Biases in Private Equity Returns

Simon Hayley (Bayes Business School)
Onur Sefiloglu (Bayes Business School)

Preliminary Draft – Comments Welcome.
Please contact the authors for an updated version before citing.
11 October 2022

Abstract

Private Equity (PE) has grown into a substantial asset class, but there remain major problems with measuring PE fund returns. Investors continue to use the internal rate of return (IRR) as a key measure of fund performance. It is well known that early returns of cash can have a substantial impact on fund IRRs, but the magnitude and causes of this effect have not previously been systematically analysed. We demonstrate that the IRR is affected by two biases: a convexity bias, and a "quit-whilst-ahead" bias arising because the returns on PE projects tend to covary with their durations. Both bias the IRRs of PE funds upwards. Using a range of parametric and non-parametric estimation techniques, we show that these biases boost fund IRRs by an average of around 3% per annum, which is a very significant proportion of the average net PE fund IRR (around 12% per annum). Fund cash multiples and PMEs become similarly biased if they are annualized to try to make them comparable with other assets. We further demonstrate that alternative performance measures which have been suggested by practitioners are also biased, which confirms how poorly understood these effects are. Failure to take proper account of these biases is likely to lead investors into badly misinformed investment decisions.

JEL Classification: G11, G12, G15
Keywords: private equity, internal rate of return, bias.

* Bayes Business School (formerly Cass), London EC1Y 8TZ, UK. E-mail: simon.hayley.1@city.ac.uk. Tel +44 20 7040 0230. The authors are grateful for comments from participants at seminars at Cork University Business School and Bayes Business School, in particular Keith Cuthbertson, Dyaran Bansraj, Nick Motson and Dirk Nitzche. The usual disclaimer applies.

Biases in Private Equity Returns

The PE sector has grown massively over recent years: US Private Equity funds now manage over $6 trillion (2021q4, Preqin). However, measuring the returns generated by PE funds remains very problematic. Funds' valuations of the assets they hold are generally regarded as unreliable, so investors rely instead on performance measures derived from the cashflows between PE funds and investors, especially the internal rate of return (IRR).

Other annualized return measures such as the Public Market Equivalent (PME) and Modified IRR (MIRR) are used in academic studies, but are not generally published by funds. Funds do publish their cash multiples, but these take no account of the time taken to return this cash to investors, so they cannot be directly compared to the annualized returns generally used for other asset classes. For these reasons, investors have little choice but to use the historic IRRs recorded by PE funds as the basis for choosing their strategic asset allocation to PE.

It is well known that IRRs can be strongly affected by early cash distributions from funds to investors. This effect is generally explained as being due to investors being unable to reinvest this cash at a rate equivalent to the IRR, but this is not a helpful explanation since it gives no account of the underlying causes of the effect, or its likely size. Instead we demonstrate two biases in these IRRs. We estimate the size of these biases, demonstrating that they are economically significant in overstating the return investors can expect from holding a strategic allocation to PE. These biases are not the result of deliberate manipulation by fund managers, but are inherent in the statistical distribution of the cashflows generated by PE projects.

These biases must be taken into account if investors are to make like-for-like comparisons between PE returns and those achieved on other assets. Failure to do so will lead investors into misinformed asset allocation decisions.

Variants of the IRR have been proposed by investment professionals to measure the extent to which funds outperform listed equity indices (e.g. direct alpha, ICM/PME, PME+ and mPME). We demonstrate in Section 8 that these measures suffer from the same biases as the original IRRs. This confirms that the effects of early cash returns on fund IRRs remains poorly understood.

2. Literature Review: Measuring PE Returns

Liquid asset classes such as exchange-traded public equity can be marked-to-market every day, making it easy to calculate returns over any chosen period. But market values are not available for the illiquid assets held by PE funds. FASB has since 2008 required US funds to release periodic fair value estimates of their assets. Before this, fund managers were generally thought to keep their valuations artificially low, often at their purchase price. However, PE fund managers are now widely believed to manipulate the valuations of their existing funds in order to encourage investors into new funds that they are starting (e.g. Jenkinson et al. (2013) and Brown et al. (2019), although Huther (2018) disagrees). The SEC has expressed concern about such manipulation. In the absence of reliable valuations, investors are forced to rely on performance measures derived from the cashflows between PE funds and their investors. The key such measures are the internal rate of return (IRR) on these cashflows and the cash multiple.

The cash multiple records the total cash returned to investors over the lifetime of the fund as a multiple of the total cash invested. The Public Market Equivalent (PME) is an alternative to the cash multiple which discounts the cashflows to a common base year before taking the ratio.[1] Academic studies have shown that the cash multiple and PME are both good indicators of relative fund performance (Harris, Jenkinson, and Kaplan, 2014), and PMEs have been used to assess the aggregate performance of the PE sector (Kaplan and Schoar, 2005). However, funds generally do not publish their performance as PMEs. They do publish the cash multiples that they achieve, and these are widely used to compare the performance of different funds, but in choosing their strategic asset allocations investors instead require a measure of annualized PE returns that they can compare with the annual returns reported for other asset classes such as bonds and listed equities. For this purpose they have little alternative but to use the IRR.

A number of papers have estimated the factor exposures of PE funds in aggregate. For buyout funds some studies find market betas above unity, some below[2], with significant

[1] The PME is defined as the present value of total cash flows released by the fund divided by the present value of the total cash invested, with present values calculated by discounting using an appropriate benchmark such as the total return on the S&P500 index (Kaplan and Schoar, 2005). A PME greater than unity shows that the fund concerned outperforms this index. As shown below, the cash multiple can be regarded as a special case of the PME, with the discount factor set to zero. Various alternative performance measures derived from fund IRRs have been proposed by practitioners. Confusingly, some of these are also referred to as the PME. We consider these in Section 8.

[2] Ewens et al. (2013), find $\beta=0.93$ for a sample of BO and VC funds. Jegadeesh, Kraussl, Pollet (2015) find $\beta=0.95$ for BO, $\beta=1.19$ for VC. Ewens, Jones, Rhodes-Kropf (2013) $\beta=0.72$ for BO, $\beta=1.23$ for

additional exposure to small (SMB) and value (HML) equity factors (Ewens et al., 2013, Jegadeesh et al., 2015 and Ang et al., 2018). Estimated alphas range from -0.3% to +1% per annum. Driessen et al. (2012) is something of an outlier reporting much larger market exposure (1.7) and a -1% alpha. These studies report higher market betas for venture capital funds, but a similar range of alphas. It is, of course, very important to distinguish risk premia from outperformance, but this is not the focus of our paper. Continued massive capital inflows suggest that the low alphas reported by these studies do not appear to have diminished investor appetite for PE, and Gompers et al (2016) notes that the simple evaluation measures used by fund managers "raises questions as to whether limited partners understand the returns are leveraged". Consistent with this, we focus on a different question: how misleading are the IRRs which are reported by PE funds?

The modified IRR (MIRR) has frequently been suggested as an alternative to the IRR, but fund managers seldom release their performance data in this form. Such MIRRs are sensitive to the level of returns which is assumed to be earned when early cash distributions are reinvested. If funds use different reinvestment assumptions, the resulting performance measures will not be comparable. Similarly, PMEs are only comparable if based on the same benchmark index. By contrast, the IRR and cash multiple are both derived directly from the fund's cashflows without the need for any judgmental inputs. This may make them seem more objective and unambiguous than the MIRR and PME, and this may help explain their continued popularity. However, whilst academics recommend alternative measures, IRRs are the only widely available estimates of PE funds' annualized returns, so investors have little choice but to use these as the basis for estimating the future returns they expect from PE.

Whatever the reasons, surveys clearly show that investors continue to rely on IRRs. Da Rin and Phalippou (2017) find that the IRR is the measure most frequently cited by investors as the most important criterion in their selection of PE funds. The cash multiple is less popular, and is not in a form that can be directly compared to the returns on other asset classes. The survey by Gompers et al. (2016) finds that PE fund managers primarily use IRRs and multiples to evaluate potential projects (with over 60% using an IRR as their "most important benchmark"), and that they believe that investors in their funds (the limited partners, LPs)

VC. Driessen, Lin, Phalippou (2012) β=1.3 for BO, β=2.73 for VC. Ang et al. (2018) derive β=1.43 for a combined BO/VC dataset, but their method assumes that cumulative log project returns are normally distributed. However, the negative covariance between project returns and durations (see Figure 1) means that even if periodic project returns were symmetrically distributed, the cumulative returns would have a substantial left skew. Ang et al. note that their simulations show that their results are not robust to such a "hold-on-to-losers" effect.

generally rely on the same measures to assess fund performance. Brown et al (2019) show that fund managers' previous IRRs affect their ability to attract commitments to future funds.

Some of the problems associated with IRRs have long been known. It is derived as the solution to a complex polynomial of discounted cash flows, so a set of cashflows may in principle have multiple IRRs, or none at all. Some argue that the IRR does not represent a rate of return at all, yet investors continue to use it as one (Phalippou, 2020). Indeed, use of the IRR has been embedded into accepted reporting practice for PE funds.[3]

It is well known that a single early cash return can have a large impact on the IRR (e.g. Phalippou 2008, 2013). Large early cashflows back in the 1990s have kept some funds' since-inception IRRs at "an artificially sticky and high level" ever since (Phalippou, 2017). Funds have also been accused of deliberately manipulating their IRRs, e.g. by requiring portfolio companies to pay a dividend to the fund as soon as they have been acquired (Rabener, 2020), or funds' recent use of lines of credit to delay calling capital from their limited partners (Albertus and Denes, 2019, Schillinger et al., 2019). These can be regarded as ways in which funds deliberately use hidden increases in leverage to boost their reported returns. Widespread use of these techniques would tend to raise the returns on all PE projects, but the biases we identify in this paper result instead from the variances and covariances of project returns and durations. These biases should thus be regarded as in addition to any increase in the IRR due to hidden increase in leverage. Furthermore, the recent use of lines of credit is likely to have had very little effect on the historical datasets we use to estimate IRR biases.

Bond yields are, of course, also calculated as the IRR of their cashflows, and are affected by the reinvestment problem: that investors will only earn a holding period return equal to this IRR if they are able to reinvest the coupons they receive at the same rate. Unlike typical bond coupons, PE funds generate stochastic cashflows. Nevertheless, the reinvestment problem is often used as a catch-all explanation for PE fund IRRs: that large early cash returns can generate high IRRs which will be misleading if investors are unable to reinvest this cash at the same IRR. This explanation is not very useful because it is couched as a counterfactual: that if investors were able to reinvest at the same rate as the IRR then this IRR would be a perfectly good measure of returns.

[3] The CFA Guidance Statement on Private Equity (2012) stated that "the basic metric and industry practice used in measuring performance in the private equity industry is the since inception internal rate of return (SI- IRR)".

By contrast, we explicitly model the size and timing of PE fund cashflows as stochastic processes and identify two biases generated by the observable statistical properties of these distributions. This approach allows us to quantify these biases using both parametric methods and simulations. Specifically, we show that PE fund IRRs are biased because the timing of the cash distributions to investors is correlated with the returns achieved: high annualised returns are typically generated by projects which mature rapidly. We term this the Quit Whilst Ahead (QWA) bias, and demonstrate how it comes about and how it differs from the other problems with the IRR. We identify a separate convexity effect which also biases PE IRRs upwards compared to the returns on other asset classes.

We derive explicit expressions for these effects, showing (i) that they generate a systematic upward bias in PE fund IRRs; (ii) that these biases arise without any deliberate manipulation by fund managers, purely as a result of the innate characteristics of the distribution of PE cash returns; (iii) that alternative measures of annualised fund returns used by practitioners are similarly biased; (iv) that the PME is not itself biased, but it will become biased as soon as it is converted into an annualised return.

Our estimates show that these biases increase fund IRRs by an average of around 3% per annum. This is economically highly significant since it accounts for a substantial proportion of the 12.2% average net IRR generated by PE funds. Correcting for these biases is likely to remove the whole extent to which these IRRs appear to outperform the returns on listed equity indices such as the S&P500. Many investors now allocate substantial proportions of their portfolios to PE, and they do this on the basis of IRRs which exaggerate the annualized returns achieved by PE. Failure to take account of the biases in these IRRs is thus likely to lead investors into badly misinformed strategic asset allocation decisions.

This analysis also contributes to three different parts of the existing literature: the literature on performance measurement, which investigates the extent to which popular performance measures are inherently biased or can be deliberately manipulated by fund managers (e.g. Ingersoll et al., 2007); the wider empirical literature on bias in the returns reported by other types of fund manager (mutual funds (Elton, Gruber and Blake, 1996, Ter Horst, Nijman and Verbeek, 2001) and hedge funds, e.g. Baquero et al., 2005, and Fung and Hsieh, 2009), and the literature which explores how the IRRs reported for other asset classes differ from the other periodic returns (Dichev, 2007, Friesen & Sapp, 2007, Dichev & Yu, 2011).

The following section derives an expression for the QWA bias, and Section 4 uses this to generate a parametric estimate of this effect. Section 5 identifies an additional convexity bias which further increases fund IRRs. Sections 6 and 7 generate alternative simulation-based estimates which confirm the scale of these effects. Section 8 considers alternative performance measures that practitioners derive from fund IRRs and finds that these fail to remove these biases. Section 9 concludes.

3. The Quit-Whilst-Ahead Bias

The IRR is defined as the discount rate which sets the present value of the cash distributed to investors equal to the present value of the cash invested. This can be parametrized as an initial investment K_0, followed by a stream of cashflows d_t (which can be positive or negative) and a final end-of-horizon payment to investors K_T:

$$K_0 = \sum_{t=1}^{T} \frac{d_t}{(1+IRR)^t} + \frac{K_T}{(1+IRR)^t} \tag{1}$$

We can also express the fund value K_t at the end of each period as a function of organic growth r_t in the fund and any periodic returns of cash to investors (d_t):

$$K_t = K_{t-1}(1 + r_t) - d_t \tag{2}$$

Following Dichev and Yu (2011) we can substitute this into equation 1 to eliminate d_t, and rearrange to give:

$$IRR = \sum_{t=1}^{T} \frac{r_t K_{t-1}}{(1+IRR)^{t-1}} \Big/ \sum_{t=1}^{T} \frac{K_{t-1}}{(1+IRR)^{t-1}} \tag{3}$$

This shows that the IRR is a weighted average of the periodic returns r_t, where the relative weight given to each return is determined by the present value of the portfolio at the start of the period (discounted at the IRR). Thus the IRR is rightly referred to as the dollar-weighted return (although, more correctly, the weights are dollar present values).

