs may be explained by factors that also explain returns of publicly traded assets, but not necessarily. In practice, an important motivation for investment in PMFs is the belief that they provide exposure to factors that are not accessible via the public capital markets. For example, PMFs are cited as a source of an illiquidity premium that accrues to investors with low liquidity needs. Although there is some evidence that some

PMFs have significant exposure to an illiquidity factor, factor proxies have been constructed from traded securities, by necessity, and the nature of the exposure to this factor may be complex.1 PMFs are also vehicles by which innovative types of investments are funded (e.g. early-stage companies, energy exploration, impact funds, crypto-currencies, direct lending). These funds may provide exposure to priced risk factors but require a governance structure that may be incompatible with public listing. Similarly, Kaplan and Stromberg (2009) find that some PMF returns vary with the equity-fixed income yield spread {a pattern consistent with an imperfect integration between the debt and equity markets. By virtue of timing these opportunities, PMFs may be exploiting a Baker, Greenwood and Wurgler (2003) type of sentiment factor. Finally, PMFs have conditional payoff patterns (Metrick and Yasuda (2009)), and their dynamic strategies may result in correlated non-linear factor exposures (Fung and Hsieh (2001)). In order to understand the potential benefits of PMFs to investors it is necessary to understand whether they deliver performance that is not spanned by standard traded factors (Stafford (2017a)).

Methodology

For readers familiar with the basics of cluster analysis, the general approach is similar. The goal is to group funds into a few basis categories that minimize within-group squared (and scaled) return differences and maximize cross-group differences. Prior academic work has done this for stocks, hedge funds and mutual funds. However, this is not so easy for non-tradable assets private capital returns are not neat and are based on periodic appraisals. The problem demands robust and relatively new tools. Think about this as K-means controlling for auto-correlation, heteroskedasticity and an unbalanced panel.

We develop a dynamic model for returns to cluster PMFs according to a statistical criterion, and thereby form a parsimonious set of factors. We refer to these private market factors as Alternative Basis Assets (ABAs), building on methodology by Brown and Goetzmann (1997) and the terminology of Ahn, Conrad, and Dittmar (2009).

Grouping PMFs into portfolios is difficult due to a partly voluntary and time-varying smoothing of NAVs which may depend on fund characteristics (e.g. fund age), and the time period. In addition, the data form a highly unbalanced panel (funds start and end at different points in time), and there are missing observations. To capture the factor structure of our data while addressing these characteristics, we extend the Grouped Fixed Effect model of Bonhomme and Manresa (2015). The underlying idea is that the time series in a panel can be grouped such that all members of a group have a common time-varying component. This group Fixed-effect should explain a large part of the cross-sectional correlation structure of the fund return time series.

In our context, one can think of fund returns being driven by a group effect stemming from the factor exposure of investments made by the fund. This amounts to assuming that funds can be grouped into several styles of investments that present similar factor exposures. To estimate our model, we follow the estimation method devised by Bonhomme and Manresa (2015). Parameter estimates are obtained by minimizing the sum of square idiosyncratic terms in both the cross-sectional and time series dimensions. The estimation is an iterative procedure similar in spirit to K-means methods. We perform extensive simulations to verify robustness.

Results

Reducing dimensionality to four "basis" portfolios [ABA for "Alternative Basis Assets"]results in groupings that do not perfectly match stylistic classifications. Instead, both style and size matter to differences in performance.

Factor Portfolio Composition

- **ABA 1** is mainly a large Buyout fund portfolio. Of interest, a large fraction of infrastructure funds is also assigned to this ABA, which supports the conjecture that many infrastructure funds are buyout funds in disguise.20 Few VC funds are assigned to this ABA.
- **ABA 2** is dominated by VC funds, and we do not observe a particular size tilt. This group also contains many expansion capital funds; these funds have an investment strategy and target companies that are similar to those of late stage VC funds. This remark holds particularly true for the smaller expansion funds, and this is exactly what we observe: it is the smallest expansion capital funds that are in the second ABA alongside VC funds. Unsurprisingly, there are few Real Asset funds in this ABA, which confirms the belief that Venture Capital and Real Assets are very distinct investments in terms of return dynamics.
- **ABA 3** is dominated by Real Estate funds. Each of the other sub-classes have between 17% and 30% of their funds in that ABA. We therefore do not see particular clusters of other funds here. This may reflect the fact that many PMFs - irrespective of their focus - own companies with significant real estate holdings and with profits that are particularly sensitive to real estate prices. One may interpret this real estate cluster as an indication that most real estate funds are real given their distinctive return dynamics.
- **ABA 4** is dominated by large Debt funds and large generalist Real Asset funds. There are also many large Buyout funds in this category. As large debt funds and large Buyout funds tend to invest in the same underlying companies with capital claims that are close to one another in the capital structure (junior debt versus preferred and common equity), this result makes intuitive sense. It is also not surprising that very few Venture Capital and Expansion funds are in this group.

Application to Portfolio Choice

In the interest of space, in this section, we do not detail much of the analysis. The tables and detailed discussion are in the paper. In it we document the relation between the ABAs and traded factors and the relation to macroeconomic factors. We test whether the ABA generated risk premia. We reject that they are spanned by traded factors. This is relevant because of recent claims that private capital returns can be replicated by passive, low cost index fund strategies. With respect to the relevance to asset allocation decisions, Figure 2 give a sense of the relevance to portfolio choice.

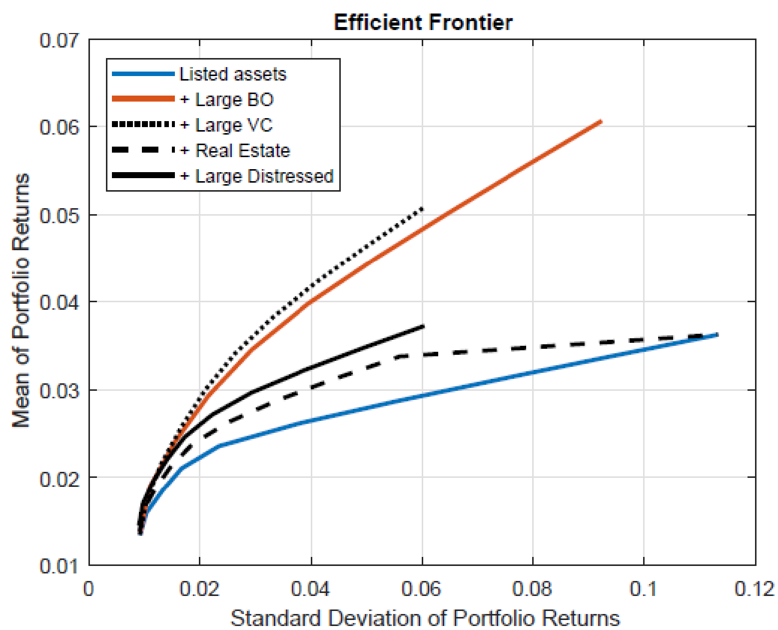

Figure 2. This figure represents the efficient frontiers of basis assets and indices, using quarterly returns over the period 1984-2016. Five portfolios are compared. The first portfolio contains listed products. The four other portfolios contain, in addition to the portfolio of listed products, the Large BO, Venture Capital, Real Estate and Large Debt ABAs.

Proposal Summary

We are excited about the project and about the results to date. The paper is strongly focused on relevance to investment practice. We believe that presentation of the paper to the Q-group will generate very useful feedback for further development of the paper and a more complete exposition of our methods will stimulate Q-group members who may be working on related questions. We look forward to the committee’s decision.

How Alternative Are Private Markets?[1]

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

June 2018

PRELIMINARY AND INCOMPLETE

PLEASE DO NOT QUOTE OR CIRCULATE

Abstract

We estimate a parsimonious set of factor portfolios out of a comprehensive panel of private market funds. We test whether these private market factors are spanned by commonly used public equity factors, and whether they are priced in the cross-section of private market funds. Four main private market factors explain returns in the cross-section; one of them, dominated by large leveraged buyout funds, generated a significant positive premium. Together with the second factor dominated by venture capital funds, it improved an investor Sharpe ratio over the past thirty years. The benefits of investing in factors related to real assets and private debt funds have been smaller but increasing over time.

[1]The authors 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. They thank Pierre Collin-Dufresne, Thierry Foucault, Emmanuel Guerre, Michel Habib, Emmanuel Jurczenko, Kevin Kaiser and Tarun Ramadorai for helpful comments and discussions. They also thank the seminar participants at Bank of Portugal, Bath U., ESSEC, INSEAD, HEC Paris, Liverpool U., Paris-Dauphine, and Zurich U. for their comments, as well as participants to the HF&PE Paris-Dauphine Conference and the ESSEC Amundi Workshop. The authors have consulted to various financial institutions; but 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. Queen Mary University of London and Centre for Economic Policy Research (CEPR), Mile End Road, London E1 4NS, United Kingdom; tel: (+44) 20 7882 8049; Email: e.gourier@qmul.ac.uk.

[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

The recent growth in institutional interest in private market funds [PMFs] has motivated a need to understand their risk, return and covariance relationships to other investments. One challenge to any top-down asset allocation approach, however, is that PMFs are primarily an ownership structure with a broad range of industry exposures, compensation schemes and financing terms. PMFs are not passive investment vehicles, but instead pursue strategies intended to profit from the purchase, active management and resale of companies. This activity is by its nature opportunistic.

The returns of some of the assets held by PMFs may be explained by factors that also explain returns of publicly traded assets, but not necessarily. In practice, an important motivation for investment in PMFs is the belief that they provide exposure to factors that are not accessible via the public capital markets. For example, PMFs are cited as a source of an illiquidity premium that accrues to investors with low liquidity needs. Although there is some evidence that some PMFs have significant exposure to an illiquidity factor, factor proxies have been constructed from traded securities, by necessity, and the nature of the exposure to this factor may be complex.[1]

PMFs are also vehicles by which innovative types of investments are funded (e.g. early-stage companies, energy exploration, impact funds, cryptocurrencies, direct lending). These funds may provide exposure to priced risk factors but require a governance structure that may be incompatible with public listing. Similarly, Kaplan and Stromberg (2009) find that some PMF returns vary with the equity-fixed income yield spread – a pattern consistent with an imperfect integration between the debt and equity markets. By virtue of timing these opportunities, PMFs may be exploiting a Baker, Greenwood and Wurgler (2003) type of sentiment factor. Finally, PMFs have conditional payoff patterns (Metrick and Yasuda (2009)), and their dynamic strategies may result in correlated non-linear factor exposures (Fung and Hsieh (2001)). In order to understand the potential benefits of PMFs to investors it is necessary to understand whether they deliver performance that is not spanned by standard traded factors (Stafford (2017a)).

