



DEGREE PROJECT IN MATHEMATICS,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2018

True risk of illiquid investments

HARALD AGERING

True risk of illiquid investments

HARALD AGERING

Degree Projects in Mathematical Statistics (30 ECTS credits)
Degree Programme in Applied and Computational Mathematics (120 credits)
KTH Royal Institute of Technology year 2018
Supervisor at COIN: Joakim Ahlinder
Supervisor at KTH: Anja Janssen
Examiner at KTH: Anja Janssen

TRITA-SCI-GRU 2018:344
MAT-E 2018:72

Royal Institute of Technology
School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci

Abstract

Alternative assets are becoming a considerable portion of global financial markets. Some of these alternative assets are highly illiquid, and as such they may require more intricate methods for calculating risk and performance statistics accurately. Research on hedge funds has established a pattern of risk being understated and various measures of performance being overstated due to illiquidity of the assets. This paper sets out to prove the existence of such bias and presents methods for removing it. Four mathematical methods aiming to adjust statistics for sparse return series were considered, and an implementation was carried out for data on private equity, real estate and infrastructure assets. The results indicate that there are in general substantial adjustments made to the risk and performance statistics of the illiquid assets when using these methods. In particular, the volatility and market exposure were adjusted upwards while manager skill and risk-adjusted performance were adjusted downwards.

Verklig risk hos illikvida investeringar

Sammanfattning

Alternativa tillgångsslag börjar utgöra en avsevärd del av globala finansiella marknader. Vissa av dessa alternativa tillgångsslag är mycket illikvida och kan som sådana kräva mer avancerade metoder för att beräkna nyckeltal för risk och utveckling mer korrekt. Forskning på hedgefonder har kunnat påvisa ett mönster där risk underskattas medan olika nyckeltal för utveckling överskattas till följd av tillgångarnas illikviditet. Målet med denna artikel är att påvisa förekomsten av sådan systematisk avvikelse samt att presentera metoder för att avlägsna densamma. Fyra matematiska metoder framtagna för att justera nyckeltal för glesa dataserier användes, och metoderna implementerades på data för tillgångar i private equity, fastigheter samt infrastruktur. Resultaten antyder att det generellt sett sker betydande justeringar av nyckeltalen för risk och utveckling för de illikvida tillgångsslagen när man tillämpar dessa metoder. Mer specifikt justerades volatiliteten och marknadsexponeringen uppåt medan förvaltarens förmåga och den riskjusterade avkastningen justerades nedåt.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Background | 1 |
| 1.2 | Purpose and Aim | 4 |
| 2 | Mathematical Theory | 4 |
| 2.1 | Market model | 4 |
| 2.2 | Nonsynchronous trading | 5 |
| 2.2.1 | Method of Scholes and Williams (1977) | 8 |
| 2.2.2 | Method of Dimson (1979) | 11 |
| 2.3 | Smoothed returns | 13 |
| 2.3.1 | Method of Getmansky et al. (2004) | 14 |
| 2.3.2 | Method of Okunev and White (2003) | 16 |
| 3 | Data | 18 |
| 3.1 | Private equity | 19 |
| 3.2 | Real estate | 19 |
| 3.3 | Infrastructure | 20 |
| 3.4 | Market Index | 20 |
| 3.5 | Risk-free asset | 20 |
| 4 | Results | 21 |
| 4.1 | Adjusted performance statistics | 21 |
| 4.2 | Portfolio optimization | 24 |
| 5 | Discussion | 28 |
| 6 | Conclusion | 30 |
| 7 | Acknowledgements | 30 |

1 Introduction

In recent years there has been a persisting trend on global financial markets of an increase in the allocation to non-traditional asset classes (Truong et al, 2015). Common to many of these alternative assets is that they trade infrequently, i.e they are illiquid. Examples of such assets are investments in private equity, real estate and infrastructure. Illiquid assets are valued infrequently which leads to sparse return time series. This in turn introduces a possibly severe source of bias in many of the conventional risk and performance measures. In general, risk is understated while performance is overstated.

Nonetheless, little has been done outside of academia in order to investigate the severity of this possible bias and to seek a remedy. As an investor it is easy to be impressed by the performance statistics of certain alternative investment opportunities while not being aware of the problems related to the exposition. Hence it is of interest to investigate whether there is a significant bias when applying conventional measures to illiquid assets and, if so, to find means of adjustment.