The IRRs calculated for a number of asset classes have been found to be significantly lower than the corresponding time-weighted (geometric mean, GM) returns: for the US equity market (Dichev, 2007), mutual funds (Friesen and Sapp, 2007) and hedge funds (Dichev and Yu, 2011). In each case the comparatively low IRRs were interpreted as evidence that investor cash flows were badly timed, buying ahead of low returns and selling ahead of high returns. Hayley (2014) showed that such differentials could instead be caused by cashflows being correlated with past returns rather than future returns (e.g. "return chasing" as investors deliberately increase their exposure following unusually strong returns), and specifically that the differential between the GM returns and IRRs for aggregate US equity markets can be entirely explained by this retrospective bias, leaving no evidence of bad investor timing.

In this paper we will demonstrate similar systematic biases in the IRRs calculated for PE funds. This context is very different, since (i) the cash committed is fixed in advance, giving investors (the limited partners, LPs) no control over the timing of the cashflows over the life of the fund, so "return chasing" behaviour cannot be at work; (ii) the difficulty of valuing PE assets before they are liquidated means that the returns generated in each period by PE funds cannot be reliably measured. Thus, in contrast to previous time series analysis, the present paper investigates whether a bias is inherent in the cross-section of the separate projects that are undertaken by a PE fund.

3.1 Modelling the Cross-Section of Returns within a PE Fund

We model each PE fund as a collection of N individual projects. For simplicity we assume that each project invests one dollar in period 0, and returns a single cash outflow to the investor at maturity T_i, when the project is liquidated. Over its lifetime, each project generates cumulative log returns $\sum_{t=1}^{t=T_i} r_{it}$.

PE assets are illiquid and hard to value objectively before they are liquidated at T_i, so we observe only the cumulative return at maturity rather than individual periodic returns r_{it} within this. Nevertheless, we will model these cumulative returns as $\sum_{t=1}^{t=T_i} r_{it}$ in order to impose the condition that these returns r_{it} have a constant mean μ_r in all periods. This allows us to rule out any effect flowing from T_i to returns (for example, that projects which have already reached a given T_i tend subsequently to generate higher or lower returns than before). The relationship can instead be regarded as flowing entirely from r_{it} to T_i: that unusually high periodic returns tend to be associated with relatively short maturities. This relationship can be thought of as a hazard rate effect: that high cumulative returns to date increase the probability that the project will mature in the next period, whilst the distribution of r_{it} remains *iid* across all projects and time periods.

We derive the present value of each project at the discount rate R:

$$PV_i = \frac{\exp\left(\sum_{t=1}^{t=T_i} r_{it}\right)}{\exp(RT_i)} = \exp\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right) \tag{4}$$

Using $e^x \approx 1 + x + 0.5x^2$ as an approximation accurate for small $\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right)$:

$$PV_i \approx 1 + \left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right) + \frac{1}{2}\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right)^2 \quad (5)$$

If the fund invests in *N* such projects, the IRR is defined as the discount rate that gives zero NPV, implying that the PVs sum to the *$N* initially invested:

$$N + \sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right) + \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right)^2 \approx N \quad (6)$$

$$=> \quad \sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right) + \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i} r_{it} - RT_i\right)^2 \approx 0 \quad (7)$$

We show below that these first and second order terms each generate distinct biases in the IRR.

3.2 First Order Bias

In this section we consider just the linear terms in equation (7), which approximate the IRR as a simple weighted average of the project returns:

$$\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i} r_{it} - T_i R\right) \approx 0 \quad (8)$$

We amend this to consider the excess returns in each period, where $\mu_r = E[r_{it}]$:

$$=> \sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it} - \mu_r) - (R - \mu_r)T_i\right) \approx 0 \quad (9)$$

$$=> \mathrm{R} - \mu_\mathrm{r} \approx \frac{\sum_{i=1}^{N}\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)}{\sum_{i=1}^{N} T_i} \quad (10)$$

We separate the right hand side into *N* separate ratios, and for each one we separate the denominator $\sum_{i=1}^{N} T_i$ into the lifetime of the project in the numerator (T_i) and the lifetimes of all other projects $T_{j \neq i}$. We then divide top and bottom by $N\mu_T$, where μ_T is the population mean of T_i.

$$R - \mu_r \approx \sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\Big/ N\mu_T}{\frac{\mu_T + \sum_{j=1}^{N-1, j\neq i} T_j}{N\mu_T} + \frac{T_i - \mu_T}{N\mu_T}}\right) \quad (11)$$

This makes the denominator approximately equal to 1, and $\frac{\mathrm{T_i} - \mu_\mathrm{T}}{N\mu_T}$ is likely to be small if *N* is large or var(T_i) small, allowing us to approximate using $\frac{1}{1+x} \approx 1 - x$:

$$=> R - \mu_r \approx \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right)\left(\frac{\mu_T + \sum_{j=1}^{N, j\neq i} T_j}{N\mu_T} - \frac{T_i - \mu_T}{N\mu_T}\right) \quad (12)$$

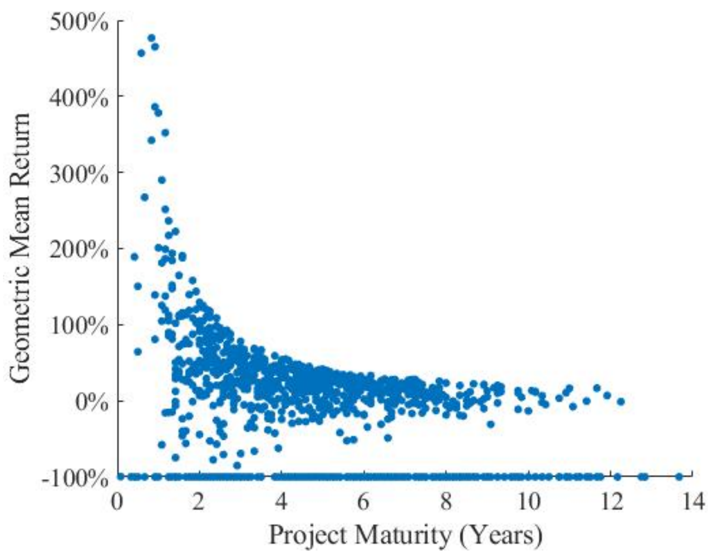
Taking expectations: $\mathrm{E}\left[\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right] = 0$, and the first fraction in the final term contains only terms in T_j $(j \neq i)$, which are independent of the terms in *i*. Hence:

$$E[R] - \mu_r \approx -\frac{1}{N^2\mu_T^2}\sum_{i=1}^{N}\left(E\left[(T_i - \mu_T)\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right]\right) \quad (13)$$

$$E[R] \approx \mu_r - \frac{1}{N\mu_T^2} cov\left(T_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right) \tag{14}$$

Thus the IRR is a biased estimator of the mean periodic return μ_r if the cumulative excess return of each project covaries with its maturity. Figure 1 shows that this correlation is very strong: deals with the highest annualised returns may be liquidated within months, whereas less successful projects last much longer and to generate lower returns. This negative correlation biases the IRR upwards, so the average IRR will be greater than the average periodic returns μ_r achieved by the fund's projects.

Figure 1: Durations And Annualised Returns of PE Deals



Source: MergerMarket (see Section 4.1 for details)

Our derivation above modelled fund managers as picking projects at random from the data shown in Figure 1 and assumed that they have no control over the size or timing of the cashflows that these projects subsequently generate for investors. Purely by good luck, some funds will find that an unusually high proportion of their initial investments generate high returns, and these tend to have short lives. If the funds reinvested the cash from these successful projects into new projects picked from the same population then these reinvestments should be expected to benefit from only an average amount of luck, so they would be likely to drag down the overall fund return. But funds do not reinvest: instead they return the cash to their investors. This avoids diluting the lucky returns in their initial projects with less lucky subsequent projects. We term this the Quit-Whilst-Ahead (QWA) bias, and it will raise average fund IRRs purely as a result of the negative correlation shown in Figure 1: that shorter-duration funds tend to generate higher annual returns.

More formally, we saw that the IRR is (to a first-order approximation) simply an equally-weighted mean of the r_{it} of every project over the lifetimes of these projects $\frac{\sum_{i=1}^{N}\sum_{t=1}^{t=T_i} r_{it}}{\sum_{i=1}^{N} T_i}$. Taken at face value, this seems like a reasonable measure, but it is biased because $\sum_{i=1}^{N} T_i$ on the denominator is endogenous to the project returns in the numerator. Hence we have the following (where the covariance with this reciprocal is positive):

$$E[IRR] \approx E\left[\frac{\sum_{i=1}^{N}\sum_{t=1}^{t=T_i} r_{it}}{\sum_{i=1}^{N} T_i}\right] = E\left[\sum_{i=1}^{N}\sum_{t=1}^{t=T_i} r_{it}\right] E\left[\frac{1}{\sum_{i=1}^{N} T_i}\right] + cov\left(\sum_{i=1}^{N}\sum_{t=1}^{t=T_i} r_{it}, \frac{1}{\sum_{i=1}^{N} T_i}\right) \quad (15)$$

If one of the initial projects has an unusually short life T_i, this increases $1/\sum_{i=1}^{N} T_i$, so such projects are on average given greater weight in the IRR calculation than they would over a fixed horizon. Figure 1 shows that this covariance is strong.

This dynamic bias is distinct from and additional to other effects which have been shown to boost PE fund IRRs where funds (i) require portfolio companies to pay a dividend to the fund as soon as they have been acquired, or (ii) use credit lines to delay calling cash from investors. These two effects are the result of deliberate behaviour by fund managers which boosts the expected IRR by increasing fund leverage, and hence increasing the expected project return in each period (μ_r, in other words E[r_{it}]). By contrast, our derivation above shows that the QWA bias boosts the expected IRR above μ_r. This bias is not caused by the deliberate behaviour of fund managers, but is inherent in the strong correlation observed in our dataset: that projects with short durations tend to generate higher annual returns. In interpreting this QWA bias, we should note:

(a) It is a first order effect. It results from the time period over which returns are calculated being endogenous to the returns achieved. It is distinct from the second order effect that we identify in Section 5 below, which is instead due to the IRR being a non-linear function of the project lives.

(b) The bias is a function of the covariance of project returns and their durations, and Figure 1 shows exactly the situation which would lead to a large positive bias: returns and project lives both have high variance and are strongly negatively correlated. We estimate the size of this bias in the following sections, using a range of different methods.

(c) This bias is not a risk-premium effect. It remains even if we set $\mu_r = 0$ in the derivation above.

(d) For simplicity the derivation above used continuous-time discounting, but this is not vital: the same result can be obtained using discrete time discounting (annually compounded returns are industry standard, e.g. under CFA GIPS), since this gives us the same first-order approximation.[4]

(e) QWA bias is declining in *N*: if a fund contains many projects, then an individual project with high r_i and low T_i will have little effect on $\sum_{i=1}^{N} T_i$ and hence little effect on the weight that this project is given in the IRR calculation, resulting in little bias.

(f) Conversely, this bias should not be presumed to be a short period (small *T*) effect, despite the $1/N\mu_T^2$ in equation (14). Suppose that the relationship between project returns and their durations is linear: $T_i = k \sum_{t=1}^{t=T_i}(r_{it} - \mu_r) + \mu_t + \varepsilon_i$ (where $\mu_r = E[r_{it}]$ and $\mu_t = E[t_i]$). Substituting this into our expression for the bias:

$$QWA\ bias \approx -\frac{1}{N\mu_T^2} cov\left(T_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right) \tag{16}$$

$$\approx -\frac{1}{N\mu_T^2} cov\left(k \sum_{t=1}^{t=T_i}(r_{it} - \mu_r) + \mu_t + \varepsilon_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right) \tag{17}$$

$$\approx -\frac{k}{N\mu_T^2} var(cumulative\ excess\ return) \tag{18}$$

If each periodic return r_{it} is *iid*, then the variance of the cumulative excess return will be proportional to T_i. Hence we would expect the bias to be proportional to $\frac{1}{N\mu_T}$. However this assumes that $\sigma_{T_i}^2$ does not vary with μ_T. If instead the *proportionate* variation in T_i is the same in populations with differing μ_T, then $\sigma_{T_i}^2$ will be proportional to μ_T and the bias will be invariate to μ_T. Thus we should not assume that the QWA bias is a short-horizon effect that shrinks as average project lives increase.

4. Quantifying the QWA Bias

Our starting point in quantifying the QWA bias is simply to evaluate the expression derived above. However, a number of simplifying assumptions were used in its derivation, and the extreme volatility exhibited by PE project returns means that these may well be imprecise. Hence, for robustness we will also use a range of different simulation techniques to estimate

---

[4] Discrete period discounting gives $PV_i = \frac{e^{\sum_{t=1}^{t=Ti} r_{it}}}{(1+R)_0^{T_i}}$. Setting $\sum PV_i = N$ gives us the same first order approximation $\sum_{i=1}^{N}\left(\sum_{t=1}^{t=Ti} r_{it} - \sum_{i=1}^{N} T_i R\right) = 0$ as above, and hence $R \approx \frac{\sum_{i=1}^{N}\sum_{t=1}^{t=Ti} r_{it}}{\sum_{i=1}^{N} T_i}$.

the biases by comparing the IRRs of simulated funds with corresponding bias-free return measures (Sections 6-7 below).