[1]See Constantinides (1986), Buss, Uppal, and Vilkov (2016), Ang, Chen, Goetzmann, and Phalippou (2018), Sorensen, Wang, and Yang (2014), Longstaff (2009), Ang, Papanikolaou, and Westerfield (2014) and Bollen and Sensoy (2016). Most recent work in this literature is by Dimmock, Wang, and Yang (2018), who solve a portfolio choice problem with one PMF that becomes fully liquid at maturity, but can be liquidated prior to maturity by paying a proportional cost.

In this paper, we use the most comprehensive panel of PMF cash flows and quarterly Net Asset Values (NAVs) available. We construct positive-weighted portfolios of funds which can be interpreted as private market factors, and which we term alternative basis assets [ABAs] for exposition simplicity. Having reduced the dimensionality of the dataset and created time-series of private market factors, we look at the shift in the efficient frontier induced by the presence of each of these factors, and we study the extent to which these factors are priced in the cross-section of PMF returns.

Our first contribution is methodological. To address the peculiarities of PMF valuation data we integrate a time-series model with lagged dependencies with a partitioning algorithm previously used in economics. Specifically, we introduce a model for PMF returns, which endogenously groups into fund portfolios that have similar return dynamics while capturing the characteristics of PMF returns: time-varying autocorrelation, different life spans, and missing observations. The model has one set of covariates that address the issue of NAV smoothing due to conservative updating. Everything else in the model, including potential exposure to public market factors, is left unspecified and captured by a group time-varying fixed effect. In short, we generate a time-series of return per group conditioning only on information on past returns.

To determine the parametric specification of the covariates, hence the nature of potential NAV smoothing, we study the autocorrelation of fund returns and find that the first two lags are significant and that the autocorrelation is higher at the time of follow-on fund fundraising, and at the fourth lag when it is the last quarter in the year. We select these four variables as covariates and the model jointly estimates the loadings on each covariate and the time-varying group fixed effects, i.e. the time-series of factor returns. This approach presents two major advantages compared to standard factor models: (i) Factors do not need to be specified ex-ante and (ii) we do not impose any linear structure at this stage.

Our second contribution is empirical. The current version of the paper analyzes the link between the factors obtained and public market factors and finds that ABAs are not well spanned by public market factors. This is consistent with some set of PMFs delivering alternative (i.e. un-spanned) factor exposures. We estimate the price of risk of ABAs by running Fama-MacBeth regressions on individual and portfolios of fund returns, grouped by (a) fund investment objectives, (b) fund size and (c) by fund geographical focus. ABA 1 has a larger price of risk compared to prices of factors obtained by pooling funds by Burgiss class. Furthermore, this premium is statistically significant. The other ABAs are not found to be priced.[2]

In order to quantify the benefits of PMF vehicles for investors, we simulate an investment strategy using realistic costs and backwards-looking information sets. A Sharpe ratio maximizing investor (as well as a power utility investor) would have held a private market portfolio that included 5 percent to 10 percent allocation to factors dominated by large Buyout [BO] and Venture [VC] funds.[3] We show the dynamics of these allocations correspond to past institutional interest in BO and VC in the 1990s and 2000s, while allocations to factors dominated by real estate and large distressed debt funds grew steadily.

The remainder of the paper is organized as follows. In Section 2 we review the institutional setting of private equity, and present our statistical model. Section 3 discusses the data and reports descriptive statistics. Section 4 reports the main empirical results, and section 6 briefly concludes.

2 Building Alternative Basis Assets

2.1 Institutional 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 mainly 60-40 US stock-bond asset allocation to an allocation that is nearly equally spread across five classes of assets: US public equity, non-US public equity, US listed debt (fixed income), non-US listed debt (fixed income), and alternative assets. Within alternatives, private market funds (PMFs) are the main piece and have grown significantly over time to reach assets under management of $5 trillion in 2016, substantially exceeding that of hedge funds ($3 trillion).[4]

[2] In the next version of the paper, we plan to conduct standard tests of asset pricing models and perform out-of-sample cross-validation analysis.

[3] In the next version of the paper we will test for alternative model specifications in which the autocorrelation function is constant, and test robustness to separation of the sample into sub-periods and sub-samples of PE funds.

[4] These statistics are similar for other types of institutional investors. The allocation of university endowments to US listed equity over the period 2006 to 2012 declined from 46% to 32% with a large part of these assets shifting to alternatives (Ang, Ayala, and Goetzmann (2014)).

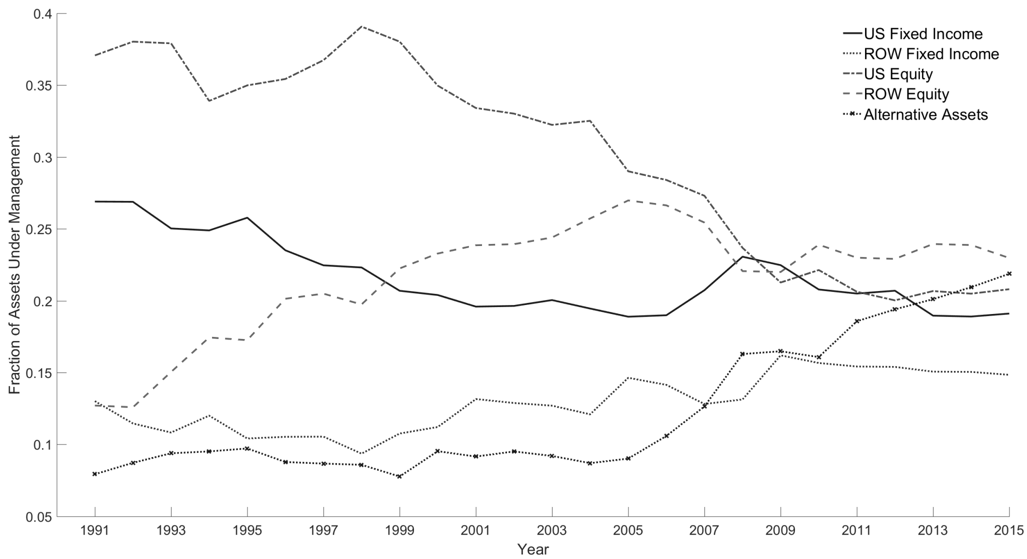

Figure 1. Asset allocation of pension funds. *Source: CEM Benchmarking*

PMFs differ from public market investment vehicles in several ways. Investors commit ex-ante to providing a given amount of capital to a PMF over a fixed period of time (e.g. five years). The quantity, amount and timing of capital calls are uncertain at inception. So are the investments, which do not necessarily match the announced intention at the time of fund raising, or do not clearly belong to a single category.

For each PMF, investors record the daily stream of capital investments and distributions, as well as quarterly NAVs reported by the fund managers. These NAVs are bound to be subjective; they result from judgements about the appropriate valuation technique and input parameters for each portfolio company. The rules that govern the estimation of NAV are SFAS 157 ("Fair Value Measurements") for the U.S. since the final quarter of 2008, and IFRS 13 ("Fair Value Measurement") since 2006 for Europe (Crain and Law (2016)). Fair value accounting has, however, been slowly implemented over time and was in partial use before these official dates.

A fair value assessment of ongoing investments 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)" (see IFRS (2012)). Note that computing a fair value should be easier in some categories than in others. For example, it is easier to mark-to-market the equity claim in the Hilton Hotel LBO than an equity claim in a toll bridge in Chicago.

2.2 Classification based on Investment Objectives

Most institutional investors use some form of top-down approach for asset allocation decisions. A variation of the Markowitz mean-variance framework is either implemented internally or via consultants, and requires as input expected returns, variance and correlation of a set of self-declared asset classes.[5] This approach determines the fraction of the portfolio dedicated to each asset class.

In practice, institutional investors access private markets via specialized fund managers, which they select. To facilitate this process, consultants and data providers categorize PMFs based on a few basic criteria thought to be relevant for assessing their returns, variances and covariances. The three main criteria are: (1) the stage of development of the companies they invest in (e.g. young vs. mature), (2) the seniority of the capital claim (e.g. junior debt vs. common equity) and (3) the industry focus of targeted investments (e.g. natural resources vs. real estate).

Leverage Buy-Out (LBO) is the largest category of PMFs. LBO funds invest in equity claims of mature companies, that are operating outside the three following industries: infrastructure (e.g. airports, utilities), real estate (e.g. shopping malls, hotels), and natural resources (e.g. oil, timber). Investments in these industries are usually LBOs as well but are, instead, assigned their own category because they are perceived as being real assets.

Some combinations of the three dimensions mentioned above are not used in practice. 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, although there are two rising categories called VC debt and RE debt that would be a good fit.

To further stress the point that the distinction between sub-categories is blurry and subjective, consider the LBO of Hilton Hotels; is this strictly a real estate acquisition? It is exposed to more

[5] This exercise is further complicated by the ex-ante uncertainty about the timing and amounts of fund investments and divestments. Investors need a model to convert capital commitments into desired PMF exposures.

business cycle risk than, say, student housing. As it happened, the Hilton LBO was sponsored by both a real estate fund and an LBO fund.[6]

In sum, commonly used classifications make intuitive sense, but are subjective and ad hoc. These classifications lead to a large number of classes, which makes any top-down analysis challenging (e.g. what is the correlation between Mezzanine funds and Real Estate funds?). Therefore it is difficult to relate all these classes to a parsimonious set of factors that will both spread returns and deliver factor premia. This paper seeks to bridge this gap by estimating a factor space within PMF returns that is based on observed PMF return dynamics.

2.3 Model

We develop a dynamic model for returns to cluster PMFs according to a statistical criterion, and thereby form a parsimonious set of factors. We refer to these private market factors as *Alternative Basis Assets* (ABAs), building on the terminology of Ahn, Conrad, and Dittmar (2009).