1.1 Background

The alternatives sector as a whole, including anything aside from the classic asset classes of bonds, stocks and certificates, is becoming increasingly popular. In fact, global alternative investments doubled between 2005 and 2011, reaching a total AUM of 6.5 trillion USD and implying a 14% annual growth rate for that period (McKinsey & Company, 2012). And the sector continues to grow. Some forecasts put the total AUM above 18 trillion USD at year 2020 (Truong et al, 2015).

Part of the explanation as to why alternatives have become more desirable could be related to low interest rates on some markets. However, traditional instruments losing their appeal to investors is unlikely to be the only reason for this re-distribution of assets. Rather, many of the alternative assets have outstanding performance statistics which naturally attracts investors. Often the volatility is significantly lower than for comparable traditional assets and at the same time the correlation with traditional assets is generally low. This indicates, as Pedersen et al. (2014) put it, that it seems like "alternative asset classes and strategies represent somewhat of

a free lunch". Such performance statistics are, however, usually a simplification.

Some of the quickly growing asset classes within the alternatives sector are connected to areas of infrequent trading and valuation. Examples of such illiquid investments are private equity, private debt, real estate, infrastructure, natural resources, venture capital etc. Looking at alternative assets it is easy to forget that a lot of them are in fact highly illiquid, being marked to market only on a quarterly or even annual basis. This naturally leads to sparse time series and is something to account for when calculating classic performance statistics such as alpha, beta, volatility and Sharpe ratio. As will be proven in Section 2, the typical case is that the volatility and beta are understated whereas the Sharpe ratio and alpha are overstated when conventional measures are used on sparse time series without adjustment. This is obviously problematic since performance statistics skewed in this way would typically indicate more attractive investment opportunities. Nevertheless, it is hard to get hold of performance statistics that have been adjusted to account for illiquidity, at least outside of academia.

The issue of seldom traded assets and sparse time series is far from new and has received some attention within academia. Already in the 1960's the works of Fisher (1966) set in motion the field referred to as nonsynchronous trading. It refers to the problems that arise when several securities are valued at different times but treated as if they were valued simultaneously. In particular, a security which traded earlier than the quoted time did not have access to the same information as any security actually traded on the quoted time. This introduces errors in the variables and bias in many regular measures. Most of all it introduces serial correlation in the return series since not all relevant information will be captured in the corresponding time period, and will instead appear with a lag in a subsequent time period. This is serial correlation in a nutshell. The issue is particularly pressing for illiquid assets which are seldom valued and hence subject to potentially very large differences between the time of the actual trade and the time quoted for trade. This issue is described in more detail in Section 2.2.

Some of the works within the field of nonsynchronous trading are still highly relevant. For instance the papers of Dimson (1979) and Scholes and Williams (1977) suggest adjustments of illiquid return series in ways which are in no way outdated. Both of their approaches manage to remove a large portion of the serial correlation and provide adjusted, consistent estimators.

Their methods are addressed in Section 2.2.

There has also been research on aspects of illiquidity which are not directly connected to the theory of nonsynchronous trading. For instance, there has been a significant research on the area of hedge funds. Asness et al. (2001) set out to remove the serial correlation exhibited in hedge fund returns. Relying on the theory founded by Dimson (1979) and Scholes and Williams (1977) they found that lagged coefficients of the market index in many cases provide explanatory power to return series of market neutral hedge funds. As a result, the original market exposures were found to be understated. Another example illustrating the problems of illiquidity is illuminated by Lo (2002). There it is shown that applying the regular formulas to a sparse hedge fund return series will overstate the Sharpe ratios. In some cases the ratios were inflated by more than 50 %.

These more recent works on illiquidity will be represented here by two papers. The first is by Getmansky et al. (2004) which propose a model almost identical to the one used by proponents of nonsynchronous trading, although with a slightly different motivation. They recognize that nonsynchronous trading is a relevant issue for hedge funds and other investments with similar trading patterns. However, they conclude that nonsynchronous effects alone can not explain all of the serial correlation and, consequently, the bias in performance measures found in return series. Rather, they propose that serial correlation in hedge fund return series can be ascribed to a more general form of illiquidity. The second paper is by Okunev and White (2003), where yet another approach is suggested. The model is based on theory originally developed for real estate appraisals and hence deals more with subjective estimations than actual market values. However, the model has proven very successful on hedge fund data.