4.1 Data on Private Equity Projects

We use the PE deal exits database from MergerMarket comprising 1,585 investment exits by 438 PE funds with purchase years from 1998 to 2018. The collected data is limited to the transaction dates, percentage of the target company equity bought or sold, and transaction values.

This database has several issues that need to be taken into account. First, it includes buyout (BO) and venture capital (VC) deals, with no way to distinguish them from each other. However, Brown et al. (2020) shows a very similar relationship to Figure 1, using an entirely different dataset comprising only BO projects. This confirms that the covariance between project durations and annualised returns is strongly negative within a population of BO projects, rather than being due to any additional heterogeneity in our dataset due to the inclusion of both VC and BO.

Second, if the purchased company is sold in multiple transactions, the database contains information on only the final tranche sold. In 63% of deals the whole of the acuired equity was sold in a single transaction, so the exit information for these deals is complete. For the remainder, we have information on the final tranche of equity, but not on earlier partial sales of equity in these firms. Our central estimates include the returns generated on these partial sales with the initial investment cost scaled down to reflect the cost of the proportion subsequently sold, on the assumption that they are representative of the whole exit transaction. For robustness we also calculated our results using only the 63% of deals which exited in one transaction. This resulted in very similar estimates for the QWA bias.

Third, the database does not provide information regarding the financing of the deals, and, as these are mostly private companies, it is not possible to obtain this information from other sources. The IRR measures the return on the equity that the fund invests, so we are forced to make an assumption about the leverage involved. For robustness, we have confirmed that our bias estimates are positive economically significant over a wide range of parameter assumptions. Our central assumptions are a leverage ratio of 60% (the average reported by S&P Global Market Intelligence, and close to the levels reported by Metrick and Yasuda, 2010, and Axelson et al., 2013), and an interest rate of 7.5% (consistent with the Libor + 4.9% reported by Axelson et al., 2013).

In common with other research in this field, we filter out projects in each financing scenario which show cash multiples greater than 6 or annualised returns greater than 500%. We confirm that on the basis of these assumptions our simulations generate a distribution of fund IRRs which is similar to those reported in other databases. For robustness we also confirm that the observed covariance of project returns and durations was not generated by filtering out outliers with exceptional cash multiples, but remains substantial and negative with alternative filters or with no filtering of multiples. Indeed, our filtering assumptions are conservative since they reduce the observed covariance.

Table 1 records the key features of the distribution of project returns generated by the Mergermarket data once we have applied the leverage, interest rate and other assumptions discussed above, and compares these with the corresponding data reported in other studies. Our assumptions generate realistic proportions of deals with negative returns, although overall our data turns out to be slightly worse-performing, with higher percentages of bankrupt deals and a lower median project multiple and median annualised return. The MergerMarket database only includes data for companies that were sold as going concerns. It excludes failing companies whose assets were sold separately. These are likely to record below-average returns but our estimates of the QWA bias will not be sensitive to the level of these returns ─ only to their covariance with T_i. The exclusion of these failed deals leaves us with a more homogeneous dataset. If anything this is likely to make our bias estimates more conservative, by removing the possibility that this additional heterogeneity tended to increase the observed covariance.

Table 1: Deal Returns and Durations

	Previous Studies	Our Data
Negative returns %	30 - 40%	38.0%
Bankrupt Deal %	17 - 20%	22.9%
Median Multiple	1.90 - 2.10	1.77
Median Annualised Return	21.0%	14.5%
Median Months Held	51.6 - 60.0	52.0

Sources: the second column shows the range reported by other studies: Lopez-de-Silanes et al. (2015), Hüther (2016) and Braun et al. (2017). The third column reports the corresponding figure implied by our MergerMarket dataset, transformed using the financing assumptions and filters discussed in the text. Each project return is calculated from a single cash investment and a single cash distribution, so the returns referred to here are annualised geometric mean (GM) returns.

Funds report their IRRs net of fees, and these fees are calculated as a function of overall fund performance. For this reason, project data is only available before fees are deducted. To simulate net fund IRRs we are forced to model this fee structure explicitly. We find that the leverage assumptions described above result in a distribution of simulated net fund IRRs which closely matches empirical data reported in other databases (Figure 3 below). We could instead fine-tune our leverage assumptions to more closely match the gross returns reported in Table 1 for other sources, and then adjust our assumed fee structure so that the resulting net fund IRRs continue to match those reported elsewhere. However, the basic parameters of the fee structure are well documented, so adjusting these would itself be unrealistic. For robustness we use a range of different estimation techniques, but our key objective is to estimate the biases that are inherent in the net IRRs reported by PE funds. For these reasons our preferred approach is to use gross data that are an adequately close ─ if not exact ─ empirical match, apply a fee structure with well-known parameters, and then confirm that the resulting distribution of net fund IRRs closely matches other sources. As a further check, in Sections 6 and 7 we derive estimates of very similar size using (i) simulations based on the MergerMarket project data, thus avoiding the inevitable simplifications that come with parametrizing this data; (ii) simulations using an entirely separate database of net fund cashflows.

4.2 Parametric Estimates of IRR biases

Our starting point is to evaluate the expression derived above: $QWA\ bias = -cov(T_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r))/N\mu_T^2$. The projects in our data have an average maturity of 4.3 years, and the covariance between these maturities and the cumulative abnormal return $\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)$ is -2.3. The first row of Table 2 evaluates this expression for different numbers of projects per fund (*N*).

Table 2 Parametric Estimates of QWA Bias

Full dataset

N:	1	2	3	4	5	6	9	12	All
Bias derived from sample covariance	12.5%	6.2%	4.2%	3.1%	2.5%	2.1%	1.4%	1.0%	0.0%
Sum(RiTi)/Sum(Ti)	35.5%	26.4%	24.1%	22.9%	22.3%	22.7%	21.8%	21.4%	20.7%
Implied Bias	14.8%	5.7%	3.4%	2.2%	1.6%	1.9%	1.1%	0.7%	0.0%

Robustness check: projects with maturity over 7.5 years eliminated

N:	1	2	3	4	5	6	9	12	All
Bias derived from sample covariance	10.3%	5.1%	3.4%	2.6%	2.1%	1.7%	1.1%	0.9%	0.0%
Sum(RiTi)/Sum(Ti)	37.5%	30.2%	27.2%	26.3%	25.8%	25.6%	25.4%	25.1%	24.5%
Implied Bias	13.1%	5.7%	2.7%	1.8%	1.3%	1.1%	0.9%	0.6%	0.0%

These positive estimates represent the degree to which fund annualised IRRs are increased by QWA bias. It is intuitive that these estimates are declining in *N*, since if each project accounts for only a very small proportion of fund value then the endogeneity which causes the bias will be correspondingly slight. Jenkinson, Kim and Weisbach (2021) reports that the median buyout fund invests in nine projects, but our derivation above assumed that projects all invest equal amounts, and that returns are independent across projects. These assumptions represent ideal diversification, whereas in practice some projects are much larger than others and systematic and sectoral risk factors may lead to significant covariance across project returns. This will increase the QWA bias, since funds will in effect behave as if they were composed of a smaller number *N* of truly independent projects (or, equivalently, a smaller number of independent principal components). Thus setting *N=9* would underestimate the likely degree of QWA bias. A simple way to correct for this is to reduce our assumption for *N*. On the other hand, these estimates are based on the covariance of gross project returns, and the option-like structure of fund fees means that the variance of net cashflows should be expected to be slightly lower. Taking both these factors into account, we argue below that *N=6* is the most appropriate assumption (implying QWA bias of +2.1%), since this matches the observed variance in empirical net fund IRRs.

Our derivation above used the approximation $1/(1+x) \approx 1-x$, which in this context could be inaccurate for small *N*. To avoid this we return to our earlier approximation $IRR \approx \sum_{i=1}^{N} r_i T_i / \sum_{i=1}^{N} T_i$. The second row of Table 2 shows that this decreases as we sum over progressively larger *N*, consistent with decreasing QWA bias. The estimate which includes all 732 projects in our sample represents a massively diversified fund which will have minimal QWA bias. Differences between this and the corresponding figures for smaller *N* represent the amount of the bias for these more concentrated portfolios. This generates QWA estimates similar to those above. Specifically, for *N=6* we obtain a bias estimate of 1.9% per annum.[5]

As described above, we already excluded extreme outliers from our dataset. As a further robustness check we removed all projects which recorded maturities greater than 7.5 years (a further 9% of our dataset). The results are shown in the second panel of Table 2. Even on this artificially conservative basis our estimates of QWA bias remain economically significant. This is reassuring in showing that the effect is not simply the result of a few extreme observations. However, the QWA bias is a covariance effect, so our estimates will inevitably be sensitive to

[5] This estimate is based on the annually-compounded project returns in our database, since this is consistent with the fact that funds report annually-compounded IRRs. Replacing these with the corresponding continuously-compounded (logarithmic) returns would artificially shrink the right tail of the return distribution, but even under this extremely conservative manipulation of the data our estimated QWA bias remains economically highly significant at 1.3% per annum.

outliers in the distribution of project returns and maturities, and artificially reducing the observed variation in our dataset will inevitably underestimate the QWA effect. Finally, note that our simulation-based estimates in sections 6 and 7 remove the need for any assumptions about parametrization and generate slightly larger bias estimates than the parametric estimates above. As a further robustness check, in the Annex we derive alternative estimates of QWA bias by explicitly modelling the extent to which the returns on all projects in a fund might be affected by systematic risk. These estimates are slightly larger than our estimates above, ranging from a likely underestimate of 1.8% per annum, to a likely overestimate of 3.9%.

Biases of the magnitudes estimated above would be economically highly significant, given that the average net IRR recorded by PE funds has been 12.2% per annum (see Table 4, below). In choosing their strategic asset allocations, investors are likely to compare this figure to the return they expect on other assets. Pension funds generally assume a total asset return of 8% per annum, and this includes significant holdings of comparatively low risk bonds, so the implied expected return on equities must be higher. A total equity return of around 10% would be consistent with the observed long-term geometric mean total return on the S&P500 index (investors who mistakenly use the arithmetic mean S&P500 index as a comparison would derive a larger figure). We cannot pin down investor expectations with any great precision, but these simple comparisons suggest that QWA bias in these IRRs could on its own account for most, if not all, of the amount by which average PE IRRs appear to outperform the return on listed equities.

Furthermore, the estimates above are for only the first order "Quit Whilst Ahead" bias resulting from the covariance of project returns and maturities. In the following section we demonstrate the existence of an additional second order bias resulting directly from the variance of project lives T_i. In subsequent sections we use a variety of simulation techniques to generate non-parametric estimates which include all biases.

5. Second-Order (Convexity) Bias

The figures derived above are estimates of the QWA bias, which is derived from the first order term of the Taylor expansion of the expression which defines the IRR. In this section we demonstrate that there is an additional second order bias which further increases fund IRRs. We use the same framework as above, but this time consider both terms in equation (7):

$$\sum_{i=1}^{N}\sum_{t=1}^{t=T_i}(r_{it}-R)+\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\approx 0 \qquad (19)$$

Decomposing the first order term, as before:

$$(R-\mu_r)\sum_{i=1}^{N} T_i \approx \sum_{i=1}^{N}\sum_{t=1}^{t=T_i}(r_{it}-\mu_r) + \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2 \tag{20}$$

Separating $\sum_{i=1}^{N} T_i$ into T_i and the lifetimes of all other projects $\sum_{j=1}^{N-1,j\neq i} T_j$, then dividing top and bottom by $N\mu_T$:

$$R-\mu_r = \sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\Big/N\mu_T}{\frac{\mu_T+\sum_{j=1}^{N-1,j\neq i}T_j}{N\mu_T}+\frac{T_i-\mu_T}{N\mu_T}}\right) + \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\Big/N\mu_T}{\frac{\mu_T+\sum_{j=1}^{N-1,j\neq i}T_j}{N\mu_T}+\frac{T_i-\mu_T}{N\mu_T}}\right) \tag{21}$$

Noting that $\frac{\mu_T+\sum_{j=1}^{N-1,j\neq i}T_j}{N\mu_T} \approx 1$, we approximate using $\frac{1}{1+x} \approx 1-x+x^2$, where $x = \frac{\mathrm{T_i}-\mu_\mathrm{T}}{N\mu_T}$.

$$R-\mu_r \approx \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right)\left(1-\frac{T_i-\mu_T}{N\mu_T}+\left(\frac{T_i-\mu_T}{N\mu_T}\right)^2\right) + \frac{1}{2N\mu_T}\sum_{i=1}^{N}\left(\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\left(1-\frac{T_i-\mu_T}{N\mu_T}+\left(\frac{T_i-\mu_T}{N\mu_T}\right)^2\right)\right) \tag{22}$$

We first consider the special case where all projects are realised at a fixed horizon $T_i = \mu_T = T^*$, noting that $\mathrm{E}\left[\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right] = 0$:

$$E[R] \approx \mu_r + \frac{1}{2NT^*}E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T^*}(r_{it}-R)\right)^2\right] \tag{23}$$

We saw above that *R* is (to a first order approximation) simply the sample mean of all the periodic returns r_{it}. Hence, we can interpret the quadratic term as the variance of these returns around their own sample mean. This has expectation $\frac{(\mathrm{N}-1)\sigma_r^2}{N}$.

$$E[R] \approx \mu_r + \frac{(N-1)\sigma_r^2}{2N} \tag{24}$$

This equation shows the "diversification return" by which the geometric mean (GM) return on a portfolio of *N iid* assets is raised above μ_r (the expected GM return on a single one of these assets). This requires careful interpretation. First, we note that the IRR of a portfolio of assets with identical maturities $T_i = T^*$ is simply the GM return.[6] Thus this effect is not

[6] The GM is derived from the ratio of the final portfolio value to its initial value: $GM = \left(\frac{\sum_{i=1}^{N} K_{iT^*}}{\sum_{i=1}^{N} K_{i0}}\right)^{1/T} - 1 \Leftrightarrow \frac{\sum_{i=1}^{N} K_{iT^*}}{(1+GM)^T} - \sum_{i=1}^{N} K_{i0} = 0$. The latter expression states that the NPV of the portfolio is zero at the discount rate GM, which is the definition of the IRR. Hence the IRR and GM are identical when all assets have equal maturity.