Grouping PMFs into portfolios is difficult due to a partly voluntary and time-varying smoothing of NAVs which may depend on fund characteristics (e.g. fund age), and the time period. In addition, the data form a highly unbalanced panel (funds start and end at different points in time), and there are missing observations. To capture the factor structure of our data while addressing these characteristics, we extend the Grouped Fixed Effect model of Bonhomme and Manresa (2015). The underlying idea is that the time series in a panel can be grouped such that all members of a group have a common time-varying component. This group fixed-effect should explain a large part of the cross-sectional correlation structure of the fund return time series. In our context, one can think of fund returns being driven by a group effect stemming from the factor exposure of investments made by the fund. This amounts to assuming that funds can be grouped into several styles of investments that present similar factor exposures.

[6] Similarly, LBOs of phone directories have classified as infrastructure even though such companies probably have different risk characteristics than, say, a toll road in New York City.

2.3.1 Specification

Consider N funds indexed by $i \in \{1, ..., N\}$ each belonging to one of G basis assets, indexed by $g \in \{1, ..., G\}$. We denote by $R_{i,t}^{obs}$ (resp. $R_{i,t}^{unsm}$) the observed (resp. unsmoothed) return process between times $t - \Delta$ and t, where $\Delta = 1$ quarter and $t = 1, ..., T$. The mapping of funds to their basis asset is given by $g : i \to j$. The returns of all funds in a given basis asset have a common component, more specifically, a time-varying group effect denoted by $\alpha_{g(i),t} \in \mathcal{A}$ (or, for simplicity, $g(i) = g_i$) and which we label *group return*. It drives the dependence structure of unsmoothed returns in our model. These returns also have an idiosyncratic component $\epsilon_{i,t}$. We model observed returns as a weighted average of past observed returns and contemporaneous unsmoothed returns:

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

$$R_{i,t}^{unsm} = \eta_i + \alpha_{g_i,t} + \epsilon_{i,t}. \tag{2}$$

Equation (1) is an extension of a classical representation of smoothed returns. Geltner (1991), Ross and Zisler (1991), Geltner (1993) used an $AR(1)$ process to represent appraisal smoothing in real estate. Similar models were 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)). Furthermore, we assume that the autocorrelation at lag l can be expressed as the sum of two components. The first component is a fund specific constant, which is the l^{th} component of the vector $\boldsymbol{\theta}_i$ and is in Θ. This component of the autocorrelation function is identified in the time series of returns. The second component is a sum of constants that are common across funds multiplied by d time- and fund-specific dummies. These constants are identified in the cross-section of funds. They form the matrix $\boldsymbol{\delta} \in \Theta^{d \times L}$. Each dummy is collected in the vector $\mathbf{D}_{i,t} \in \mathbb{R}^d$. Let us denote by $\mathbf{X}_{it}$ the vector of lagged returns: $\mathbf{X}_{it} = (R_{i,t-1}, ..., R_{i,t-L})'$. We have:

$$\sum_{l=1}^{L} \theta_{i,l,t} \cdot R_{i,t-l}^{obs} = (\boldsymbol{\theta}_i' + \mathbf{D}_{i,t}' \boldsymbol{\delta}) \mathbf{X}_{it}. \tag{3}$$

Dummy variables can account for fund characteristics (age, size), or market conditions (sign of contemporaneous returns, recession). For example, PMF managers may have stronger incentives to smooth NAVs during recessions. Such a behavior would be captured in the following model:

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

and the null hypothesis mentioned above would be written as $\delta > 0$.

Both the idiosyncratic term $\epsilon_{i,t}$ and the group return $\alpha_{g_i,t}$ are left unspecified. Therefore, our model belongs to the class of latent factor models, but does not suffer from the common disadvantage that one has to specify factors ex-ante, and to impose a linear relation between returns and factors. Instead, the trajectories of group returns are fully estimated as part of the estimation algorithm.[7]

2.3.2 Estimation

To estimate our model, we follow the estimation method devised by Bonhomme and Manresa (2015). Parameter estimates are obtained by minimizing the sum of square idiosyncratic terms in both the cross-sectional and time series dimensions:

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

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

Note that in the absence of autocorrelation ($\theta_{i,t} = 0$), the estimation coincides with the standard minimum sum-of-squares partitioning problem and is equivalent to the correlation-based clustering method of Ahn, Conrad, and Dittmar (2009).

In Appendix A, we present the three sets of conditions that ensure consistency of the estimators, as well as a discussion of these assumptions and of the robustness of the estimates to the presence of i) autocorrelation, ii) cross-correlation between group and idiosyncratic terms, and across idiosyncratic terms, and iii) heteroscedasticity and non-zero higher (than three) order moments in

[7] The assumption that each fund belongs only to a single group may be relaxed methodologically with a more complex mixture model, but it has the virtue of estimation simplicity, and has been shown to work well in applications to publicly traded equities.

the idiosyncratic component.

The minimization problem (4) is solved by alternatively selecting the optimal grouping of funds conditionally on the model parameters and trajectories of the group returns, and computing the optimal parameters and group returns conditionally on the grouping of funds into basis assets. The number of groups is chosen by minimizing the Bayesian Information Criterion (BIC).[8] In addition to this criterion, we analyze the impact of adding a group on the correlation between group returns.

If the estimated number of groups exceeds the true number, the estimator of θ is consistent, however the estimated group effects suffer from a finite T bias, with bias term of order $O(1/\sqrt{T})$. In contrast, if we work with a number of groups that is smaller than the true number, and if the group returns are correlated with the covariates, the estimator of θ becomes inconsistent, due to omitted-variable bias. Estimating the number of groups with the BIC criterion provides an upper bound of the true number of groups when the length of the time series is smaller than the number of funds, which is true in our case.

Details on the algorithm are provided in Appendix C. The accuracy of the estimation is studied using Monte Carlo simulations, and is discussed in Appendix D.

3 Data

3.1 Datasets

We use a comprehensive dataset of private market funds, which is collected and maintained by Burgiss. This dataset contains timed cash flows as well as a time series of net asset values (NAVs) for 4,282 private equity funds as of December 2017. We believe this is the largest dataset of private equity fund cash flows ever analyzed. Moreover, as detailed in Harris, Jenkinson, and Kaplan (2014), the Burgiss data offers many advantages compared to that of other commercial providers, including improved quality of the data. For each fund, the dataset contains the amount of committed capital (a.k.a. *fund size*), the geographical area and currency of investments, as well as the year in which

[8] The BIC penalizes over-fitting using a purely statistical approach. In our context, it may be relevant to also use an economic criterion. For example, we may stop adding portfolios when the increase in the maximum Sharpe Ratio falls below a certain threshold.

a fund first drew down committed capital (a.k.a. *vintage year*).

Consistent with the literature and practitioner reports, both the number and size of funds increase over time to peak in 2007 (non-tabulated). We select funds with a 2010-US-dollar fund size of at least $10 million and with vintage year between 1984 and 2012.[9] Stopping with the 2012 vintage year ensures that most funds in the dataset have finished their investment period.

Recently, Burgiss went through their database and via a labor-intensive process, classified each fund in one of twelve asset classes. In a nutshell, they drew on their extensive experience to determine the nature and number of distinct asset classes they perceived to be present in the data; and decided ex-post on which fund pertains to which class. These asset classes are themselves associated with one of four overarching primary classifications: *Debt*, *Equity*, *Real assets*, *Generalist*.[10]

Returns on mutual funds (e.g. those of Dimensional Fund Advisors and Vanguard) and indices (e.g. FTSE REITS) come from WRDS/CRSP. Macroeconomic variables (credit spread, inflation, industrial production growth,...) come from FRED (Federal Reserve Economic Data). Returns on standard equity factors come from academics' personal websites (Ken French [11], David Hsieh [12]).

3.2 Fund characteristics and descriptive statistics

As funds are spread out across many countries of investment focus, we group them into three regions: U.S., Europe (includes the U.K. and Scandinavia, but excludes eastern Europe), and the rest of the world (includes Canada, Australia, Japan). Similarly, for fund currencies, we assign funds to one of three currencies (USD, Euro, other). As fund size is not comparable over time, for each vintage, we categorize funds into quartiles. Small funds have a fund size that belongs to the first quartile of fund sizes with the same vintage; mid funds, large funds and mega funds have a fund size that is respectively in the second, third and fourth quartile of their vintage year.

Table 1 shows the distribution of funds across the twelve categories, and across regions of investment

[9] Harris, Jenkinson, and Kaplan (2014) point out that 1984 is the first year with a meaningful number of funds.

[10] Burgiss defines as a *Generalist* a fund that invests in two or more categories or invests less than 70% of capital in a single category. We exclude 44 funds classified as *Not elsewhere classified*; these funds belong to an emerging or less meaningful category that is not yet recognized as a discrete category in the taxonomy structure. We also exclude 12 funds classified as *Unknown*; these funds have insufficient information.

[11] Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[12] Available at `http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-FAC.xls`

focus, fund currencies, and fund sizes. One third of the sample is composed of Venture Capital (VC) funds, and another third of the sample is composed of Leveraged Buy-Out (LBO) funds. About half of the rest of the sample is made of Real Estate (RE) funds. Each of the other fund categories represents less than 5% of the funds in the sample. The numbers of observations in these categories vary between 17 (Generalist Real Assets) and 224 (Generalist Equity).

There are notable geographical tilts. More than 80% of Debt funds and VC funds are US-focused, whereas the samples of Infrastructure funds and LBO funds are more spread-out across geographical areas (about one quarter of the funds are focused on Europe and slightly less than one fifth of the funds are focused on the rest of the world). Expansion capital funds (a.k.a. growth capital) are most common in emerging markets.[13] 85% of the funds are in USD even though only 73% of the funds focus on the U.S. Even some Europe-focused funds are denominated in USD. Only 6% of the funds in the sample are in a currency other than USD and Euro. This tilt may reflect the fact that the dataset contributors are mainly located in the U.S. and have a preference for USD funds.

Both RE funds and Natural Resource funds are equally distributed across size categories. Half of infrastructure funds, LBO funds and distressed debt funds are mega funds. In contrast, VC funds tend to be small and mid-sized: only 8% are mega funds.