This paper will present and contrast different takes on illiquidity and the issue of how to adjust conventional measures to better fit reality when assets are traded infrequently. The methods of Dimson (1979), Scholes and Williams (1977), Getmansky et al. (2004) and Okunev and White (2003) will be examined in detail and have their methods evaluated on up-to-date data. Each approach will be dedicated a section presenting the theory and another section evaluating the method on data.

1.2 Purpose and Aim

This paper sets out to investigate how conventional performance and risk measures can be adjusted to enable more fair comparisons between liquid and illiquid assets. There is existing theory to do this, much of which has also proven useful on hedge fund data. However, there are not yet any well known applications to the kind of illiquid assets treated in this paper. Hence, the aim of this work is to present relevant theory and evaluate it on alternative assets data.

2 Mathematical Theory

This section presents the theory from four prominent papers on the subject of adjusted return series. There is a subdivision between these papers where the works of Dimson (1979) and Scholes and Williams (1977) sort as *Nonsynchronous trading* while the works of Getmansky et al. (2004) and Okunev and White (2003) sort as *Smoothed returns*. The first subsection will state the basic market model which is common to all of the methods discussed subsequently. After that an exposé of the different methods is presented, each with its own assumptions and suggested formulas based on the common market model.

2.1 Market model

There are different ways of modelling a financial market and the returns of assets traded on the market. One common way to model the return R_t of an asset is to assume a simple linear model

$$R_t = \alpha + \beta M_t + \epsilon_t. \quad (1)$$

Here, R_t is the true return, α is the intercept, β represents the correlation with the market index M_t and ϵ_t is an error term. The return R_t is referred to as *true* since it represents the actual return of an asset during a specified time span, as opposed to the *observed* return which for various reasons might differ from the true return. The details will be laid out in the upcoming section. The market index M_t represents the return of an index relevant to the market in question. This model composes the core of the theory in each of the methods discussed below. Assumptions regarding distributions and correlations between the constituents

of (1) differ slightly between the methods and will therefore be presented in the corresponding subsections.

The real issue is not associated with this particular choice of model, which is prevailing through much of the literature and is not very controversial. In fact, Stapleton and Subrahmanyam (1983) showed that a linear one-factor market model such as (1) is sufficient for deriving the widely accepted relationship between returns and market beta in the Capital Asset Pricing Model introduced in its original form by Sharpe (1964). Instead, the problem is that in reality one does not observe the true returns R_t . Rather, what we measure are the observed returns R_t^o . It is with respect to the observed returns different authors have devised various assumptions and models. In particular, it is an open question concerning how to best link the observed returns to the true returns.

The following sections include methods which suggest different solutions to the problem of adjusting the risk and performance measures with respect to illiquidity by assuming certain relations between observed and true returns. These methods have been divided into the subgroups of nonsynchronous trading and smoothed returns.

2.2 Nonsynchronous trading

As mentioned briefly in Section 1, nonsynchronous trading is related to the case where assets are valued at different times but treated as if they were all valued at the same time. Based on this there will arise a deviance between true and observed returns. Adopting the notation of Section 2 in Scholes and Williams (1977) the setting is the following. Suppose that there are $n = 1, \dots, N$ securities and that true returns are to be calculated over time intervals $[t - 1, t]$ for $t = 1, \dots, T$. However, a security typically trades at stochastic discrete times and has a price quoted only at the time of the actual trade. Such discontinuous trading patterns introduce errors in the market model (1).

To be more specific, consider any time interval $[t - 1, t]$. During any such time window there is either a trade and a corresponding price is set, or there is no trade. If a trade takes place it is assumed that it is conducted at a random time $t - s_{nt}$ where s_{nt} is the remainder of time period t in which no trade occurs for security n and $0 \leq s_{nt} \leq 1$ (Figure 1). If there is, for example,

a trade reported in two consecutive periods for security n , then the *observed* rate of return r_{nt}^s can be calculated over the time period $[t - 1 - s_{nt-1}, t - s_{nt}]$. The crucial observation is that r_{nt}^s in general differs from the *true* return series r_{nt} which is taken over $[t - 1, t]$.

The issue of nonsynchronous trading is still relevant, and maybe the issue is more pressing than ever with a market where highly illiquid alternatives are gaining ground. The more illiquid the security, the longer the possible space between actual trade and ascribed time of trade, hence increasing the risk for a large deviance between observed and true returns. There has been research on the subject of nonsynchronous trading and two of the most cited works are presented here.