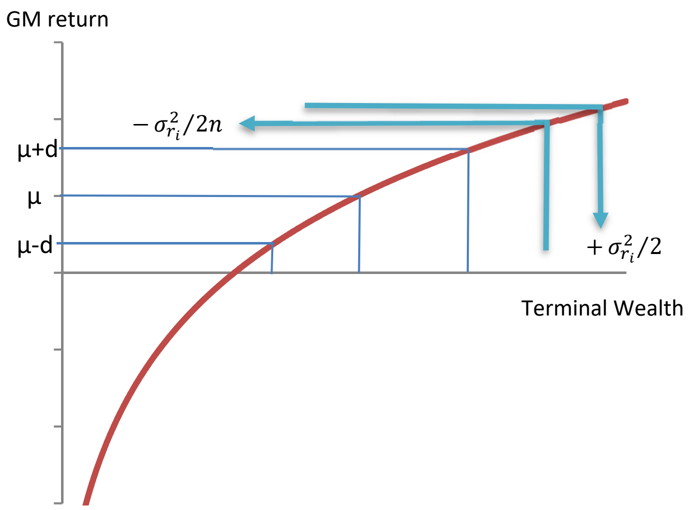
specific to the IRRs that are the focus of this paper. It is relevant whenever we compare the GM returns of different assets or portfolios.

This effect can be regarded as simply an application of the well-known relationship between the geometric and arithmetic means: $\mathrm{E[GM]} = \mathrm{E[AM]} - \sigma_r^2/2$. This relationship comes about because the GM return is a concave function of the final portfolio value: $GM = (K_T/K_0)^{\frac{1}{T}} - 1$ in discrete time (or $GM = \frac{1}{T} log(K_T/K_0)$ in continuous time). For example, if the distribution of asset returns is symmetrical, then the concavity of this function generates an upward skew in the corresponding distribution of terminal asset values, because compounding a slightly above-mean return $\mu + d$ increases the terminal value by more than a below-mean return $\mu - d$ would reduce it. This upward skew boosts the expected terminal value of each asset by $\sigma_{r_i}^2/2$, as can be seen in the textbook result that the mean of a lognormal distribution with mean return μ and variance $\sigma_{r_i}^2$ is $e^{\mu+\sigma_{r_i}^2/2}$, but its median is e^{μ}.

We reverse the process when we calculate the GM returns corresponding to these asset values: the upward skew disappears as the curvature of the return function punishes positive outliers with particularly large K_T, thus reducing the mean return by $-\sigma_{r_i}^2/2$. This effect has been termed "volatility drag".

However, these two effects no longer cancel out when we consider a portfolio of assets. The concavity of the return function: (i) gives an upward skew to the distribution of final values which increases the expected value of <u>each asset</u> by half its variance; (ii) reduces the expected return corresponding to this total final value by half of the variance of the <u>portfolio</u>. Thus for a portfolio of N *iid* assets, the expected final portfolio value is boosted by $\sigma_{r_i}^2/2$, but the portfolio return associated with this distribution is reduced by $\sigma_{r_i}^2/2N$. The expected portfolio return is the expected asset return plus $\frac{\sigma_{r_i}^2}{2}\left(1 - \frac{1}{N}\right)$, as shown in equation (23) above.

Figure 2: Illustration of Volatility Drag and "Diversification Return"



If we are comparing portfolios with similar levels of diversification and containing assets with similar variances then the GM returns of these portfolios would give a like-for-like comparison. But with different levels of diversification return these GMs give a misleading comparison. For example, a particularly highly diversified mutual fund will, all else equal, tend to record a higher average GM return than other funds which invest with equal skill in a smaller number of similar assets. This differential may be misinterpreted as outperformance by the better-diversified fund. Portfolio diversification is obviously desirable, but investors do not need to invest in highly diversified funds to achieve this − they can instead invest in a highly diversified selection of less diversified funds. Comparing GM returns may thus encourage investors to choose inappropriate funds. Cuthbertson et al. (2016) showed that confusion about this diversification return has led to dynamic trading strategies being recommended which have no impact on the expected terminal value of the portfolio, but merely boost the GM return by reducing volatility drag.

This issue is of much wider scope than the PE IRRs that are our focus in this paper. However, we need to take it into account since our goal here is to identify any additional biases which affect the IRRs of PE funds when compared to the GM returns calculated for more liquid assets such as exchange-traded equities. Phalippou (2020) reports that PE fund managers do indeed cite the GM returns on other asset classes as benchmarks against which their fund IRRs should be compared.

Our assumption above that all projects are realised at a fixed horizon T^* made the fund IRR identical to the GM return. Relaxing this assumption shows how variation in T_i affects the

IRR. In the slightly more general case where T_i varies, but is not correlated with either r_{it} or $\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2$, most of the terms in equation (22) have zero expectation (since terms in $T_{j(j\neq i)}$ are independent of the terms in i and $E\left[r_{it}-\mu_r\right]=0$), but the second order product term does not:

$$E[R]\approx\mu_r+\frac{1}{2N\mu_T}\sum_{i=1}^{N}E\left[\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\right]E\left[\left(1+\left(\frac{T_i-\mu_T}{N\mu_T}\right)^2\right)\right] \tag{25}$$

$$\approx\mu_r+\frac{(N-1)\sigma_r^2}{2N}\left(1+\frac{\sigma_t^2}{N\mu_T}\right) \tag{26}$$

Thus, even in the absence of the covariance which generates QWA bias, the variance of T_i increases the expected IRR. Our dataset shows that these project maturities vary widely, with a mean of 4.3 years, but a standard deviation of 2.3 years. However, before using this expression to quantify this effect, we should recall that it is based on the approximation $\frac{1}{1+x}\approx 1-x+x^2$, which is adequate to demonstrate the existence of this effect, but may be a poor approximation, since the higher order terms in this Taylor series could also be significant. In particular, the distribution of project maturities T_i is positively skewed, with fewer of the extremely short durations which would maximise this convexity effect. Furthermore, shorter-maturity projects record more volatile annualised returns than longer projects (see Figure 1), and this negative correlation will reduce the positive skew in the resulting distribution of terminal values. Hence the positive effect on the expected IRR is likely to be smaller than the $\left(1+\frac{\sigma_t^2}{N\mu_T}\right)$ derived above. To avoid an inappropriate approximation, we recall that this effect comes about because we derived our expression for the expected IRR by dividing both sides of equation 20 by $\sum_{i=1}^{N}T_i$:

$$(R-\mu_r)\approx\frac{\sum_{i=1}^{N}\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)}{\sum_{i=1}^{N}T_i}+\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2}{2\sum_{i=1}^{N}T_i} \tag{27}$$

If $\sum_{i=1}^{N}T_i$ is correlated with the cumulative returns then we have QWA bias. If there is no such correlation then the linear term on the right has zero expectation, leaving:

$$E[R-\mu_r]\approx E\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2}{2\sum_{i=1}^{N}T_i}\right]=\frac{1}{2}E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\right]E\left[\frac{1}{\sum_{i=1}^{N}T_i}\right] \tag{28}$$

When T_i is fixed at $T_i=\mu_T=T^*$, this simply becomes $\frac{1}{2NT^*}E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)\right)^2\right]$ giving us the expression above for the diversification return for the portfolio geometric mean return calculated over a fixed horizon. But when T_i varies independently of returns, this boosts the diversification return simply because the reciprocal is a convex function such that $E\left[\frac{1}{\sum_{i=1}^{N} T_i}\right] > \frac{1}{E[\sum_{i=1}^{N} T_i]}$ (given that $T_i > 0$). Our data on project returns shows that for N=6 the left hand side of this inequality is 5% larger than the right, so independent variance in T_i in our dataset should be expected to boost the diversification return to 7.05% from its original 6.71% (derived by evaluating $\frac{(\mathrm{N}-1)\sigma_{\mathrm{r}}^2}{2N}$ using the observed 16.1% maturity-weighted variance of project returns). This represents a more modest increase (+0.34%) in the expected IRR than the QWA bias discussed above, but: (i) it represents an entirely separate effect, which will arise even in the absence of QWA bias; (ii) it should be considered a bias since it boosts the IRRs recorded for portfolios of assets with varying T_i above those (GMs) of portfolios with entirely similar return characteristics but a fixed horizon.

6. Simulation Estimates of the Biases

The parametric bias estimates derived above were based on a number of assumptions. We now turn to simulation techniques as an alternative. These avoid the simplifications that are inherent in parametrizing the data. Furthermore, our estimates above were for only the first and second order biases, whereas simulations will also include any higher order effects.

We demonstrated that QWA bias and convexity bias both arise because of variation in the maturities (Ti) of the projects in which PE funds invest, either directly (convexity bias), or because they are correlated with project returns (QWA bias). Both biases will be absent in a strategy which invests in PE over an exogenously fixed horizon, immediately re-investing any early cash distributions into similar PE investments. Comparing simulated fund IRRs (without reinvestment) with the simulated returns of funds which reinvest in similar projects over a fixed horizon will thus give us an estimate of the combined effect of QWA and convexity biases.

To confirm this intuition we again assume each PE fund comprises *N* projects which each invest $1, generate periodic returns r_{it} and mature at time T_i. Now we assume that the cash released by these projects at maturity is reinvested at rate R^* until a fixed horizon T^*, at which point their terminal values are measured. Discounting these terminal values at discount rate *R* gives: $\sum_{i=1}^{N} \mathrm{NPV_i} = \sum_{i=1}^{N}\left(e^{\sum_{t=1}^{t=T_i}(r_{it}-R)+\sum_{t=T_i}^{t=T*}(R^*-R)} - 1\right) = 0$ as the expression which defines *R*=IRR. Consider the first two terms in the Taylor expansion of this exponential function:

$$\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)+\sum_{t=T_i}^{t=T^*}(R^*-R)\right)+\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)+\sum_{t=T_i}^{t=T^*}(R^*-R)\right)^2\approx 0 \quad (29)$$

We decompose the first term, substituting: $\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)+\sum_{t=T_i}^{t=T^*}(R^*-R)\right)=\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)+\sum_{t=T_i}^{t=T^*}(R^*-\mu_r)-NT^*(R-\mu_r)\right)$, where $\mu_r=E[r_{it}]$. Back in section 3.2, when we reached this stage in the derivation of the IRR we found that obtaining a first-order solution for R required us to divide by $\sum T_i$ (see equation 10). This gave rise to the QWA bias because the $\sum T_i$ in the denominator covaries with the project returns in the numerator. This time we are dividing by an exogenous *NT**, so the IRR calculated over this fixed horizon does not suffer from QWA bias:

$$R\approx\mu_r+\frac{1}{NT^*}\left(\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)+\sum_{t=T_i}^{t=T^*}(R^*-\mu_r)\right)+\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-R)+\sum_{t=T_i}^{t=T^*}(R^*-R)\right)^2\right) \quad (30)$$

We saw earlier that convexity bias is a function of the variance of the project maturities T_i, so this too will be zero when the IRR is calculated over a fixed horizon. The IRRs calculated without reinvestment suffer from both QWA and convexity biases, but equivalent IRRs calculated with reinvestment to a fixed horizon suffer neither of them. Thus the difference between the IRR and the return on a corresponding fixed-horizon strategy can be used to estimate the size of the biases inherent in the IRR.

It is worth noting that the total return indices which are calculated for other assets similarly assume that all cash distributions (such as dividends or coupons) are immediately reinvested into the same asset. Using the same approach for PE will ensure a like-for-like comparison with these other assets. Such reinvestment means that there are no intermediate cashflows so, as we saw in the previous section, the IRR becomes identical to the geometric mean return.

Such fixed-horizon reinvestment strategies could be modelled using either fixed or stochastic returns during the reinvestment period. For robustness we take both approaches in the following sections.

6.1 Fixed Rate Reinvestment

The most obvious way to model fixed-horizon returns is to calculate the modified IRR (MIRR) which assumes that a steady rate of interest is earnt on any cash returned to investors before the end of the time horizon. The textbook approach is to calculate the MIRR using the risk-free interest rate (typically proxied as the yield on short-term government bonds). This is likely to be very low compared to the returns expected on PE investments. The MIRR calculated on this assumption would represent the return on a dynamic strategy which shifts from a high-risk asset (PE) into cash over the course of the investor's horizon. It is not clear why investors would choose such a strategy, so although it is an investible strategy, we have no reason to think that it represents a sensible benchmark. Choosing such an inappropriate reinvestment rate would understate the expected returns that should be expected from maintaining an exposure to PE. Larocque et al. (2022) compares fund IRRs with the "multiple return" calculated from the fund's cash multiple and lifetime as $MOIC^{1/T} - 1$, but this is simply the MIRR calculated with a reinvestment rate of zero. Under this extreme assumption they find that the mean fund IRR exceeds the mean multiple return by almost 8% per annum.

Phalippou (2008) instead suggested using a fixed reinvestment rate of 8% (a figure commonly used as the hurdle rate used in calculating fund performance fees), or the total return on a popular listed equity index such as the S&P500. However, these may still be excessively low if PE funds earn additional risk premia, such as those resulting from a CAPM beta greater than unity, an illiquidity premium (as was identified by Franzoni et al., 2012) or a small capitalisation equity premium (e.g. Jegadeesh, Kraussl and Pollet, 2015).

The project returns in our dataset incorporate any such risk premia. The weighted average of the annualised discrete time gross GM returns achieved by each project in our dataset (weighted by their lifetimes (T_i) and gross of fees) is 20.7% per annum. This represents a first-order approximation of the average return that would be achieved if we invested sequentially in projects picked at random from this dataset (the T_i-weighting reflects the fact that the probability that at any given moment we are invested in a high-return project is relatively low because such projects tend to have short lives). This figure incorporates no QWA bias, because it measures the return on a strategy which involves no early returns of cash (hence no "quitting").