3.3 Time series of fund returns

To compute the time-series of fund returns, we need to use quarterly NAVs and therefore can only work at quarterly frequency. However, cash flows are recorded daily, which forces us to make an assumption regarding the re-investment and financing policy for these intra-quarter cash flows. Our default approach is simple. We use a Modified Internal Rate of Return (MIRR) to bring all the

[13] In fact, most funds operating in emerging markets are expansion capital funds because it is difficult to either purchase controlling stakes in companies, or use leverage, or both in these geographies. Investments in these geographies are minority equity claims in mature but growing companies, which is what is referred to as expansion capital. These funds tend to be smaller (due to these growing companies being relatively small). Most of these funds, however, are in USD and their returns should, therefore, be sensitive to the change in the value of the USD compared to emerging market currencies.

Table 1. Distribution of funds across styles and characteristics
This table presents the distribution of funds across the twelve private equity styles defined by Burgiss. Column 2 contains the number of funds in each style category (N) and Column 3 shows the number of funds as a percentage of the whole sample ($N\%$). Columns 4 to 13 contain the number of funds within each category as a percentage of the total number of funds in that category. Fund size categories are based on same vintage year fund size quartiles.

			Geographical Focus			Fund Currency			Fund Size			
	N	N %	U.S.	Europe	RoW	USD	EURO	Other	Small	Mid	Large	Mega
Debt												
Mezzanine	141	3	87	9	4	87	9	4	21	23	30	26
Distressed	117	3	79	8	14	97	2	2	11	12	21	56
Generalist	30	1	87	10	3	90	7	3	20	20	27	33
All	288	7	84	8	8	91	6	3	17	18	26	39
Real Assets												
Real Estate	596	14	73	12	15	87	8	5	19	25	32	24
Nat. Resources	174	4	80	1	19	91	0	9	32	25	25	18
Infrastructure	74	2	59	23	18	74	15	11	11	18	22	50
Generalist	17	0	100	0	0	100	0	0	12	53	0	35
All	861	21	74	10	16	87	7	6	20	25	29	26
Equity												
Vent. Capital	1302	31	82	6	12	93	4	3	38	33	21	8
Buyout	1286	31	63	22	14	73	16	11	15	20	25	39
Exp. Capital	72	2	51	3	46	94	4	1	26	28	31	15
Generalist	224	5	67	12	22	86	7	8	28	22	27	24
All	2884	70	72	14	15	83	10	7	27	26	23	23
Generalist												
Generalist	115	3	81	6	13	97	3	0	17	17	25	40
All	4148	100	73	12	14	85	9	6	25	25	25	25

intra quarter cash flows to the end of the quarter using an 0.021% daily compounding rate.[14] The amount we obtain at the end of the quarter is added to the fund NAV and a quarterly return can then be computed using this total end of quarter value and the NAV at the start of the quarter.

PMFs being structured as closed-end funds, NAVs have no direct impact on investors' wealth.[15] However, PMF managers may purposely smooth NAVs with the aim of facilitating investor relationship management (e.g. avoid negative returns).

We use the panel of fund returns to test several models for the structure of the autocorrelation. Results are shown in Table 2. The first specification uses a standard AR(4) model. The first two lags are economically significant and of similar magnitude. The next two lags are half the magnitude of the first two. We thus work with an AR(2) from there on. We note that the economic magnitudes are smaller than one might expect, with an autocorrelation coefficient of 0.1.[16]

Smoothing should, however, depend on fund characteristics and time period. We estimate the same panel regression but with a series of cross-effect on the first two auto-regressive coefficient. The second specification studies the impact of the fund classification. As the average length of the time-series is only of 30 quarters, we use only four dummy variables: BO, VC, RE, and Debt. The coefficients are very close to one another. Buyout funds only have a slightly lower autocorrelation. We conclude that autocorrelation is not classification dependent.

The third specification studies the impact of fund age. We create four categories: less than two years, between two and four, between four and six and above six. For the first lag autocorrelation is the same in each age bracket except for years 2 to 4, which is when the fund is getting fully invested and make some first cash distributions. It is also the time at which the follow-on fund is

[14] The daily rate coincides to an 8% annual rate, which is the usual fund hurdle rate, i.e. the return at which funds start to earn a performance based compensation. All statistics and tests have also been computed with a plain 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 (if we used IRR there would be outliers as some large distributions may happen a few days after the start of the quarter and generate an artificially high IRR).

[15] In open-ended funds, investors can buy and sell their fund stakes at a price equal to the NAV on a daily basis. In contrast, 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 implicit a wealth transfer between buying and selling investors (unlike for open-ended funds).

[16] These magnitudes are smaller to those documented for hedge funds (see Getmansky, Lo, and Makarov (2004)) and are smaller to those observed for aggregate private market indices such as that of Cambridge Associate private equity induces. We are undertaking simulation exercises to understand better why autocorrelation is smaller in this panel estimation setting compared to in a time-series with aggregated returns setting.

being raised. The peak in smoothing may then be the result of fund-raising efforts. At times of fund-raising good and bad news may be more slowly incorporated into NAVs. If the fund managers are doing well, they rather be conservative to avoid negative surprises later on and pumping up returns has little benefits since they are doing well. If the fund managers are not doing well then NAVs are optimistic in order to still be considered as a possible investment option by potential investors.[17]

The fourth specification studies the impact of fund size. Although the coefficients differ from one size category to the other, we do not see a clear pattern. Results are similar with fund region of investment focus in the fifth specification.

Finally, we look at time-dependence in autocorrelation (specification 6). We use as a cross effect the following variables: pre- post- 2006,[18], the contemporaneous return on the S&P 500 index, a dummy variable for NBER-classified recession, and a dummy variable for fourth quarter. The latter dummy variable takes the value one for the last quarter of every year (End of Year) and crossed with the fourth lag in an AR(4); it is added to capture the fact that most NAVs are audited at the end of the year. There are no clear patterns except for the End of Year effect. Funds with a good return in the fourth quarter of a given year are more likely to have again a good fourth quarter the following year. This effect is not present at other horizons or return lags. Some funds may differ in how they adjust the NAV during the year, with some funds updating only at the fourth quarter each year, which means that on average their fourth quarter returns are persistently higher than those of their peers.

From these results, we determine the specification we retain to model autocorrelation: We include two lags, a fourth lag crossed with fourth quarter indicator variable (end-of-year), and the first lag crossed with a dummy variable that is one if the fund is between two and four years old (zero otherwise). We now have selected the Parametric Structure of the Covariates and can estimate the model described above in section 2.3

[17] In the next version, we plan to use a more direct measure of fundraising time. Note however that there is little variation on time elapsed between two subsequent funds, it is almost invariably between two and four years.

[18] Crain and Law (2016) found that after 2008, both LBO funds and VC funds were more likely to update the quarterly valuations of their investments, which indicates that information was incorporated more quickly into valuations.

Table 2. Parametric Structure of the Covariates

This table reports the coefficients resulting from OLS panel regressions with quarterly fund returns as dependent variables. EoY is a dummy variable that is 1 for the last quarter of every year, and zero otherwise.

Autocorrelation coefficients											
(1) AR(4)		(2) AR(2) crossed with objective		(3) AR(2) crossed with fund age		(4) AR(2) crossed with fund size		(5) AR(2) crossed with fund geography		(6) AR(4) crossed with time periods	
R_{t-1}	0.09	$R_{t-1} \times$ ...		$R_{t-1} \times$ ...		$R_{t-1} \times$ ...		$R_{t-1} \times$ ...		$R_{t-1} \times$ ...	
		...BO	0.08	...<2y	0.09	...Small	0.11	...US	0.1	...Pre-2006	0.08
		...VC	0.11	2<.<4y	0.17	...Mid	0.06	...Europe	0.08	...$SPX < -10\%$	0.09
		...RE	0.11	...4<.<6y	0.10	...Large	0.08	...RoW	0.11	...$SPX > 10\%$	0.16
		...Debt	0.11	...>6y	0.07	...Mega	0.13			...NBER recession	0
		...Other	0.10								
R_{t-2}	0.09	$R_{t-2} \times$ ...		$R_{t-2} \times$ ...		$R_{t-2} \times$ ...		$R_{t-2} \times$ ...		$R_{t-2} \times$ ...	
		...BO	0.11	...<2y	0.03	...Small	0.03	...US	0.09	...Pre-2006	0.1
		...VC	0.11	...2<.<4y	0.11	...Mid	0.12	...Europe	0.12	...$SPX < -10\%$	0.12
		...RE	0.12	...4<.<6y	0.1	...Large	0.05	...RoW	0.01	...$SPX > 10\%$	0.02
		...Debt	0.1	...>6y	0.06	...Mega	0.12			...NBER recession	0.03
		...Other	0.09								
R_{t-3}	0.04										
R_{t-4}	0.05									$R_{t-4} \times EoY$	0.1

4 Empirical Results

4.1 Description of Alternative Basis Assets

Our estimation indicates that the optimal number of Alternative Basis Assets (ABAs) is about twenty. Such a large number of ABAs makes any portfolio optimization exercise difficult. However, we observe that although increasing the number of ABAs improves the granularity of the grouping (e.g., correlation between ABAs goes down), it does not alter our main results. For simplicity, we present descriptive statistics for the case with four ABAs, and then show the same descriptive statistics for the case with seven ABAs.[19]

In Table 3, we show how funds in each sub-class are distributed across the four ABAs. In addition, we show the average fund size in each sub-sample (funds belonging to both a given sub-class and a given ABA). The case of Mezzanine funds shows the role played by fund size in the return dynamics. Mezzanine funds are quite evenly spread across all four ABAs: 18% of Mezzanine funds are assigned to the first ABA, 30% are assigned to the second ABA and to the third ABA, and the remaining 23% are assigned to the fourth ABA. However, the Mezzanine fund assigned to the first ABA have an average size of $368 million whereas the Mezzanine fund assigned to the fourth ABA have an average size that is more than twice as much ($810 million).

Other sub-classes tend to be more polarized. Large distressed debt funds tend to be assigned to the fourth ABA (with the large mezzanine funds). Half of Real Estate funds are in the third ABA. Half of Natural Resources funds are in the fourth ABA. Half of Venture Capital funds are in the second ABA and nearly half of Buyout funds are in the first ABA.