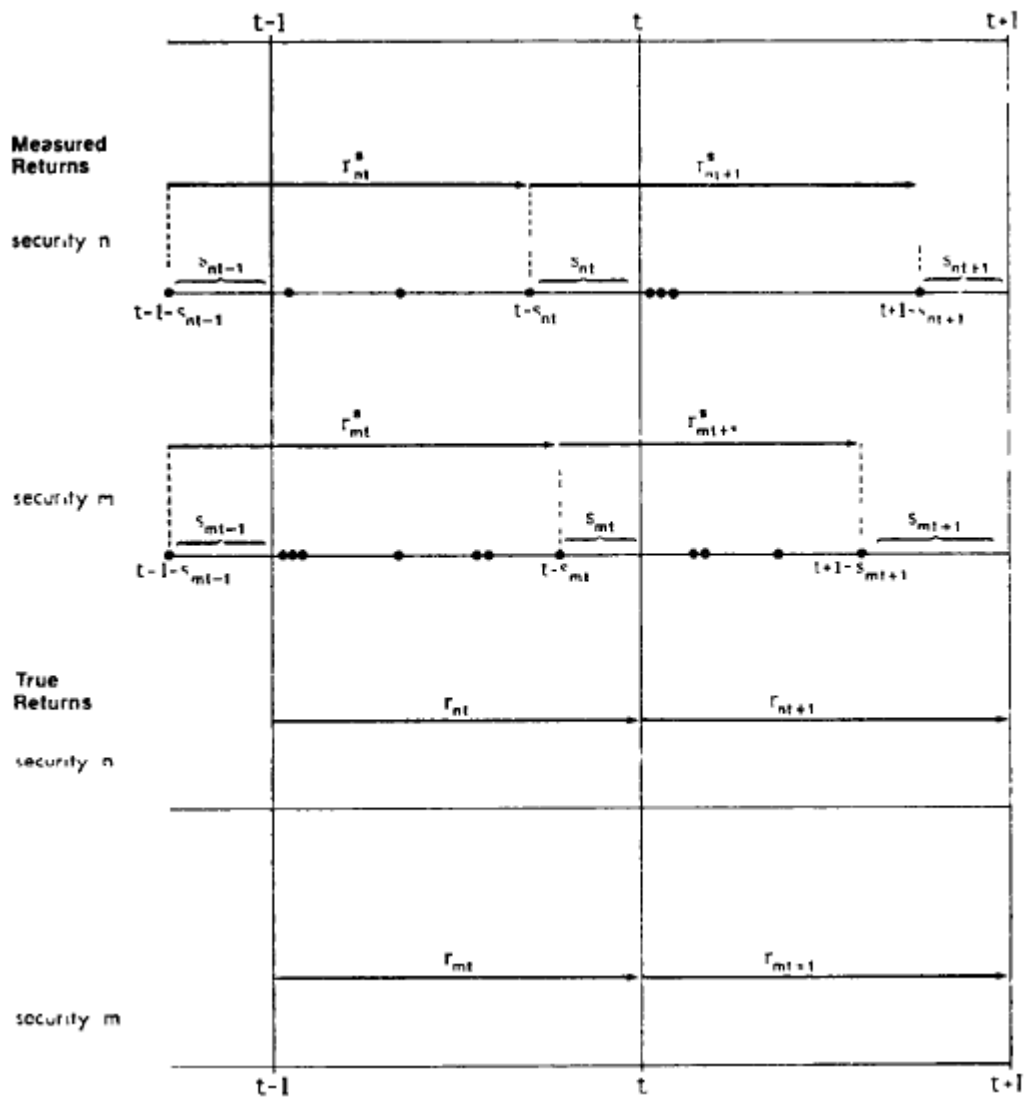


Figure 1: How nonsynchronous trading splits measured (observed) returns and true returns. Picture from Scholes and Williams (1977).