However, using this figure would lead to misleadingly low MIRRs because this return is constant, and so will not generate the diversification return seen on the stochastic returns on the initial investments. To show this, we return to equation (30) and set our reinvestment rate equal to the mean periodic return $R^* = \mu_r = E[r_{it}]$. Taking expectations:

$$E[R] \approx \mu_r + \frac{1}{2T^*} E\left[\left(\sum_{t=1}^{t=T_i} (r_{it} - R) + \sum_{t=T_i}^{t=T^*} (R^* - R)\right)^2\right] \qquad (31)$$

We saw in Section 3.2 that to a first-order approximation, the IRR is the sample mean of the periodic returns r_{it}, so the cross-product above has approximately zero expectation:

$$\approx \mu_r + \frac{1}{2T^*} E\left[\left(\sum_{t=1}^{t=T_i}(r_{it} - R)\right)^2\right] + \frac{1}{2T^*} E\left[\left(\sum_{t=T_i}^{t=T^*}(R^* - R)\right)^2\right] \quad (32)$$

The first quadratic term is half the variance of these returns around their sample mean $\left(= \frac{N-1}{2N} T_i \sigma_r^2\right)$. As we saw earlier, this is the diversification return caused by the variance of the returns on the initial investment projects. However, the second expectation term is much smaller: it is half the variance of this sample mean R around μ $\left(= \frac{1}{2N}(T^* - T_i)\sigma_r^2\right)$. Thus using a fixed reinvestment rate which does not correct for this diversification return will lead to MIRRs which are systematically lower than the corresponding IRRs. This reduction should not be regarded as the correction of a bias, since the diversification return represents the increase in the mean terminal wealth due to the upward skew caused by compounding: an effect which is of genuine value to investors. To ensure that our MIRR is a like-for-like comparison with the returns calculated on stochastic assets held throughout the investment horizon we should either (i) use a fixed reinvestment rate comprising: the risk-free rate + appropriate risk premia + $\frac{N-2}{2N}\sigma_r^2$ to compensate for the missing diversification return (we take this approach here); or (ii) use stochastic reinvestment returns with appropriate mean and variance (we simulate these in Section 7).

To derive reinvestment rates which include an appropriate diversification return, we consider the 30.4% gross IRR generated by a portfolio which invests equally in every project in our database. This IRR includes minimal QWA and convexity biases (since these both fall with *1/N*, and our dataset includes hundreds of projects), but does include diversification return. By contrast, the 20.7% T_i-weighted average project GM return contains no diversification return or convexity bias (since it is the average of single-project returns, and for *N=1* these effects are both zero). The difference between these two returns thus represents the maximum possible diversification return for a large *N* portfolio. For smaller portfolios we include a fraction $\frac{\mathrm{N}-2}{2N}$ of this additional return. Table 3 compares (i) the mean IRR for funds each containing *N* randomly-selected projects from our dataset; with (ii) the corresponding MIRR that would result if early returns of cash from these projects were reinvested at a constant return. These IRRs and MIRRs include the same risk premia and equivalent levels of diversification return (increasing in *N* as diversification improves), but the IRRs will be boosted by QWA and convexity biases, whilst the MIRRs will not. The difference between these figures represents the combined size of these two biases.

Choice of time horizon is also important. The conventional approach would be to choose a horizon at least as long as the longest project. However, this would make our estimates very sensitive to any error in the reinvestment rate used: an excessively low rate would reduce the MIRRs and overestimate the biases. We also modelled returns over a very short horizon of three years. Many projects will mature after this horizon so, rather than reinvesting, the MIRR calculation discounts project terminal values back to the equivalent PV at the three year horizon. An excessively low reinvestment (discount) rate would thus overstate the MIRRs and understate the biases. We can observe this effect in Table 3, where for our central case *N*=6: a 7.5 year horizon with our lower reinvestment rate (20.7%) generates lower MIRRs than with the higher (30.4%) rate, and hence a higher bias estimate. Using the 3 year horizon this difference is reversed, since these reinvestment rates will frequently be discount rates, and a lower discount rate applied to future project liquidation values boosts the MIRR, implying lower bias. Our central estimates are based on a horizon of 4.3 years. This is the mean project life, so using this horizon implies that there will be as much discounting back as reinvesting, implying that the average MIRRs calculated over this horizon should be least sensitive to the possibility that we are using an inappropriate reinvestment rate. For this reason we consider them the most robust estimates.

Table 3: IRR-MIRR Median Differentials for Simulated Funds (per annum)

Reinvestment Rate	Time Horizon (years)	Bias (N=3)	Bias (N=6)	Bias (N=9)	Bias (N=12)
20.7%	3.0	4.3%	1.3%	0.7%	0.2%
Mid	3.0	4.6%	2.0%	1.3%	0.8%
30.4%	3.0	5.8%	2.9%	1.9%	1.4%
20.7%	4.3	6.8%	4.7%	3.7%	3.4%
Mid	4.3	6.7%	3.8%	2.8%	2.5%
30.4%	4.3	5.5%	2.7%	1.4%	1.3%
20.7%	7.5	9.3%	7.2%	7.0%	6.4%
Mid	7.5	8.3%	5.4%	4.5%	4.2%
30.4%	7.5	4.4%	2.3%	1.4%	1.1%

The "Mid" reinvestment rate varies with N. As derived above, it includes $(\mathrm{N}-2)/2N$ of the difference between the low (20.7%) and high (30.4%) rates.

Table 3 shows that our estimated biases are positive across this very wide range of assumed parameters for the number of projects per fund, the time horizon and the reinvestment rate. The results also confirm that bias is smaller for funds containing more projects (higher *N*) as the QWA bias is reduced. Jenkinson, Kim and Weisbach (2021) reports that the median buyout fund invests in nine projects. But, as discussed above, drawing nine projects at random from our historic dataset represents an unrealistic degree of diversification, due to systematic risk and projects of unequal size. We correct for this by setting $N=6$. This generates a realistic standard deviation of fund IRRs (Figure 3, below), whilst avoiding the need to specify the risk factors which account for this systematic risk. This assumption gives us our central estimate of 3.8% per annum bias: somewhat greater than the sum of our parametric estimates above for QWA bias (1.9%) and convexity bias (0.34%).

As a further robustness check, we repeat these MIRR calculations after removing any projects which recorded lives greater than 7.5 years. This only slightly reduces the central bias estimate (for $N=6$) to 3.5% per annum, confirming that this bias is not simply the result of a few exceptionally long-lived projects.

7. Stochastic Reinvestment

In this section we compare the IRRs of conventional PE funds (which distribute the proceeds of mature projects to investors), with the returns generated if this cash is reinvested into new projects until a fixed horizon. We saw above that such a fixed-horizon return will be free from QWA and convexity biases, so the difference between the estimates derived on these different assumptions will again give us an estimate of the size of these biases.

For robustness we will derive estimates using two very different approaches and different datasets: (i) simulating reinvestment by each fund into new projects (in effect simulating the returns on "evergreen" funds); (ii) simulating the returns achieved by an investor who reinvests the cash distributed by a mature fund into a new PE fund. These approaches both remain fully invested over a fixed horizon, and so will be free from QWA and convexity biases.[7]

[7] As a further robustness check we tried shuffling the data to remove the covariance of project returns and maturities by matching each r_i with a randomly drawn Ti. However, this generates some projects which generate extremely large annual returns over a period of many years, resulting in massive dispersion of the final values generateed by different projects. This resulted in correspondingly far greater volatility drag and diversification return than is seen in actual fund data. In removing the first-order QWA bias this massively increased the second order bias, and was thus did not generate useful results.

7.1 Simulating PE Fund IRRs Using Project-Level Data

We first consider the returns generated by a PE fund which immediately reinvests the proceeds from each maturing project into a new project. These stochastic project returns could be modelled either parametrically or by bootstrapping historic PE returns. However, in either case, generating returns over a fixed investment horizon raises problems.

Consider an investment in an initial project which generates an annualised return r_1 and matures at T_1. These values are then fed into the IRR calculation, but to generate the corresponding annualised return up to our fixed horizon T^* we need to simulate the return r_2 which is earned when this cash is reinvested. We know that r_2 is strongly correlated with the maturity T_2 of this second project, so if we select a project (either parametrically, or directly from our historical dataset) with exactly the required maturity $T_2 = (T^* - T_1)$, the return on this project will be conditional on T_1, and hence on r_1. Thus simulating a return which exactly fits the horizon T^* threatens to produce very misleading results by generating autocorrelation of returns.

To avoid this we must select reinvestment projects unconditionally. We draw these projects from the same pool as we used to select the initial projects with replacement. Once the final project extends beyond our pre-defined horizon we discount its future liquidation value back to our horizon. Thus stochastic modelling still requires the use of an appropriate discount rate to generate our end-of-horizon valuations. Given this issue, taking a parametric approach to generating these reinvestment returns has little to recommend it since it introduces the additional risk of misparamaterisation without avoiding the need for discounting.

Our simulations use the same project-level data as above. We simulate 10,000 funds with a life of 10 years. Each fund invests in private equity deals drawn from the project database with replacement. As discussed in section 4.1, we chose leverage parameters which generated net IRRs with a distribution in line with that observed empirically.

We assume that capital is called 2.5 years after being committed by investors (similar to the figure reported by Metrick and Yasuda (2010) although, of course, the IRR is a calculated from the first cashflow rather than the commitment of capital). This leaves 7.5 years for the deals to be liquidated and the proceeds distributed back to the investors. Durations of the deals that exceed 7.5 years are capped at 7.5 years, keeping their IRRs fixed and recalculating their multiples accordingly (for robustness, in ongoing work we model alternative assumptions).

The IRRs quoted by PE funds are calculated from the net-of-fees cash returned to investors, so to generate comparable data we also need to simulate these fees. There are four broad groups of fund fees for private equity funds: Management fees, transaction costs, carried interest and monitoring fees:

a) A typical buyout fund charges a management fee of 2% on the committed capital (during the investment period) and on the invested capital (during the harvesting period);
b) Each project is charged a transaction cost of 1% of the deal value (after 50% reimbursement to the investor, Metrick and Yasuda, 2010). For simplicity, we combine these with management fees as "Fixed Fees" in the tables below;
c) A carried interest of 20%, based on a hurdle rate of 8%, with catch-up, claw-back and a waterfall structure;[8]
d) Metrick and Yasuda (2010) reports that monitoring fees are 0.40% of the firm value annually, but 80% of these fees are reimbursed to the investors. Since the remaining effect is negligible, we do not take these fees into consideration.

Our base case assumptions are leverage ratio 0.60 (defined as debt/assets), interest rate 7.5% and 2.5 years between commitment of capital and first call. As discussed above, we assume that each fund contains 6 independent projects, since this makes allowance for an appropriate degree of systematic risk. Table 4 shows that our simulated fund fees are very similar to those documented by Phalippou (2009). Figure 3 compares the Net IRR distribution of our simulated funds with the Net IRR distribution of the funds in the Preqin Cash Flow database. This shows that our base case assumptions produce a realistic distribution of net fund IRRs.

Table 4: Actual and Simulated Fund Net IRRs and Manager Fees

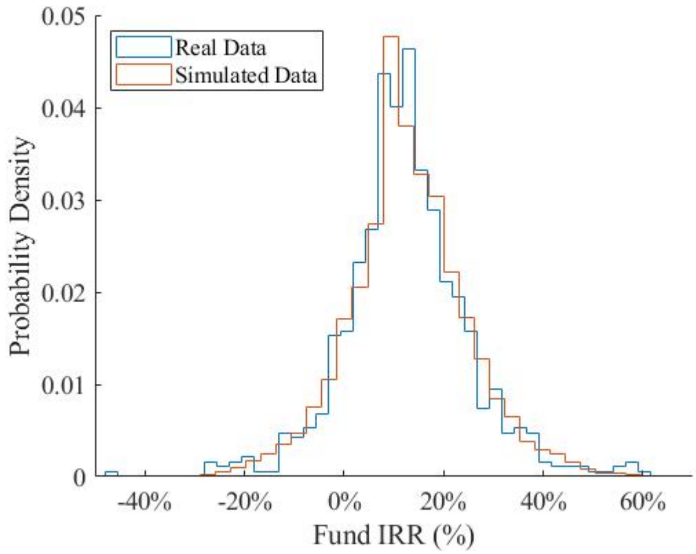
Panel A: Net IRR	Actual	Simulated	Difference
Mean	12.7%	12.7%	0.0%
Median	12.2%	12.5%	-0.3%
St. Dev.	13.1%	13.5%	-0.3%

[8] According to this structure, as long as the combined fund returns do not exceed a pre-defined hurdle rate (almost always 8%), all fund profit goes to the fund investors (limited partners). If fund return exceeds the hurdle rate, catch-up provision kicks in and fund managers retain the excess profits until the total profits are shared as 80/20 between investors and fund managers. For the further excess profits, 80/20 split is maintained. Generally the profit split is made on a deal level, and a claw-back provision, results in a final rebalancing at the end of the fund life based on the total fund returns.

Panel B: Manager Fees	Benchmark	Simulated	Difference
Fixed Fees	4.8%	5.3%	-0.6%
Performance Fees	2.0%	3.0%	-1.0%

Sources: Panel A actual: Preqin. Panel B benchmark: Phalippou (2009).

Figure 3: Simulated vs. Actual Fund Net IRRs (%)



Source: Preqin buyout fund net cashflows (1980-2005), simulations based on Mergermarket (1998-2018) deal data.

We will compare these simulated IRRs with the simulated returns of PE funds which reinvest the proceeds of their initial deals into new projects drawn with replacement from the same pool of PE projects until the end of our fixed horizon (10 years) so that the fund remains fully invested throughout this period.