To sum up, the first ABA is mainly a large Buyout fund portfolio. Of interest, a large fraction of infrastructure funds is also assigned to this ABA, which supports the conjecture that many infrastructure funds are buyout funds in disguise.[20] Few VC funds are assigned to this ABA.[21]

[19] The BIC statistic decreases only marginally after four and even less so after seven.

[20] Infrastructure is commonly associated with stable cash flows, and often perceived as a substitute for bond investments. As a result, a fund buying companies with high amounts of leverage may find it an easier sale to go under the label infrastructure rather than buyout even if the fund targets companies with a non-negligible business cycle risk.

[21] It would be interesting to know whether these VC funds invest mainly in late stage, i.e. at the mature companies end of the young companies spectrum. We do not have such information, but these VC funds are only slightly larger than the rest of the VC funds, making this conjecture unlikely.

The second ABA is dominated by VC funds, and we do not observe a particular size tilt. This group also contains many expansion capital funds; these funds have an investment strategy and target companies that are similar to those of late stage VC funds. This remark holds particularly true for the smaller expansion funds, and this is exactly what we observe: it is the smallest expansion capital funds that are in the second ABA alongside VC funds. Unsurprisingly, there are few Real Asset funds in this ABA, which confirms the belief that Venture Capital and Real Assets are very distinct investments in terms of return dynamics.

The third ABA is dominated by Real Estate funds. Each of the other sub-classes have between 17% and 30% of their funds in that ABA. We therefore do not see particular clusters of other funds here. This may reflect the fact that many PMFs - irrespective of their focus - own companies with significant real estate holdings and with profits that are particularly sensitive to real estate prices. One may interpret this real estate cluster as an indication that most real estate funds are *real* given their distinctive return dynamics.

The fourth ABA is dominated by large Debt funds and large generalist Real Asset funds. There are also many large Buyout funds in this category. As large debt funds and large Buyout funds tend to invest in the same underlying companies with capital claims that are close to one another in the capital structure (junior debt versus preferred and common equity), this result makes intuitive sense. It is also not surprising that very few Venture Capital and Expansion funds are in this group.

Overall, even though many funds are reclassified, i.e. there is not a perfect mapping between objective-based classifications and ABAs, we do see patterns that depend on both fund objective and fund size. This finding confirms that existing fund classifications, although helpful, may not be a sufficient criterion to parsimoniously capture the different return dynamics, hence diversification benefits, of the different PMFs.

Remarkably, when looking at all the PMFs, we see that they are nearly equally distributed across the four ABAs. We note differences in terms of fund size in each ABA: the first and fourth ABAs contain funds that are on average twice as large as those in the second and third ABAs. The volatility is similar across ABAs although the large Buyout one (ABA 1) and the VC one (ABA 2) have more volatile returns than the other two. The Sharpe Ratios differ widely but higher order moments of the time-series of returns may bias this picture. The large Buyout ABA and the large Debt ABA have the highest Sharpe Ratios. The Venture Capital ABA has a low Sharpe Ratio due to its high volatility. Correlations with stock-market indices are similar across ABAs, although the VC ABA seems to be more correlated with the stock-market index than other ABAs.

Finally, ABA returns are positively correlated with one another, but the average correlation is modest at 51%. The Real Estate ABA and the large Buyout ABA are the most highly correlated; whereas the Real Estate ABA and the Venture Capital ABA are the least correlated.

Table 4 shows the same statistics as Table 3, but setting the number of ABAs to seven. The first four ABAs are similar to the four ABAs we just described. The first four ABAs are basically large Buyout and Infrastructure, Venture Capital and Expansion Capital, Real Estate, and Debt (no longer with a size tilt though).

Interestingly, the rest of the ABAs show the type of funds that were next in line in terms of distinctive return dynamics. The fifth ABA is dominated by Natural Resources and General Real Asset funds. The sixth ABA is a collection of relatively small funds spanning all sub-classes. In particular, the sixth ABA contains the 20% smallest infrastructure funds. Finally, the seventh ABA is dominated by Mezzanine funds.

Increasing the number of ABAs improves the granularity of the decomposition. The average correlation between ABAs, for example, decreases to 41% when we have seven ABAs (compared to 51% with four ABAs). Sharpe Ratios are still varying significantly across ABAs, with the large Buyout ABA having a much higher Sharpe Ratio than the rest of the ABAs despite having a volatility in line with the average of the other ABAs. The correlation with the stock-market index is similar to what we found with four ABAs, except that now the Real Estate ABA and the Debt ABA have correlations that are close to zero with the stock-market index.

4.2 ABA loadings on standard asset pricing factors

In this sub-section, we test whether standard asset pricing models capture the ABAs' return dynamics. We consider linear models from the public equity literature (e.g., 5-factor Fama-French models, augmented with the liquidity factor of Pastor and Stambaugh (2003)). In addition, we use

Table 3. Descriptive Statistics - The case of FOUR Alternative Basis Assets

This table reports descriptive statistics on the funds in each of four Alternative Basis Assets (ABAs) when estimating the model described by equations (1) and (2). We show the fraction of funds in each sub-classes that is assigned to one of four ABAs. Next to each fraction we report the average fund size in the corresponding sub-sample (e.g. Mezzanine funds assigned to ABA 1).

Panel A: Sub-class classification and ABAsassignment

	ABA 1		ABA 2		ABA 3		ABA 4	
	N%	*Size*	*N%*	*Size*	*N%*	*Size*	*N%*	*Size*
Mezzanine	18	368	30	589	30	543	23	810
Distressed	12	1292	11	627	25	886	52	1365
Generalist Debt	20	1178	27	267	30	796	23	848
All Debt	16	766	22	557	28	698	35	1152
Real Estate	17	626	15	487	52	579	16	789
Natural Resources	15	418	12	365	25	479	47	624
Infrastructure	34	1255	17	1316	22	1633	26	1186
Generalist Real Asset	24	1219	6	224	18	298	53	1140
All Real Assets	18	707	14	550	43	612	24	776
Venture Capital	14	289	53	267	18	225	15	222
Buyout	40	1301	18	848	19	821	22	1040
Expansion Capital	28	690	38	344	19	635	15	906
Generalist Equity	32	718	27	503	17	322	23	749
All Equity	27	995	35	420	19	520	19	720
Generalist Private Equity	28	1530	10	1198	31	885	31	1174
All Private Equity funds	25	956	29	448	25	580	21	801

Panel B: ABA return characteristics

		ABA 1	ABA 2	ABA 3	ABA 4
Volatility		3.82	4.18	3.13	2.99
Sharpe Ratio		1.41	0.64	0.96	1.41
Correlation with S&P 500		0.25	0.34	0.22	0.26
Correlation matrix	ABA 1	1.00	0.51	0.63	0.57
	ABA 2	0.51	1.00	0.37	0.49
	ABA 3	0.63	0.37	1.00	0.52
	ABA 4	0.57	0.49	0.52	1.00

Table 4. Distribution of funds across SEVEN Alternative Basis Assets

Same table as the previous one, but with seven ABAs instead of four ABAs.

Panel A: Composition of basis assets

	ABA 1		ABA 2		ABA 3		ABA 4		ABA 5		ABA 6		ABA 7	
	N%	*Size*	*N%*	*Size*	*N%*	*Size*	*N%*	*Size*	*N%*	*Size*	*N%*	*Size*	*N%*	*Size*
Mezzanine	15	435	13	464	15	404	9	1284	8	886	4	352	35	554
Distressed	20	1151	8	1096	8	813	36	1166	11	2046	7	802	10	734
Generalist	3	95	17	1219	20	340	10	1027	17	968	13	309	20	821
All Debt	16	801	11	761	13	501	20	1185	10	1402	6	543	23	609
Real Estate	17	658	10	485	33	692	9	717	10	360	8	469	13	629
Natural Resources	9	543	8	285	16	417	9	700	40	658	5	452	12	286
Infrastructure	24	1314	13	1518	11	1832	11	1110	16	1732	20	857	7	1045
Generalist	6	3004	6	224	0	0	12	706	53	961	18	213	6	2324
All Real Assets	15	748	10	565	27	698	9	753	18	639	9	533	13	595
Venture Capital	14	309	35	251	10	196	10	262	9	243	9	238	11	277
Buyout	30	1146	12	789	11	1143	11	1364	15	854	11	1177	11	953
Expansion Capital	20	735	24	445	5	377	7	354	20	618	12	424	11	948
Generalist	19	723	18	459	6	543	11	447	16	1015	18	527	12	308
All Equity	22	859	23	390	10	672	11	797	13	661	11	701	11	588
Generalist	17	1622	16	1018	16	1372	21	800	17	1604	4	809	9	567
Generalist	28	1530	10	1198	31	885	31	1174						
All PE	20	856	19	439	14	693	11	839	14	726	10	664	12	592

Panel B: Descriptive statistics of group returns for each basis asset

		ABA 1	ABA 2	ABA 3	ABA 4	ABA 5	ABA 6	ABA 7
Volatility		3.74	4.11	3.08	5.55	4.32	4.73	2.98
Sharpe ratio		1.41	0.73	0.62	0.65	0.82	0.47	1.15
Correlation with S&P 500		0.28	0.36	0.07	0.10	0.25	0.34	0.26
Correlation matrix	ABA 1	1.00	0.61	0.51	0.38	0.38	0.60	0.61
	ABA 2	0.61	1.00	0.25	0.22	0.25	0.52	0.50
	ABA 3	0.51	0.25	1.00	0.29	0.42	0.35	0.45
	ABA 4	0.38	0.22	0.29	1.00	0.36	0.47	0.34
	ABA 5	0.38	0.25	0.42	0.36	1.00	0.26	0.43
	ABA 6	0.60	0.52	0.35	0.47	0.26	1.00	0.39
	ABA 7	0.61	0.50	0.45	0.34	0.43	0.39	1.00

the Fung and Hsieh (2001) model designed to capture the option-like return dynamics of Hedge Funds. As Hedge Funds are usually grouped with Private Equity funds under the banner Alternative Assets, it is interesting to study the extent to which our ABAs load on the same risk factors as those shown to capture Hedge Fund return dynamics. Results are shown in Table 5.