2.2.1 Method of Scholes and Williams (1977)

Scholes and Williams stress the assumption that the prices of securities are following a lognormal distribution in the market model (1). Assuming $n = 1, \dots, N$ securities and time periods $t = 1, \dots, T$, this means that the true returns R_{nt} are joint normally distributed with constant means μ_n , constant variances σ_n^2 and constant covariances σ_{nm} . (Note that we abandon the notation given in Figure 1 in order to conform with the rest of the paper.) This yields the market model

$$R_{nt} = \alpha_n + \beta_n M_t + \epsilon_{nt} \quad (2)$$

which is simply the market model (1) considered for each of the n securities. The coefficients are as usual defined as $\alpha_n = \mu_n - \beta_n \mu_M$ and $\beta_n = \sigma_{nM} / \sigma_M^2$, where subscript M denotes the market index.

Due to nonsynchronous trading, observed returns will in general differ from true returns (Figure 1). The observed returns are generated by another process

$$R_{nt}^o = \alpha_n^o + \beta_n^o M_t^o + \epsilon_{nt}^o \quad (3)$$

where the coefficients are given by $\alpha_n^o = E[R_{nt}^o] - \beta_n^o E[M_t^o]$ and $\beta_n^o = \text{Cov}(R_{nt}^o, M_t^o) / \text{Var}(M_t^o)$. These expressions do in general not coincide with the corresponding expressions appearing in (2).

The difference between equations (2) and (3), i.e. the difference between true and observed returns, will now be investigated more thoroughly. If we impose the further assumption that the non-trading periods $S_t \equiv (s_{1t}, s_{2t}, \dots, s_{Nt})$ are independent and identically distributed (IID) random variables it can be shown that the expectation of observed returns is unchanged with respect to true returns, whereas the variance of the observed returns in general differ from the variance of true returns. Beginning with the expected value, the derivation makes use of the law of total expectation by conditioning on the periods of non-trading S_t . After realizing that the period of trading corresponding to R_{nt}^o is $1 - s_{nt} + s_{nt-1}$ (see Figure 1) the derivation goes

$$\begin{aligned} E[R_{nt}^o] &= E[E[R_{nt}^o | S_t, S_{t-1}]] = \\ E[(1 - s_{nt} + s_{nt-1})\mu_n] &= (1 - E[s_{nt} - s_{nt-1}])\mu_n = \mu_n. \end{aligned} \quad (4)$$

This proves that the expected value of observed returns is indeed equal to the expected value of true returns. The calculation of the variance is a bit more involved. First we notice that the period of overlap between two observed returns R_{nt}^o and R_{mt}^o , conditional on S_t and S_{t-1} , has a length given by

$$\begin{aligned} & \min(1 - s_{nt}, 1 - s_{mt}) + \min(s_{nt-1}, s_{mt-1}) \\ & = 1 - \{\max(s_{nt}, s_{mt}) - \min(s_{nt-1}, s_{mt-1})\}. \end{aligned} \quad (5)$$

Using this expression for the time overlap and invoking the law of total covariance we get

$$\begin{aligned} \text{Cov}(R_{nt}^o, R_{mt}^o) &= E[\text{Cov}(R_{nt}^o, R_{mt}^o | S_t, S_{t-1})] + \\ & \text{Cov}(E[R_{nt}^o | S_t, S_{t-1}], E[R_{mt}^o | S_t, S_{t-1}]) \\ &= E[(1 - \{\max(s_{nt}, s_{mt}) - \min(s_{nt-1}, s_{mt-1})\})\sigma_{nm}] + \\ & \text{Cov}((1 - s_{nt} + s_{nt-1})\mu_n, (1 - s_{mt} + s_{mt-1})\mu_m) \\ &= (1 - E[\max(s_{nt}, s_{mt}) - \min(s_{nt-1}, s_{mt-1})])\sigma_{nm} + \\ & \text{Cov}(s_{nt} - s_{nt-1}, s_{mt} - s_{mt-1})\mu_n\mu_m. \end{aligned} \quad (6)$$

The covariance between any lagged returns is found by changing the expression for the time overlap in this derivation. Introducing the coefficient of variation $\nu_n \equiv \sigma_n/\mu_n$, the correlation coefficient $\rho_{nm} \equiv \sigma_{nm}/\sigma_n\sigma_m$ and simplifying the covariance term in (6) the relationship can be written more compactly as

$$\begin{aligned} \text{Cov}(R_{nt}^o, R_{mt}^o) &= \{1 - E[\max(s_{nt}, s_{mt}) - \min(s_{nt}, s_{mt})]\} + \\ & 2 \text{Cov}(s_{nt}, s_{mt})/(\rho_{nm}\nu_n\nu_m)\}\sigma_{nm}. \end{aligned} \quad (7)$$

From this expression it is apparent that the observed covariances in general differ from the true ones. In order to find an expression for the variance we simply set $n = m$ in (7) to get

$$\text{Var}(R_{nt}^o) = \{1 + 2 \text{Var}(s_{nt})/\nu_n^2\}\sigma_n^2. \quad (8)$$

Again, this shows that there in general is a discrepancy between observed and true variances.