For robustness we introduce different levels of return persistence in the deals in which funds invest. Such persistence might be due to manager skill, or differences in the risk premia earned by funds specialising in different sectors or within sub-periods of our dataset. Performance persistence should be expected to reduce bias by reducing the cross-sectional variance of the returns on the projects in which funds invest. Our first reinvestment model (Model 1) assumes no performance persistence: successive deals are independent. Model 2 introduces a high level of performance persistence by drawing the reinvestment deals from the initial set of invested deals. Thus a fund which initially invested in successful deals is likely to continue to do so. Model 3 generates moderate performance persistence by drawing projects for reinvestment from one of two deal clusters. The first cluster includes all the deals drawn by the successful half of the PE funds, the second cluster includes the deals drawn by the other half of the PE funds. Both clusters include all deals, but the successful cluster has a higher share of the better ones.

We quantify the resulting persistence using a simple regression model of the annualised return of reinvestment deal k invested by fund f against the return on the fund's preceding deal $DealGMreturn_{f,k} = \beta DealGMreturn_{f,k-1} + \varepsilon$. Model 1 generates β=0, Model 2 β=0.21, Model 3 β=0.13. By comparison, Braun, Jenkinson and Stoff (2017) look at the performance of successive projects by a given manager and find a regression coefficient of 0.102.[9] Thus our different models introduce unrealistically little persistence (Model 1), unrealistically strong persistence (Model 2), and a more plausible central scenario (Model 3).

For each of these models we simulate the net IRRs of PE funds containing (i) nine projects per fund (the observed median for buyout funds), and (ii) six projects per fund (which, as discussed above, generates a more realistic level of variance in our simulated IRRs).

We saw above that QWA and convexity biases are both due to variation in the maturities of the individual project returns within the fund (and hence in the timing of the cashflows returned to investors). We thus estimate the sum of these two biases as the difference between (a) the fund IRR without reinvestment (subject to both biases), and (b) with reinvestment over a fixed horizon (which removes both biases).

The results are shown in Table 5. As expected, a larger number of projects per fund results in lower bias, but despite this – and other differences between the models – the resulting bias estimates are in the relatively narrow range 3.2% to 4.4%. Biases of this magnitude will clearly be highly economically significant to investors since they account for a significant proportion of the average net IRR of 12.2%.

These estimates are all slightly higher than the parametric estimates in Sections 4 and 5 (1.9% QWA plus 0.34% convexity bias). This can be attributed to the fact that these simulations: (i) avoid the simplifying assumptions required in our derivations and (ii) allow possible higher-order effects in addition to the first order (QWA) and second order (convexity). The fact that our simulations model net-of-fees fund IRRs and our earlier parametric estimates

[9] Braun, Jenkinson and Stoff (2017) regress the PMEs of successive fund deals, whereas we use deal GM returns (IRRs), but as long as the market return (which is netted off the fund return when the PME is calculated) is assumed to have near-zero autocorrelation the regression coefficients will be comparable to ours.

were based on gross project returns makes little difference, since the assumed leverage in the latter was calibrated to match the empirically-observed volatility in net fund IRRs.

Table 5: Simulated Median Net Fund Returns and Implied Bias

	N = 6			N = 9		
	Model 1(no persistence)	Model 2 (high persistence)	Model 3 (medium persistence)	Model 1(no persistence)	Model 2 (high persistence)	Model 3 (medium persistence)
Median IRR (no reinvestment)	12.3%	12.3%	12.3%	12.4%	12.4%	12.4%
Median return over fixed horizon (with reinvestment)	7.9%	8.9%	8.2%	8.9%	9.3%	8.9%
Median Bias	4.4%	3.4%	4.1%	3.5%	3.2%	3.5%

It is also worth emphasizing that Table 5 reports the median bias for each scenario (our earlier parametric estimates were, by their nature, means). The mean simulated biases are substantially higher, but we are again keen to be conservative in our estimates. For robustness we also repeated these simulations whilst excluding all deals recorded after 2014 (in order to avoid any risk of self-selection resulting from deals which were not yet complete being omitted from the MergerMarket dataset). This left our central bias estimate unchanged at 3.4%.

Table 6 shows the corresponding bias estimates for different deciles of the IRR distribution, with substantial bias in all deciles.[10] We would expect this sorting to reduce the variance of project returns and maturities within each decile. The fact that this results in only a slight reduction in the overall median (from 3.4% to 3.0%) is a reassuring sign that our earlier estimates were not badly affected by the pooling of potentially heterogenous projects (e.g. projects from different vintages, or from sectors with persistently different risk premia).

Table 6: Median Simulated Bias by IRR Decile (N=6, high persistence)

Max IRR	2	3	4	5	6	7	8	9	Min. IRR
2.9%	2.3%	2.0%	1.4%	1.9%	2.3%	3.3%	3.9%	5.7%	5.7%

[10] For this decomposition to be meaningful we need to apply the high persistence assumption (otherwise unluckily low IRRs record higher fixed-horizon returns as reinvestments tend to be less unlucky – this is simple mean reversion rather than an indication of bias). With high persistence, reinvestment will on average be equally lucky as the initial project choices, although with some variation depending on the frequency with which the best projects among the initial projects are picked again for reinvestments.

7.2 Simulated Investor Recommitment Strategy

In this section we simulate an investor who invests in a PE fund and subsequently reinvests the cash distributions from this fund into a new PE fund. This is another way of modelling the return generated by maintaining an exposure to PE over a fixed time horizon, which will be free of convexity and QWA biases. This alternative approach allows us to check the robustness of our earlier simulations using different assumptions and an entirely separate database.

We use the Preqin database of fund cashflows. This contains the data on 4,355 PE funds (with vintages between 1980 and 2018), including the dates and amounts of the cash inflows to the PE funds (capital calls), and cash distributions to the investors (net of fund manager fees). From the 30 different fund strategies included in the database, we limit our attention to buyout funds (consistent with our project-level simulations, which were calibrated to match the empirical distribution of net BO fund IRRs). We also filter out all funds with a vintage later than 2005, since these might still make significant future cash distributions (the typical private equity fund has a life of 10-13 years).[11] Consistent with other studies we also filter out a small number of funds with extreme IRRs and very short durations, leaving a sample of 369 buyout funds.

Table 7 summarizes the main characteristics of this data. Mean (Median) IRR of PE funds is 12.7% (12.2%), in line with the summary statistics in Gupta, Nieuwerburgh (2019). The average fund calls 86% of committed capital, and distributes close to double this amount, but with substantial variation — the worst performing funds end up close to complete failure. On a value-weighted basis, these funds manage investor funds for an average of five years.

Table 7: Characteristics of Net Fund Cashflows (Buyout Funds)

	Mean	Median	Min	Max
IRR	12.7%	12.2%	-46.8%	102.3%
Contributions (% of commitments)	8.5	8.6	3.6	14.6
Distributions (% of commitments)	16.3	15.2	0.5	117.2
Duration (years from first call)	5.0	4.9	1.3	10.7
Vintage	1998.9	1999	1980	2005

Source: Preqin. Note that this data excludes completely failed funds, although in practice these are rare – most funds make at least some cash distributions.

[11] Kaplan and Schoar (2005), Phalippou and Gottschalg (2009) similarly eliminate funds younger than 10 years. We are more stringent in removing live funds, since we require fund cashflows for our simulations. For this reason we wish to remove live funds which include a final valuation generated by the fund managers. For this reason we eliminate funds younger than 13 years, considering the possibility of 3-years fund life extension.

Investors commit to a PE fund, remit capital to the fund when it is called, and receive distributions when the projects are exited. After the final cash flow the fund IRR is calculated using the complete set of net cash flows. In the counterfactual scenario, distributions received by the investor are committed to another PE fund. Such "recommitment" strategies are widely discussed in the practitioner literature (e.g. Cardie et al., 2000, de Zwart et al., 2012, Nevins et al., 2004, Oberli, 2015). Endowment funds are known to reinvest in this way, since they have allocations to PE which are not only substantial, but also very stable over time (Azlen and Zermati, 2017). More generally, such reinvestment is likely to be implicit in investors' strategic asset allocation decisions, which typically assume that the chosen asset allocations are maintained throughout the investment horizon.

In practice, investors commit to multiple funds simultaneously, in order to diversify returns and make their aggregate cashflows more predictable. We consider an investor who commits to a single fund, and when sufficient cash has been distributed by this fund, reinvests into another single fund. This is a less attractive strategy, but it generates returns which are a like-for-like comparison with the IRRs reported by individual funds. By contrast, a portfolio of such funds would generate additional diversification return which would interfere with this comparison.

Fund cashflows can extend considerably further than the maturities recorded for individual projects, so in order to allow investment in a sequence of funds we model an investment horizon of 40 years — substantially longer than for our project-level simulations. At the end of this horizon the investor stops reinvesting, then all remaining distributions are collected and the resulting cashflows are used to calculate the annualised IRR. This leaves only slight residual variation in the effective horizon (final cashflows are received between 40 and 45 years), so the scope for QWA or convexity bias is very small. The difference between this figure and the average PE fund IRR gives us an estimate of the systematic biases. Unlike the project cashflows above, the Preqin dataset records the cash distributions from each fund net of fund manager fees, so no additional fee calculation is necessary.

We aim to keep the investors' average cash holdings close to zero, even though the cash calls by funds and cash distributions from funds are unpredictable. This assumes that investors sometimes borrow (or sell other non-private assets from elsewhere in their portfolios) in order to provide the cash called by a fund, but this assumption is required since positive average cash holdings would dilute investor returns, leaving us comparing the IRR generated on a fund fully exposed to PE with the return on a reinvested strategy with significant cash holdings or net borrowing.

Investors are assumed to collect distributions from their most recent fund for 4 years before making commitments to a new fund. At this point they commit a multiple of the cash they hold into a new fund. The value of this multiple (a=1.55) is chosen to give an average net cash holding close to zero across these simulations. By the time this cash is called, funds are likely to have received further distributions from their previous investment. For robustness we use different techniques for dealing with the short-term fluctuations in the cash held by the investor. Our first method calculates the IRR over the full investment horizon, including cash distributions and cash calls as positive/negative cashflows. This IRR might in principle be biased, but any such biases are likely to be small because the strategy remains on average fully invested in PE over this long (and almost entirely exogenous) horizon. By contrast, the IRRs published by funds will be affected by the very large variance (and covariance) of the project lives shown in Figure 1 above. Over 50,000 simulations, our reinvested investor IRR averaged 11.4%, compared to the average published IRR of 13.5%, suggesting a bias of just over 2.1%.[12]

As an alternative which avoids using an IRR, we instead assumed that cash balances were subject to a borrowing/lending rate of 10% until the end of the horizon. This generated a long term investor return (in effect a MIRR) which averaged 11.2% and suggested a bias of 2.2%. Assuming zero interest resulted in a lower long term investor return and hence a higher estimated bias. These simulations remain preliminary, since we are investigating alternative simulation techniques which might be less granular, and hence generate smaller absolute cash balances. However, whilst some simulated paths generate positive cash balances, others were negative (implying borrowing), and with the average cash holding across our simulations close to zero, there should be little net effect due to simple leverage differences.

These results are broadly consistent with the bias estimates that we derived in earlier sections. This is reassuring, since here we are using a different method and an entirely separate database of net fund cashflows. Furthermore, we know that these latest estimates are likely to be conservative, since they compare the IRRs of single funds against a reinvested strategy which is slightly better diversified (accumulated cash is reinvested into a single new fund, but the existence of residual cashflows from earlier funds will give an element of additional

[12] For VC funds the bias is larger, at around 4%, consistent with the greater variance of VC fund returns.

diversification). This will add some further "diversification return" to the returns of the reinvested strategy, and hence will reduce our estimated bias.

These estimates can be regarded as giving equal weight to all funds in our dataset, since our simulated investors make the required commitment to each new fund regardless of the size of this fund. For robustness we repeated this analysis using just the cashflows of the largest 50% of funds (thus removing the smaller funds that would arguably have been most overweighted). This generated marginally larger bias estimates.

Phalippou and Gottschalg (2008) notes that the IRRs recorded by individual funds are sometimes aggregated together into an average for the PE sector as a whole, but that this average will be a misleading indicator of the returns investors can expect to achieve since funds which record higher IRRs tend to be of relatively short duration. To correct for this (and the different sizes of funds) they weight fund IRRs by the PV of cash invested and the fund's duration, resulting in "a sort of IRR per year and per dollar invested" which is 2.42% per annum lower than the simple average of the fund IRRs. This differential is related to QWA bias to the extent that it is the covariance of the project returns and durations which leads to the observed covariance of fund IRRs and durations, although the latter could also result from sectoral, vintage or compositional differences within the dataset. However, QWA bias is distinct, since it not just a matter of aggregation − each individual fund IRR should be regarded as having potentially been boosted by QWA bias. Furthermore, duration-weighting fund IRRs simplifies each fund into a single cashflow returned after the calculated duration, thus removing some of the effect of the covariance of individual project returns and maturities *within* individual funds. It also removes the convexity effect identified above.

We have now estimated QWA and convexity bias both parametrically and by simulation, including simulations based on two entirely separate databases, and using a wide range of different assumptions. This is grounds for confidence that the biases that we have identified are robust and economically highly significant.

8. Alternative Return Measures - Biases in Other Measures of Annualized Returns

Having established that the IRR is systematically biased, we now investigate alternative measures that have been suggested for measuring PE returns. This gives a useful additional perspective on the underlying factors causing QWA bias.