Few loadings on the standard equity factors are statistically significant, but the economic magnitudes are not negligible. We note that the *Large Buyout* ABA loads positively on the HML factor, which is consistent with Buyout funds investing into value companies. Similary, the *Venture Capital* ABA loads negatively on the HML factor, confirming a growth-tilt of Venture Capital investment strategies. All the ABAs load negatively on the size factor. This result is surprising given that companies targeted by private equity funds are small in comparison to publicly listed companies.

The five factors of Fung and Hsieh (2001) are the returns obtained by constructing so-called look-back straddles on five different option markets: Bond, Commodity, Currency, Short Term Interest Rate, and Stock Index respectively. All the markets except currency are related to the investment strategy of PE funds. However, as PE funds do not trade dynamically, but instead buy and hold their capital claims, we may not expect many significant loadings. Results confirm this intuition although interest rate strategies seem related to the *large Buyout* ABA and the *Real Estate* ABA.[22] The fact that large Buyout and Real Estate funds use significant amounts of leverage and regularly refinance their underlying investments may explain this relation.

4.3 ABA loadings on macroeconomic factors

We now examine the loading of each ABA on a set of macroeconomic factors. We include only one of these factors at a time (hence, basically capture the correlation between each ABA and each factor). Results are shown in Table 6.

We start with some macroeconomic variables. Inflation is a variable of interest to most institutional investors as they search for good inflation hedges and may anticipate PE funds, in particular Real Assets, to be good hedges against inflation. We find no significant loading on inflation by any of our ABAs.

[22]Note that the loadings are negative but this factor delivers negative returns.

Table 5. Alternative Basis Asset Loadings on Standard Models

This table reports the results from regressions of the excess return for each of the four Alternative Basis Asset on the factors of standard asset pricing models. In Panel A, the factors include the 5 Fama-French factors (Small Minus Big, High Minus Low, Profitability and Investment) plus the liquidity factor of Pastor and Stambaugh (2003). In Panel B, the factors are those Fung and Hsieh (2001). These factors are returns from lookback straddle option trading strategies.

Panel A: Loadings on Fama-French and Pastor-Stambaugh equity factors

	1. Large Buyout	2. Venture Capital	3. Real Estate	4. Large Debt
S&P 500	0.17	0.04	0.02	0.16 (*)
SMB	-0.23	-0.75 (*)	-0.21 (*)	-0.25 (*)
HML	0.22	-0.40	0.22 (*)	-0.03
Profitability	-0.13	0.53	-0.24 (*)	-0.15
Investment	-0.44	-0.36	-0.26	0.03
Liquidity	0.18	-0.26	0.09	0.04
Adjusted R^2	0.07	0.13	0.10	0.10

Panel B: Loadings on Fung-Hsieh Hedge Fund factors

	1. Large Buyout	2. Venture Capital	3. Real Estate	4. Large Debt
Bond option	-0.03	0.06	0.00	0.00
Commodity option	0.03	0.04	0.02	-0.02
Currency option	-0.02	-0.07	-0.01	0.01
Short term interest rate option	-0.03 (*)	-0.06	-0.08 (***)	-0.02
Stock index option	0.13 (*)	-0.36 (*)	0.04	0.04
Adjusted R^2	0.09	0.26	0.33	0.03

Industrial production growth has a significant and large positive effect for all ABAs. This indicates that all PE fund returns are highly sensitive to business cycles.

Loadings on credit spreads are overall negative. The negative sign is intuitive: when credit spread increases, risky credit becomes more expensive, risk premia increases, and as a result existing risk project witness a decrease in value, particularly so if they are financed with debt. It may be surprising that the large Buyout ABA does not load more on credit spread. However the credit spread captures the spread between BAA and AAA rated debt. Large buyout funds are usually more exposed to High Yield debt (a.k.a. junk bonds). When we use T.Rowe high yield returns, we find that large Buyout returns are indeed positively related to High Yield bonds; and are therefore lower when interest rate on high yield debt increases. Also consistent with these observations is that there is no relation between ABA 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 ABA returns (although this is the return on the VIX index that is used here and interpretation may change). By design, PE fund returns may resemble those of options on listed equity. As mentioned above, Venture Capital returns have features similar to those of out-of-the-money equity call option; whereas Buyout returns 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 option. We could include change in VIX. However, when we use returns on Put options and Collar options, we do not see much correlations, except for the positive relation between Collar returns and the large Buyout ABA returns. We may use more option return series in the future.

4.4 ABA risk premiums

Our ABAs are positive-weighted portfolios of funds which can be interpreted as private market factors, as they capture, by construction, the dispersion in time series of PMF returns. We test whether these ABAs are priced in the cross-section of PMFs by running Fama-MacBeth regressions. We first run these regressions at the fund level, but find that no ABA has a statistically significant price of risk. This is not surprising as running regressions at the fund level is subject to substantial noise. Following the asset pricing literature, we group PMFs into portfolios according to (i) their

Table 6. Unconditional Hedging Benefits of Alternative Basis Assets

This table reports the loadings of ABA returns on macro-economic factors and publicly listed assets.

Panel A: Loadings of ABA returns on macroeconomic factors

	1. Large Buyout		2. Venture Capital		3. Real Estate		4. Large Debt	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
Inflation	0.86	0.00	0.34	-0.01	0.29	-0.01	0.78	0.00
Indus. prod. growth	2.22 (**)	0.07	3.34 (**)	0.06	1.68 (***)	0.16	1.97 (***)	0.16
Credit spread	-2.56	0.00	-9.09 (*)	0.04	-6.93 (***)	0.23	-1.40	0.00
VIX	-0.02	0.00	0.01	-0.01	0.00	-0.01	-0.03	0.01

Panel B: Loadings of ABA returns on publicly listed assets

	1. Large Buyout		2. Venture Capital		3. Real Estate		4. Large Debt	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
T.Rowe High Yield	0.55 (**)	0.08	0.09	-0.01	0.10	0.00	0.48 (***)	0.11
T-bond 10 years	0.08	-0.01	-0.49	0.00	-0.01	-0.01	-0.25	0.01
Put	0.33	0.02	0.33	0.00	0.11	0.00	0.21	0.02
Collar	0.40 (*)	0.04	0.25	0.00	0.10	0.00	0.24 (*)	0.04
FTSE REITs	0.12	0.01	-0.22	0.00	0.05	0.00	0.07	0.00
DFA Microcap	0.11	0.01	0.03	-0.01	0.01	-0.01	0.09	0.02
DFA Value	0.07	0.00	-0.07	-0.01	0.03	-0.01	0.07	0.00
LPX VC	0.07	0.01	0.32 (**)	0.08	0.05	0.01	0.09 (*)	0.06
LPX Mezzanine	0.16	0.05	0.05	0.01	0.08	0.01	0.08	0.01
LPX 50	0.13 (*)	0.05	0.34 (**)	0.07	0.07	0.02	0.12 (**)	0.08
BB Commodity	0.27 (**)	0.06	0.29	0.01	0.20 (**)	0.07	0.27 (***)	0.11
MSCI Nat. Res. Index	0.27 (*)	0.10	0.10	0.02	0.17 (*)	0.07	0.24 (**)	0.16
SPG Infra Index	0.42 (**)	0.12	0.19	0.05	0.29 (**)	0.11	0.30 (**)	0.12

investment objective – we consider the 12 Burgiss categories of funds –, (ii) their size quartile and (iii) their geographical focus. Table 7 reports the prices of risk of each of the four ABAs as well as their Newey-West standard error, t-statistic and p-value. These prices of risk are compared to those of factors built by aggregating funds based on whether they are categorized as Debt, Equity, Real asset or Generalist by Burgiss. Only the first ABA, dominated by large buyout funds, is found to be significantly priced in the cross-section of PMFs. It has a (quarterly) risk premium of 3.4%. All other factors considered, ABAs and Burgiss-based factors, have a risk premium that is statistically insignificant.

Table 7. Prices of risk

This table reports the prices of risk of ABAs and of factors obtained by pooling PMFs together based on whether they are categorized as Debt, Equity, Real asset or Generalist by Burgiss. Newey-West standard errors, t-statistics and p-values are provided.

Panel A: ABA prices of risk

	ABA 1	ABA 2	ABA 3	ABA 4
Risk premium	0.034	0.001	-0.031	0.002
Standard error	0.016	0.016	0.019	0.012
t-stat	2.108	0.068	-1.587	0.158
p-value	0.035	0.946	0.112	0.875

Panel B: Prices of risk of factors based on Burgiss primary categories

	Equity	Debt	Real assets	Generalist
Risk premium	0.002	0.016	0.028	0.008
Standard error	0.009	0.012	0.032	0.014
t-stat	0.260	1.363	0.882	0.588
p-value	0.795	0.173	0.378	0.557

5 ABAs and Portfolio Diversification

In this section we first examine the relation between the returns of our basis assets and those of several listed products, with the aim of understanding which listed product may capture part of the risk-return profile of PMFs. We then conduct two types of portfolio allocation exercises to study the diversification benefits of PMFs. The first exercise is static while the second one is dynamic.

5.1 Relation between ABAs and Listed Assets

Our benchmark portfolio contains a collection of listed indices and mutual funds which 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 Buyout 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). Finally we use some widely used indices on Commodity and Natural Resources (whose returns may relate to those of Natural Resources funds) and an Infrastructure index.

5.2 Static portfolio exercise

As outlined in Huberman and Kandel (1987), an investor with a mean-variance utility function is indifferent with respect to holding an additional asset if the mean-variance frontier of the initial set of assets (called benchmark portfolio) coincides with the mean-variance frontier of that benchmark portfolio augmented with the additional asset. In that case, there is mean-variance spanning of the additional asset by the benchmark portfolio.

We gather in the benchmark portfolio all listed products considered in the last section, which were shown to have returns statistically related to those of the basis assets.

To visually evaluate the performance gains coming from PE funds by comparing the mean-variance frontier formed the benchmark portfolio to the frontier of that same portfolio augmented with the basis assets, one by one. Portfolio weights are imposed to be positive.

Substantial portfolio improvements accrue through the inclusion of the *Large Buyout*, *Venture Capital* and *Large Debt* basis assets, as displayed on Figure 2. The maximum Sharpe ratio increases from 1.55 for listed products to 1.75 when including the basis assets. Results are less clear for the *Real Estate* asset.

We complement visual evidence by running mean-variance spanning tests. There is mean-variance spanning if the minimum variance stochastic discount factors that price the benchmark portfolio,

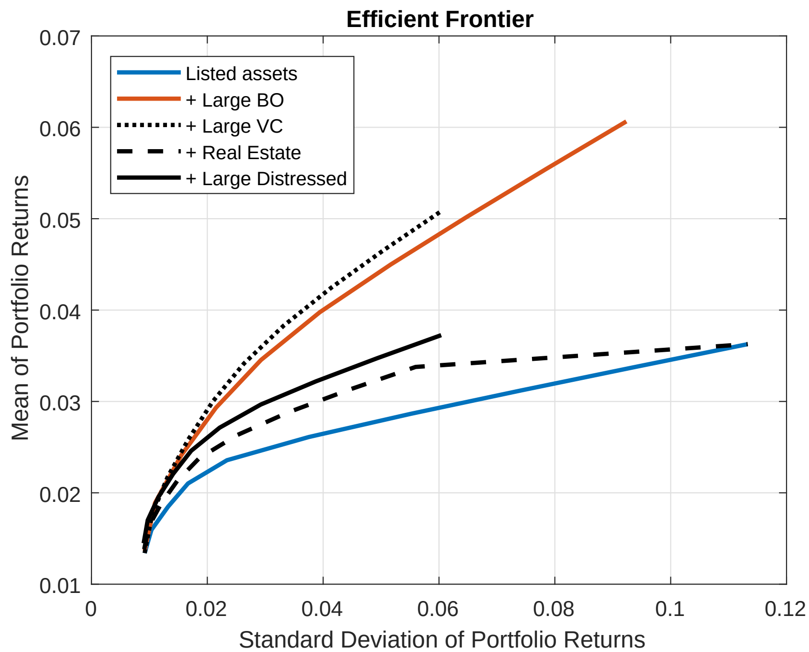

Figure 2. This figure represents the efficient frontiers of basis assets and indices, using quarterly returns over the period 1984-2016. Five portfolios are compared. The first portfolio contains listed products. The four other portfolios contain, in addition to the portfolio of listed products, the Large BO, Venture Capital, Real Estate and Large Debt ABAs.

also price the additional asset (Ferson, Foerster, and Keim (1993), Bekaert and Urias (1996)). Furthermore, the stochastic discount factors associated with mean-variance optimizing behavior have the lowest variance among all admissible stochastic discount factors and are linear in asset returns, Hansen and Jagannathan (1991). Mean-variance spanning can therefore be tested using regression analysis of the additional asset's excess returns on the benchmark portfolio's excess returns. Under the null hypothesis of mean-variance spanning, the intercept of this regression is equal to zero. We calculate the Wald test statistics for the null hypothesis that the benchmark portfolio spans each of the basis assets.[23] This hypothesis is rejected for all basis assets except the *Real Estate* asset, with test statistics reported in Table 8.[24]

To summarize, our analysis reveals that investing in large funds allows achieving better diversification than investing in small funds. This result can be rationalized within the model of Berk and Green (2004) and Chung, Sensoy, Stern, and Weisbach (2012). Talented managers may capture

[23]Test statistics are based on a Newey-West covariance matrix with four lags.

[24]These results still hold when considering a utility maximizing investor with power utility and risk aversion level of 5 (DeRoon, Nijman, and Werker (1996)).

Table 8. Which Alternative Basis Asset is Redundant?
P-value of the hypothesis test: $\mathcal{H}_0$: the returns of basis assets are spanned by the benchmark portfolio of listed assets. Test statistics are based on a Newey-West covariance matrix with four lags.

	4 basis assets		7 basis assets	
Alternative Basis Asset	p-value	hypothesis test	p-value	hypothesis test
1. Large Buyout	0.00	Null rejected	0.00	Null rejected
2. Venture Capital	0.00	Null rejected	0.00	Null rejected
3. Real Estate	0.07	Null not rejected	0.10	Null not rejected
4. Large Debt	0.00	Null rejected	0.00	Null rejected
5. Natural Resources			0.00	Null not rejected
6. Small funds			0.05	Null not rejected
7. Mezzanine funds			0.00	Null rejected

the returns to their skills by growing the size of the fund. In this setup, both the likelihood of raising a follow-on fund and the size of the follow-on fund depend on performance. Assuming decreasing returns-to-scale, Chung, Sensoy, Stern, and Weisbach (2012) find that the indirect pay for performance is of the same order of magnitude as the direct pay coming from carried interest.

However, it may not be possible for investors to take advantage of the added value of large funds. Hochberg, Ljungqvist, and Vissing-Jørgensen (2014) develop a model with information frictions where LPs learn to distinguish skill from luck in a good performing GP. When a new fund is raised, they use their information advantage over other investors to hold up the GP. If inside LPs refuse to invest, outsiders will draw the conclusion that the GP is unskilled and will also not invest. Within such model, larger funds are managed by the most talented GPs, however LPs need to first invest into small funds to hope to have access to such large funds in the future.

5.3 Dynamic portfolio exercise

In order to study how the diversification benefits offered by the basis assets evolved with time, we perform a dynamic portfolio analysis. We consider a mean-variance investor, who, each quarter, forms her optimal portfolio based on the conditional moments of the instruments at hand in the first five years (or, in the beginning of the time series, using all the history of data). We add transaction costs to reflect the differences in the liquidity of listed assets and PE funds. We transaction costs to 2% for listed assets, 5% when buying PE funds and 15% when selling it. The latter number reflects the average cost of selling a PE fund on the secondary market. Figure 3 reports the trajectories of the optimal allocations in each basis asset.

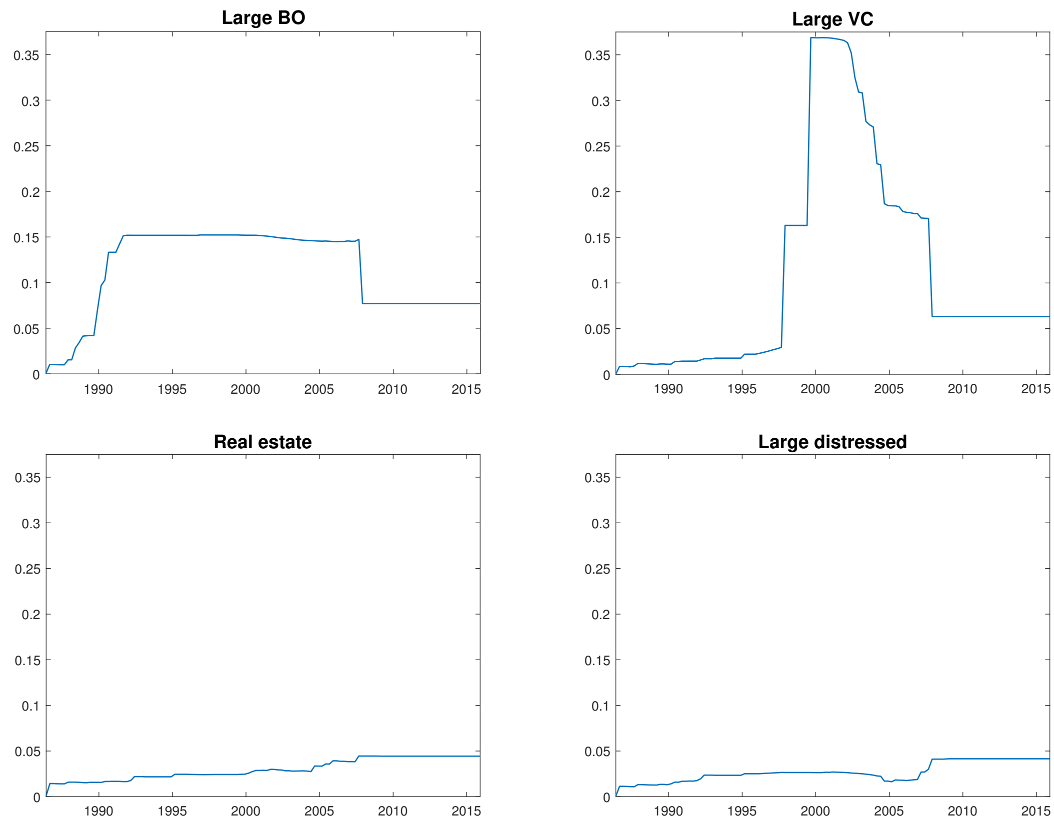

Figure 3. Optimal allocations in basis assets 1 to 4, for a mean-variance investor who forms her optimal portfolio every quarter, based on the conditional moments of all assets at hand in the past five years. Transaction costs are set to 2% for listed assets, 5% for buying PE funds and 15% for selling them.

The benefits of investing in the first basis asset *Large Buyout* are most apparent in the nineties and beginning of the 2000s. During that period, the optimal allocation is around 15% of the portfolio, which is substantially more than allocations in all other assets. For reference, all allocations in listed assets are below 5% in 2016. However, the performance of this basis asset suffered from the financial crisis, and so did the optimal allocation, which then fell to a stable level between 5 and 10%. Similarly, the second basis asset *Venture Capital*, being dominated by VC funds, had booming returns at the end of the nineties, due to the dot-com bubble. The optimal allocation follows this pattern and increases from below 5% to above 35% within 2 years. This period was however followed by a dramatic burst of the bubble in the beginning of the 2000s, which led to the collapse of the optimal allocation in basis asset 2, back to a level below 20%. The financial crisis then triggered a new large decline in the optimal allocation to a level comparable to the allocation in *Large Buyout*, between 5 and 10%. In contrast, the optimal allocations in *Real Estate* and *Large Debt* debt remain small over the entire time series, but steadily increase to reach a level close to 5% in 2016. The latter allocation slightly declined during the financial crisis but recovered swiftly.

6 Conclusion

This paper introduces a new method to recategorize PMFs into portfolios of funds with similar return trajectories. We term these portfolios Alternative Basis Assets [ABAs]. This grouping allows us determine more precisely the characteristics of the PMFs that have delivered on the promise of alternative assets.