The PME (Public Multiple Equivalent) derived by Kaplan and Schoar (2005), is widely used in academic studies, but seldom calculated by funds. It is defined as the ratio of two present values, using a time-varying discount rate which reflects the return achieved each period in a benchmark asset, typically the total return on an equity index such as the S&P500. A PME greater than one thus represents fund outperformance of this benchmark:

$$PME \equiv \frac{PV(total\ cash\ returned\ to\ investors)}{PV(total\ cash\ invested)} \tag{33}$$

The cash multiple (Multiple of Invested Capital, MOIC) can be seen as a special case of the PME where the discount rate is set to zero. This measure is widely cited by funds, although it cannot be directly compared to the annualized returns recorded for other asset classes. To investigate the properties of the PME, it is useful to make the same simplifying assumption that we did for the IRR in section 3: that all the investor cash (K_0) is called by the fund at time $t=0$. All cashflows returned to investors in subsequent periods are discounted back to $t=0$ using the time-varying discount rate M_t:

$$PME = \frac{1}{K_0}\sum_{t=1}^{T}\frac{d_t}{(1+M_1)\ldots(1+M_t)} \tag{34}$$

Substituting for d_t, using the identity $d_t = K_{t-1}(1+r_t) - K_t$:

$$PME = \frac{1}{K_0}\sum_{t=1}^{T}\frac{K_{t-1}(1+r_t)-K_t}{(1+M_1)\ldots(1+M_t)} \tag{35}$$

$$PME = \frac{1+r_1}{1+M_1} - \frac{K_1}{K_0(1+M_1)} + \frac{K_1(1+r_2)}{K_0(1+M_1)(1+M_2)} - \frac{K_2}{K_0(1+M_1)(1+M_2)} + \frac{K_2(1+r_3)}{K_0(1+M_1)(1+M_2)(1+M_3)} \ldots - \frac{K_{T-1}}{K_0(1+M_1)(1+M_2)\ldots(1+M_{T-1})} + \frac{K_{T-1}(1+r_T)}{K_0(1+M_1)(1+M_2)\ldots(1+M_T)} - \frac{K_T}{K_0(1+M_1)(1+M_2)\ldots(1+M_T)} \tag{36}$$

We can ignore the last term, since the investment horizon T will be extended until all assets have been written off or distributed to investors, so $K_T \equiv 0$.

$$PME = 1 + \frac{r_1-M_1}{1+M_1} + \frac{K_1(r_2-M_2)}{K_0(1+M_1)(1+M_2)} + \frac{K_2(r_3-M_3)}{K_0(1+M_1)(1+M_2)(1+M_3)} \ldots \frac{K_{T-1}(r_T-M_T)}{K_0(1+M_1)(1+M_2)\ldots(1+M_T)} \tag{37}$$

The PME is thus a weighted average of the excess returns (in excess of the benchmark return M_t) recorded by the fund in each period. Recall the corresponding equation for the IRR:

$$IRR = \sum_{t=1}^{T}\frac{r_t K_{t-1}}{(1+IRR)^{t-1}} \Big/ \sum_{t=1}^{T}\frac{K_{t-1}}{(1+IRR)^{t-1}} \tag{38}$$

The PME and IRR can both be regarded as weighted averages of the periodic returns r_t. In each case, the weight given to each periodic return is a function of the present value of the fund at the start of each period. However, the weights for the PME are functions of variables which are all simultaneous or prior to the periodic return in question, so these weights will not subsequently be revised as successive terms are added. By contrast, we can see that the IRR gives weights to each r_t which depend on asset values in all periods — including those subsequent to the return in question (equation 3). This leads to retrospective adjustment of these weights, and hence QWA bias. Thus the IRR exhibits QWA bias, but the PME does not. The same is true for the cash multiple, where $M_t=0$ for all t.[13]

However the weights on the excess returns in equation (37) sum to much greater than unity. This reflects the fact that the PME is a multiple, not an annualized return. By contrast, for a meaningful measure of *annualized* returns these weights must sum to unity, since a performance measure with $\Sigma w_t \neq 1$ would be a biased estimator of the expected excess return (assuming that returns are drawn from a stable distribution). But setting $\Sigma w_t=1$ introduces QWA bias, because it means that earlier w_t must be adjusted retrospectively as subsequent periodic returns and distributions are added into the calculation.

The weights given to later returns r_t must be a function of the cash distributions that have been made over periods 1 to t-1. Otherwise, substantial weight would be given to periodic returns even after the fund has already returned most of its assets to investors, leaving little value in the portfolio. To be meaningful, an annualised performance measure must possess two properties: that w_t should be a function of prior d_t, and that $\Sigma w_t=1$. However, together these conditions imply that the weight given to early period returns must also be a function of *subsequent* distributions. This retrospective adjustment will give rise to QWA bias if distributions are correlated with earlier returns, just as we saw above for the IRR.[14]

[13] In theory in a fund manager might be able to bias the MOIC: for example, if the first project in which the fund invested very rapidly generated an exceptionally large multiple, then the fund manager will know that subsequent investments are likely to dilute this exceptional multiple with more average values. Hence the fund's overall multiple could be boosted by abandoning any further investments and winding the fund up prematurely. However, this seems unlikely in practice, since the amount of cash committed to the fund by LPs is fixed in advance. Failure to invest this full amount is likely to disappoint the LPs, and would, of course, dramatically reduce the fees earned by the fund manager. Thus it seems unlikely that fund managers would choose to exploit this bias. This comes in dramatic contrast to the QWA bias analysed in this paper which does not depend on discretionary fund manager action, but is instead driven by the negative correlation between fund returns and duration shown in Figure 1, which appears to be a fundamental characteristic of the PE sector.

[14] As discussed earlier, reliable periodic valuations of PE projects are not available until the project matures and is realised as cash or listed equity. We can infer the average periodic return r_i on this project between initial investment and maturity, but we have no reliable means of measuring how this return might have varied over time (r_{it}). Nevertheless, the cross-sectional correlation shown in Figure 1 between the maturity of individual projects

This has a very important implication: there is no point in looking for alternative functions which, like the IRR, seek to measure annualized fund returns using only the fund's periodic cashflows. Any such return would require weights w_t which are adjusted to reflect cash calls and distributions and, coupled with the requirement $\Sigma w_t=1$, this requires retrospective adjustment of earlier weights. This leads to an important conclusion: that QWA bias does not arise from the particular functional form of the IRR ─ instead it follows directly from the properties that we would demand of any meaningful measure of annualized returns.

Consistent with this, it is straightforward to show that the various alternative measures of annualized returns that have been suggested within the PE industry all suffer from QWA bias. These have been suggested as means of generating IRRs that compare fund cashflows with those on listed equities, giving a measure of the annualized outperformance by the fund. Confusingly, the first of these alternative measures is also referred to as the public market equivalent (PME), but it is generally attributed to Long and Nickels, so we will refer to it here as PMELN. PMELN compares the IRR of the fund cashflows with the IRR that would have been recorded for a similarly-timed set of investments in listed equities. The value of each cash call made by the fund is assumed to evolve in line with the total return on an equity index such as S&P500. Subsequent cash distributions by the fund are treated as the sale of these equities, leaving a reduced asset value invested in the equity index. These cashflows plus the residual asset value at the end of the investment horizon are then used to calculate a benchmark IRR. The difference between the fund's conventional IRR and this benchmark IRR is taken as a measure of fund outperformance.

One known problem with this approach is that the cash distributions of a fund which significantly outperforms the equity index may generate a negative residual asset value, implying that the fund is then being benchmarked against a short position in the equity index. This problem has been addressed by introducing a scaling factor for the implied equity flows (e.g. the Capital Dynamics PME+ measure and the Cambridge Associates mPME).

A more fundamental question is whether these measures suffer from QWA bias. They each compare the fund IRR with the IRR of equivalently-timed investments in public equity.

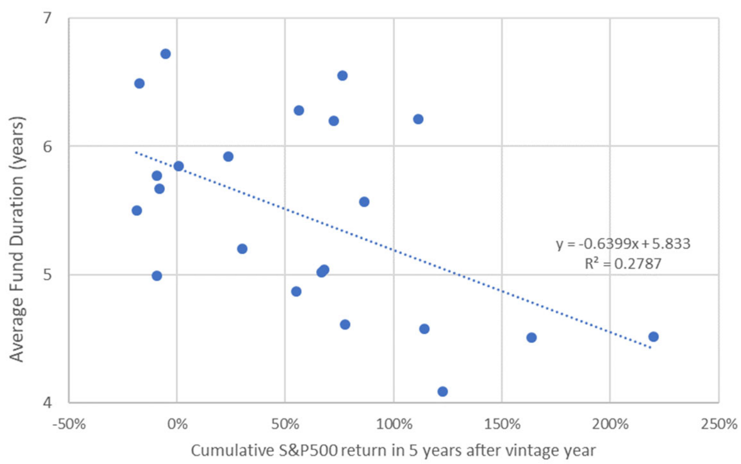
---

and their average return will itself translate into a time series correlation between r_t and d_t for the fund in aggregate. High return projects tend to mature quickly, and after they mature the average return on the fund's remaining projects will fall. Thus substantial returns of cash to investors on average tend to be followed by lower subsequent aggregate fund returns, leading to a positive QWA bias.

We know that the fund IRR suffers significant bias as a result of the correlation between project returns and project maturities. Only if the second component (the IRR of the public equity equivalent) is equally biased will these two biases net off, leaving the overall measure unbiased.

This requires that the covariance between the returns and durations of the projects within the fund must be entirely due to systematic risk rather than idiosyncratic risk. This seems extremely unlikely, since (i) annualized returns on PE projects show far greater volatility than publicly-listed equity (annualised standard deviation of returns on the S&P500 has been around 12% compared to the (maturity-weighted) standard deviation of project annualised returns of 64% in our dataset); (ii) estimates of the market beta for fund returns have been around unity (see section 2). Hence the majority of the volatility in project returns is idiosyncratic. It remains theoretically possible that the observed covariance of PE project returns with their maturities comes about because maturities covary massively with the market return, but are independent of idiosyncratic risk. To assess this, we can regress the average duration of funds within each vintage against subsequent equity index returns. Figure 4 shows that there is some systematic effect of strong market returns on durations, but this is far smaller than would be required to generate the observed covariance of project returns and durations without any effect from idiosyncratic returns (to generate the observed 64% standard deviation of returns, the regression coefficient would need to be greater than 5).

<u>Figure 4: Average Fund Duration by Vintage Year vs. Subsequent Equity Returns</u>



Direct Alpha is another performance metric used by practitioners. This is calculated by deflating each of the cashflows into and out of the fund by the cumulative return on the equity index up to that point. The IRR of these deflated cashflows is then calculated, and is interpreted as representing the annualized excess return generated by the fund. This figure will suffer from QWA bias for exactly the same reasons as PMELN, since the deflated project returns are likely to show a strong negative covariance with T_i. The fact that these alternative measures suffer from QWA bias shows that the sources of this bias have not previously been well understood.

Conclusion

The IRR has long been known to be sensitive to the impact of early cash returns. However the reasons for this have not previously been fully analyzed. We identify two systematic upward biases in the IRRs quoted by PE funds. These arise because the timing of the cash distributions made by funds to investors (i) is stochastic (resulting in convexity bias), and (ii) covaries with the returns achieved up to this date ("Quit Whilst Ahead" bias). By contrast, the returns on more liquid assets are calculated over exogenously fixed horizons, and so are free from these biases. Thus the IRRs quoted by PE funds are upwardly-biased compared with the returns on other asset classes such as listed equity, with which they are likely to be compared. Survey evidence clearly shows that investors continue to regard fund IRRs as a key factor when deciding to make commitments to funds.

We quantify the biases in PE IRRs using a range of parametric and simulation techniques. These show robust upward bias that raises a typical fund IRRs by an average of around 3% per annum. This is economically highly significant to investors, compared to the average net fund IRR of 12.2% per annum. These biases are not just the result of a small number of extreme datapoints: consistent with other studies we removed extreme returns and durations from our dataset of PE projects. We also reported the median biases in our simulations (the means are substantially higher, implying that individual fund IRRs can in some cases be substantially higher). We considered alternative scenarios which (i) removed projects with the longest lives; (ii) modelled the performance of individual deciles of the IRR distribution (consistent with persistent differentials in manager skill levels or the risk premia involved); (iii) checked our project-based simulations against an entirely separate dataset of net fund cashflows by modelling an investor reinvesting in successive funds.

QWA bias does not arise from the particular functional form of the IRR, it follows directly from the properties that we would demand of any meaningful measure of annualized returns. Variants on the IRR have been suggested by practitioners (Direct Alpha, ICM/PME, PME+, mPME). We show that these are similarly biased. The fact that practitioners have proposed such biased measures demonstrates that the biases in the IRR have not previously been properly understood.