We find that the estimated ABAs do not coincide with standard fund categories and that the returns of basis assets are not well spanned by standard listed equity factors. Private and public equity are, therefore, not perfect substitutes.

Using mean-variance analysis, we find that the ABAs that improve most an investor's opportunity set contain mainly large BO funds as well as VC funds. The diversification benefits of real assets are shown to be substantially smaller, although they have improved with time.

References

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

Ang, A., A. Ayala, and W. N. Goetzmann, 2014, "Investment Beliefs of Endowments," Columbia Business School Research Paper No. 13-72.

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

Bauwens, L., and J. V. K. Rombouts, 2007, "Bayesian clustering of many GARCH models," *Econometric Reviews*, 26, 365386.

Bekaert, G., and M. Urias, 1996, "Diversification, Integration, and Emerging Market Closed-End Funds," *Journal of Finance*, 51(3), 835–870.

Berk, J. B., and R. C. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112(6).

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.

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

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.

Chung, J.-W., B. A. Sensoy, L. Stern, and M. S. Weisbach, 2012, "Pay for Performance from Future Fund Flows: The Case of Private Equity," *Review of Financial Studies*, 25(11), 32593304.

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.

DeRoon, F., T. Nijman, and B. Werker, 1996, "Testing for Spanning with Futures Contracts and Nontraded Assets," Working Paper, Tilburg University.

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.

Ferson, W., S. Foerster, and D. Keim, 1993, "General Tests of Latent Variable Models and Mean-Variance Spanning," *Journal of Finance*, 48(1), 131–156.

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

Geltner, D., 1991, "Smoothing in appraisal-based returns," *Journal of Real Estate Finance and Economics*, 4, 327345.

——— , 1993, "Estimating market values from appraised values without assuming an efficient market," *Journal of Real Estate Research*, 8, 325345.

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.

Hansen, L., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(21), 225–262.

Hansen, P., N. Mladenovic, and J. Perez, 2010, "Variable neighbourhood search: methods and applications," *Annals of Operations Research*, 175, 367407.

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

Hochberg, Y. V., A. Ljungqvist, and A. Vissing-Jørgensen, 2014, "Informational Holdup and Performance Persistence in Venture Capital," *Review of Financial Studies*, 27(1), 102152.

Huberman, G., and S. Kandel, 1987, "Mean-variance spanning," *Journal of Finance*, 42, 873–888.

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), 17911823.

Kaplan, S., and P. Stromberg, 2009, "Leveraged Buyouts and Private Equity," *Journal of Economic Perspectives*, 23(1), 121–146.

Kat, H., and C. Brooks, 2002, "The statistical properties of hedge fund index returns and their implications for investors," *Journal of Alternative Investments*, 5, 2644.

Korteweg, A., and S. Nagel, 2016, "Risk-Adjusting the Returns to Venture Capital," *Journal of Finance*, 71(3), 14371470.

Korteweg, A., and M. Sorensen, 2010, "Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies," *Review of Financial Studies*, 23, 37383772.

Liao, T., 2005, "Clustering of Time Series DataA 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, p. 281297. 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, 153168.

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

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., 2017a, "Replicating Private Equity with Value Investing, Homemade Leverage, and Hold-to-Maturity Accounting," .

——— , 2017b, "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.

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. *Let $M > 0$ be some constant.*

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

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

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

d *For all* $g \in \{1, ..., G\}$: $\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(\epsilon_{it}\epsilon_{is}\alpha_{gt}^0\alpha_{gs}^0)\right| \leq M$.

e $\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(v_{it}v_{is}R_{it}R_{is})\right| \leq M$.

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

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

h *Let* $\overline{X \circ D}_{g\wedge\tilde{g},t}$ *denote the mean of* $\mathbf{X}_{it} \circ \mathbf{D}$, *where* $\circ$ *denotes the element-by-element product, in the intersection of groups* $g_i^0 = g$, *and* $g_i = \tilde{g}$. *Let* $\hat{\rho}$ *be the minimum eigenvalue of the following matrix, where the infimum is taken over all possible groupings* $\gamma = \{g_1, ..., g_N\}$:

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

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

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

j $\forall i$, *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, 1.e and 1.g. Endogeneous covariates would be ruled out, but lagged returns interacted with dummies satisfy these conditions. Assumption 1.f 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.h 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.j requires that the lagged returns exhibit sufficient variation over time, for each fund, to identify the fund-specific autocorrelation component. Finally, Assumption 1.i 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 goes as in Bonhomme and Manresa (2015), with

$$\hat{Q} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\epsilon_{it} + \alpha^0_{g_i,t} - \alpha_{g_i,t}\frac{1-\sum_{l=1}^L \theta_{i,l,t}}{1-\sum_{l=1}^L \theta^0_{i,l,t}} + \frac{[(\boldsymbol{\theta}_i^{0\prime} - \boldsymbol{\theta}_i^{\prime}) + \mathbf{D}_{it}^{\prime}(\boldsymbol{\delta}^0 - \boldsymbol{\delta})]\mathbf{X}_{it}}{1-\sum_{l=1}^L \theta^0_{i,l,t}}\right\}^2$$

$$\tilde{Q} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\alpha^0_{g_i,t} - \alpha_{g_i,t}\frac{1-\sum_{l=1}^L \theta_{i,l,t}}{1-\sum_{l=1}^L \theta^0_{i,l,t}} + \frac{[(\boldsymbol{\theta}_i^{0\prime} - \boldsymbol{\theta}_i^{\prime}) + \mathbf{D}_{it}^{\prime}(\boldsymbol{\delta}^0 - \boldsymbol{\delta})]\mathbf{X}_{it}}{1-\sum_{l=1}^L \theta^0_{i,l,t}}\right\}^2.$$

□

B Data cleaning

The original dataset contains 4455 funds. We eliminate funds with vintage year prior to 1984 and after 2012 (69 funds), funds with nothing paid-out (1 fund), funds with size smaller than 10 million USD (27 funds) and funds that Burgiss categorized as *Unknown* or *Not Elsewhere Classified* (76 funds). This first selection leaves a total of 4282 funds in the database, with average size slightly smaller than the original dataset (677 million USD and 661 million USD, respectively). Next, we eliminate quarterly returns that are larger than 100% or smaller than -50%, and only keep funds for which at least 10 pairs of consecutive returns could be calculated. This filter eliminates 60 funds. These funds have similar characteristics to the rest of the sample (non-tabulated).

C Model estimation

We estimate model (1)-(2) using starting values for θ and 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 θ and $\alpha_{g_i,t}$ based on the new estimate for g_i, by minimizing the total sum of squared errors across funds:

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

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.

The k-means algorithm which is used to allocate funds to ABAs, 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, that does not depend on how the problem is initialized.

D Numerical 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)-(2). 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.

E Optimal dynamic allocations

This appendix reports results from a dynamic optimal allocation problem with no transaction cost; shown in the Figure below.

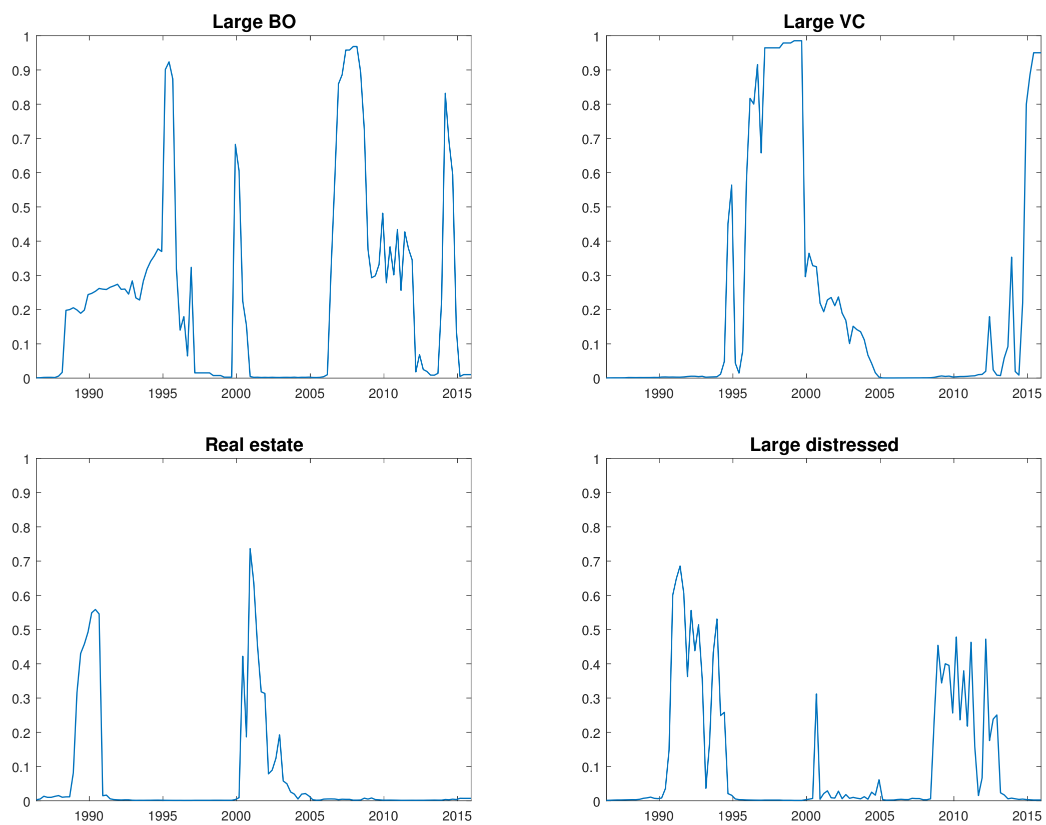

Figure 4. Optimal allocations in the four basis assets, for a mean-variance investor who forms her optimal portfolio every quarter, based on the conditional moments of all assets at hand in the past five years. Transaction costs are set to zero.

F Related Literature

This appendix discusses related papers. 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)).[25] 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 (2017b)).

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 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

[25] As no risk adjustments are made, the debate has concentrated on which benchmark is most appropriate.

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 the stream of papers that model 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, accomodate 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.[26] 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

[26] 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)).

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.

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.