ANNEX: Modelling Systematic Risk

In this section we check the robustness of our estimates of QWA bias above by explicitly modelling the extent to which the returns on all projects in which a fund invests might be affected by systematic risk. We return to our first order approximation of the IRR as simply a weighted average of the project returns, but this time systematic risk means that T_j will not be independent of r_{it}, since each T_j is correlated with $\sum_{t=1}^{t=j} r_{jt}$, and systematic risk means that each r_{jt} is correlated with r_{it}.

$$R - \mu_r \approx \frac{\sum_{i=1}^{N}\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)}{\sum_{i=1}^{N} T_i} = \sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\big/_{N\mu_T}}{1+\frac{\sum_{j=1}^{N-1,j\neq i}(T_j-\mu_T)}{N\mu_T}+\frac{T_i-\mu_T}{N\mu_T}}\right) \tag{A1}$$

$$\approx \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right)\left(1-\frac{\sum_{j=1}^{N-1,j\neq i}(T_j-\mu_T)}{N\mu_T}-\frac{T_i-\mu_T}{N\mu_T}\right) \tag{A2}$$

This expression identifies an additional element of bias reflecting the covariance of $\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)$ with the lives T_i of each of the *(N-1)* other projects in the fund. We can estimate this additional bias by approximating that the relationship between project durations and returns is linear: $(T_j-\mu_T) = k\sum_{t=1}^{t=T_j}(r_{jt}-\mu_r)+f_j$, where $k<0, f_j$ are *iid*, $E[f_j]=0$:

$$R - \mu_r \approx \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right)\left(1-\frac{\sum_{j=1}^{N-1,j\neq i}\left(k\sum_{t=1}^{t=j}(r_{jt}-\mu_r)+f_j\right)}{N\mu_T}-\frac{T_i-\mu_T}{N\mu_T}\right) \tag{A3}$$

We identify systematic risk by substituting $\sum_{t=1}^{t=T_j}(r_{jt}-\mu_r) = \sum_{t=1}^{t=T_j}\left(\beta(r_{mt}-\mu_{r_m})+e_{jt}\right)$, where r_{mt} is the market return, β is the CAPM coefficient, e_{jt} is *iid* and $E[e_{jt}]=0$.

$$R - \mu_r \approx \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}\beta(r_{mt}-\mu_{r_m})+e_{it}\right)\left(1-\frac{\sum_{j=1}^{N-1,j\neq i}\left(k\sum_{t=1}^{t=T_j}(\beta(r_{mt}-\mu_{r_m})+e_{jt})+f_j\right)}{N\mu_T}\right) - \frac{1}{N\mu_T}\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right)\left(\frac{T_i-\mu_T}{N\mu_T}\right) \tag{A4}$$

Taking expectations, noting that $E\left[\sum_{t=1}^{t=T_i}(r_{mt}-\mu_{r_m})\right]=0$, and that the idiosyncratic errors e_{it} and f_j are independent of r_{mt}:

$$E[R]-\mu_r \approx -\frac{k\beta^2}{N^2\mu_T^2}\sum_{i=1}^{N}\left(E\left[\sum_{t=1}^{t=T_i}(r_{mt}-\mu_r)\sum_{j=1}^{N,j\neq i}\sum_{t=1}^{t=T_j}(r_{mt}-\mu_r)\right]\right) - \frac{1}{N^2\mu_T^2}\sum_{i=1}^{N}\left(E\left[(T_i-\mu_T)\sum_{t=1}^{t=T_i}(r_{it}-\mu_r)\right]\right) \tag{A5}$$

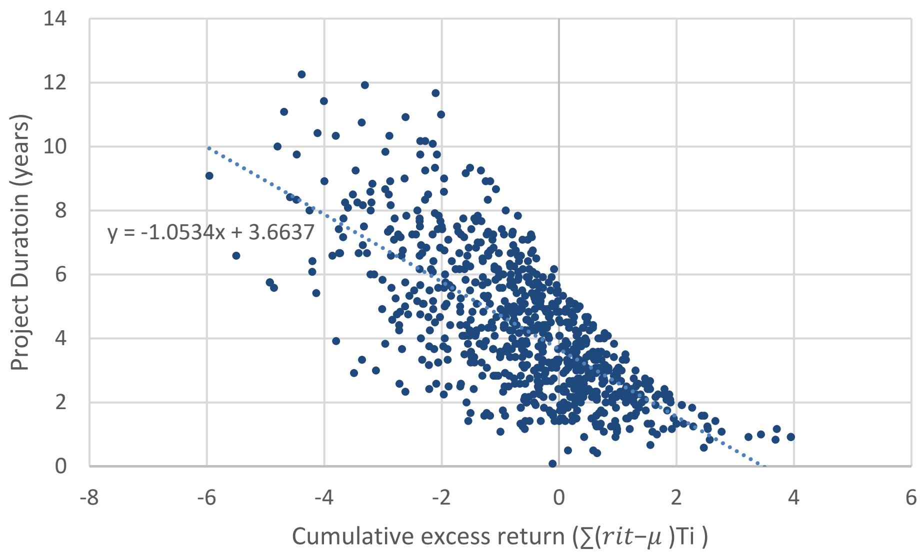
Thus a component of the bias is related to the variance of the market return. We can approximate that $E\left[\sum_{t=1}^{t=T_i}(r_{mt}-\mu_r)\sum_{t=1}^{t=T_j}(r_{mt}-\mu_r)\right] \approx \mu_T^2\sigma_{r_m}^2$, giving:

$$E[R] - \mu_r \approx -\frac{(N-1)k\beta^2}{N}\sigma_{r_m}^2 - \frac{1}{N\mu_T^2}cov\left(T_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right) \quad \text{(A6)}$$

This expression decomposes the bias into systematic and idiosyncratic risk components. As noted in the literature survey, previous studies report β estimates ranging either side of unity, so we use β=1. Simple OLS estimation of our linear model of project durations gives a coefficient *k*= -1.05 (Figure A1) and, as above, $cov\left(T_i, \sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right)$= -2.30. Taking *N*=6 gives an estimated bias of 3.9% (*N*=9 implies 3.0%).

This is a useful cross-check, although we should regard it as a likely overestimate of QWA bias, since (i) the squared residuals in $E\left[\sum_{t=1}^{t=T_i}(r_{mt} - \mu_r)\sum_{t=1}^{t=T_j}(r_{mt} - \mu_r)\right]$ will only be variances for *min(Ti, Tj)*, since the excess market returns during the period where only one of the two projects is still live will be uncorrelated with the earlier returns (given our assumption of serial independence); and (ii) high market returns tend to reduce both T_i and T_j as project exits are achieved more quickly (although this endogeneity will be of limited magnitude since the majority of the volatility of project returns is idiosyncratic).

Figure A1: Project Durations and Cumulative Excess Returns



Source: MergerMarket, with debt assumptions imposed as discussed in Section 4.

More generally, we could model the bias as a function of a number of independent risk factors: $E[R] - \mu_r \approx -k\beta_1^2\sigma_{r_{m1}}^2 - k\beta_2^2\sigma_{r_{m2}}^2 - k\beta_3^2\sigma_{r_{m3}}^2 \ldots$ The corresponding portfolio variance is $\beta_1^2\sigma_{r_{m1}}^2 + \beta_2^2\sigma_{r_{m2}}^2 + \beta_3^2\sigma_{r_{m3}}^2$ ..., so our assumption that T_i is a linear function of $\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)$ implies that the bias is simply proportional to the portfolio variance: $E[R] -$

$\mu_r \approx -k\sigma_{r_p}^2$. Indeed, we could derive this result directly by substituting $T_i - \mu_T = k\sum_{t=1}^{t=T_i}(r_{it} - \mu_r) + f_i$ into equation (A3) above:

$$E[R - \mu_r] \approx \frac{-k}{N^2\mu_T^2} E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{t=T_i}(r_{it} - \mu_r)\right)\left(\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{t=j}(r_{jt} - \mu_r) + f_j\right)}{N\mu_T}\right)\right] \approx -k\sigma_{r_p}^2 \quad \text{(A7)}$$

This allows us a simpler approach to estimating the bias. It is based on the restrictive assumption of a linear relationship between T_i and r_i, but it has the advantages (i) that it allows us to be completely agnostic about the risk factors which generate this variance; (ii) that we can apply it to the observed variance of net fund IRRs, thus to some extent taking into account the non-linear effect of performance-related fees, which will tend to make net returns less volatile than gross returns. We use the same estimate of k as above (-1.05), since the aggregate sensitivity of project durations to these unspecified risk factors must be the same as their average sensitivity to project returns r_{it}, regardless of how we choose to decompose the return volatility. Combining this with the observed variance of net fund IRRs (standard deviation=13.1%, see Table 4) gives us a bias of 1.8%. However, k was of necessity estimated using gross project returns. To the extent that the option-like fee structure reduces the variance of net IRRs, the regression coefficient would need to be larger to be consistent with the observed variance of T_i. Thus the bias estimate this gives us should be regarded as an underestimate.

Our bias estimates inevitably vary when we use different techniques. These estimates were based on a number of assumptions that might be imprecise, especially given the extreme volatility of PE project returns. However, even with these inherent uncertainties, our estimates are fairly consistent, ranging from a likely underestimate of 1.8% per annum to a likely overestimate of 3.9%.

REFERENCES

Albertus, J. F., & Denes, M. (2019). Distorting private equity performance: The rise of fund debt. Unpublished Working Paper.

Ang, A., Chen, B., Goetzmann, W. N., & Phalippou, L. (2018). Estimating private equity returns from limited partner cash flows. The Journal of Finance, 73(4), 1751-1783.

Azlen,M W and Zermati, I "Investing Like the Harvard and Yale Endowment Funds" Alternative Investment Analyst Review Vol 6, issue 3 2017

Baquero, G., Ter Horst, J., & Verbeek, M. (2005). Survival, look-ahead bias, and persistence in hedge fund performance. Journal of Financial and Quantitative analysis, 40(3), 493-517.

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

Braun, R., Jenkinson, T. & Stoff, I. 2017, "How persistent is private equity performance? Evidence from deal-level data", Journal of financial economics, vol. 123, no. 2, pp. 273-291.

Brown, G. W., Gredil, O. R., & Kaplan, S. N. (2019). Do private equity funds manipulate reported returns?. *Journal of Financial Economics*, *132*(2), 267-297.

Brown, G., Harris, R., Jenkinson, T., Kaplan, S., Robinson, D., 2015. What do different commercial data sets tell us about private equity performance? Unpublished working paper. University of North Carolina.

Brown, G. W., Harris, R. S., Hu, W., Jenkinson, T., Kaplan, S. N., & Robinson, D. T. (2020). Private equity portfolio companies: A first look at Burgiss holdings data. Available at SSRN 3532444.

Cardie, J. H., Cattanach, K. A., & Kelly, M. F. (2000). How Large Should Your Commitment to Private Equity Really Be? The Journal of Wealth Management, 3(2), 39-45.

Da Rin, M., & Phalippou, L. (2017). The importance of size in private equity: Evidence from a survey of limited partners. Journal of Financial Intermediation, 31, 64-76.

de Zwart, G., Frieser, B., & van Dijk, D. (2012). Private equity recommitment strategies for institutional investors. Financial Analysts Journal, 68(3), 81-99.

Dichev, I. D. "What Are Stock Investors' Actual Historical Returns? Evidence from Dollar-Weighted Returns." *American Economic Review*, 97 (2007), 386-401.

Dichev, Ilia D., and Gwen Yu. "Higher risk, lower returns: What hedge fund investors really earn." *Journal of Financial Economics* 100, no. 2 (2011): 248-263.

Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). Survivor bias and mutual fund performance. The review of financial studies, 9(4), 1097-1120.

Friesen, Geoffrey C., and Travis RA Sapp. "Mutual fund flows and investor returns: An empirical examination of fund investor timing ability." *Journal of Banking & Finance* 31, no. 9 (2007): 2796-2816.

Fung, W., & Hsieh, D. A. (2009). Measurement biases in hedge fund performance data: an update. Financial Analysts Journal, 65(3), 36-38.

Gompers, P., Kaplan, S. N., & Mukharlyamov, V. (2016). What do private equity firms say they do? *Journal of Financial Economics*, *121*(3), 449-476.

Harris, R. S., Jenkinson, T., & Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, *69*(5), 1851-1882.

Hayley, Simon. "Hindsight effects in dollar-weighted returns." *Journal of Financial and Quantitative Analysis* 49, no. 1 (2014): 249-269.

Hüther, Niklas. "Raising funds on performance: Are private equity returns too good to be true." Unpublished working paper (2016).

Ingersoll, J.; M. Spiegel; W. Goetzmann and I. Welch. "Portfolio Performance Manipulation and Manipulation-Proof Performance Measures." *Review of Financial Studies,* 20 (2007),1503-1546.

Jegadeesh, N., Kräussl, R., & Pollet, J. M. (2015). Risk and expected returns of private equity investments: Evidence based on market prices. The Review of Financial Studies, 28(12), 3269-3302.

Jenkinson, T., Kim, H., & Weisbach, M. S. (2021). Buyouts: A primer (No. w29502). National Bureau of Economic Research.

Jenkinson, T., Sousa, M., Stucke, R., 2013. How fair are the valuations of private equity funds? Unpublished working paper. Oxford University *Available at SSRN 2229547*.

Kaplan, Steven N., and Antoinette Schoar. "Private equity performance: Returns, persistence, and capital flows." The journal of finance 60, no. 4 (2005): 1791-1823.

Korteweg, A., & Sorensen, M. (2017). Skill and luck in private equity performance. *Journal of Financial Economics*, *124*(3), 535-562.

Larocque, S., Shive, S., & Stevens, J. S. (2022). Private Equity Performance and the Effects of Cash-Flow Timing. The Journal of Portfolio Management, 48(9), 86-102.

Long, A. M., & Nickels, C. J. (1996). A private investment benchmark. Working paper.

Lopez-de-Silanes, F., Phalippou, L. & Gottschalg, O. 2015, "Giants at the Gate: Investment Returns and Diseconomies of Scale in Private Equity", Journal of financial and quantitative analysis, vol. 50, no. 3, pp. 377-411

Metrick, Andrew, and Ayako Yasuda. "The economics of private equity funds." *The Review of Financial Studies* 23, no. 6 (2010): 2303-2341.

Nevins, D., Conner, A., & McIntire, G. (2004). A portfolio management approach to determining private equity commitments. The Journal of Alternative Investments, 6(4), 32-46.

Oberli, A. (2015). Private Equity Asset Allocation: How to Recommit?. The Journal of Private Equity, 18(2), 9-22.

Phalippou, L. "The Hazards of Using IRR to Measure Performance: The Case of Private Equity." *Journal of Performance Measurement*, 12 (2008), 55-66.

Phalippou, L. "Private Equity Laid Bare". (2017) Createspace Independent Publishing Platform

Phalippou, L. "An Inconvenient Fact: Private Equity Returns and the Billionaire Factory." *The Journal of Investing* 30.1 (2020): 11-39.

Phillips, B., Pukthuanthong, K., & Rau, P. R. (2012). Do Mutual Fund Investors Chase False Returns? *The 2012 Annual Meeting of the Academy of Behavioral Finance & Economics (p. 111)*.

Rabener, Nicolas. "Private Equity Is Still Equity, Nothing Special Here." *The Journal of Investing* 30.1 (2020): 89-99.

Schillinger, P., Braun, R., & Cornel, J. (2019). Distortion or cash flow management? Understanding credit facilities in private equity funds. Understanding Credit Facilities in Private Equity Funds.

Sirri, E. R., & Tufano, P. (1998). Costly Search and Mutual Fund Flows. *Journal of Finance* 53(5), 1589-1622.

Ter Horst, J. R., Nijman, T. E., & Verbeek, M. (2001). Eliminating look-ahead bias in evaluating persistence in mutual fund performance. Journal of Empirical Finance, 8(4), 345-373.