Having established that there are errors present in the observed return series the next step is naturally to develop a remedy. Performing an ordinary least squares fit directly on the observed

return series would yield biased and inconsistent estimates. The suggested remedy involves the following additional coefficients:

$$\beta_n^{o-} \equiv \frac{\text{Cov}(R_{nt}^o, M_{t-1}^o)}{\text{Var}(M_{t-1}^o)} \quad (9)$$

$$\beta_n^{o+} \equiv \frac{\text{Cov}(R_{nt}^o, M_{t+1}^o)}{\text{Var}(M_{t+1}^o)} \quad (10)$$

$$\rho_M^o \equiv \frac{\text{Cov}(M_t^o, M_{t-1}^o)}{\text{std}(M_t^o) \text{std}(M_{t-1}^o)}. \quad (11)$$

The objective is to be able to find a relationship which links the observed and true coefficients. The derivation is lengthy and only the key step will be shown here. Starting from the expression (7) of the covariance we get

$$\begin{aligned} \text{Cov}(R_{nt}^o, R_{mt}^o) &= \{1 - E[\max(s_{nt}, s_{mt}) - \min(s_{nt}, s_{mt})]\} \sigma_{nm} + 2 \text{Cov}(s_{nt}, s_{mt}) \mu_n \mu_m \\ &= \sigma_{nm} - \{E[\max(s_{nt} - s_{mt}, 0)] \sigma_{nm} + \text{Cov}(s_{nt}, s_{mt}) \mu_n \mu_m\} - \\ &\quad \{E[\max(s_{mt} - s_{nt}, 0)] \sigma_{nm} + \text{Cov}(s_{nt}, s_{mt}) \mu_n \mu_m\} \\ &= \text{Cov}(R_{nt}, R_{mt}) - \text{Cov}(R_{nt}^o, R_{mt-1}^o) - \text{Cov}(R_{nt-1}^o, R_{mt}^o). \end{aligned} \quad (12)$$

To see why the last equation holds, simply redo the derivation (6) but for new expressions of the time overlap. The important observation is that this equation provides a relationship between true and observed returns. It can also be used to find adjusted volatility and in turn the adjusted Sharpe ratio. What remains of this derivation is a large portion of algebra using the expressions (9)-(12) along with the definitions of α_n^o and β_n^o in (3). However, we skip straight to the result which is

$$\alpha_n^o = \alpha_n + (\beta_n - \beta_n^o) \mu_M \quad (13)$$

and

$$\beta_n^o = \beta_n - (\beta_n^{o-} + \beta_n^{o+} - 2\beta_n \rho_M^o). \quad (14)$$

These equations provide an explicit relationship between true and observed coefficients. Before solving for the consistent estimators, consider the equations (13)-(14) along with the definitions (9)-(10). For a security n which trades infrequently the covariance with a lagged market index

is likely to be large since all relevant information might not be available at the contemporaneous time. Remember Figure 1 which illustrates this issue. This implies a large β_n^o and in turn that $\beta_n^o < \beta_n$ and $\alpha_n^o > \alpha_n$. In other words, a security with a sparse return series will have its beta understated and its alpha overstated. This is problematic since investors within alternatives are typically looking for investments with low market risk and skilled managers, which is precisely what low beta and high alpha would indicate.

Finally we can solve for the consistent estimators. Let b denote estimated values for β . Then it follows from (13)-(14) that a pair of consistent estimators for the true returns are given by

$$\hat{\alpha}_n = \frac{1}{T-2} \sum_{t=2}^{T-1} R_{nt}^o - \hat{\beta}_n \frac{1}{T-2} \sum_{t=2}^{T-1} M_t^o \quad (15)$$

$$\hat{\beta}_n = \frac{b_n^- + b_n + b_n^+}{1 + 2\rho_{\hat{M}}}. \quad (16)$$

The only restriction for this model to be valid is that the non-trading periods S_t are IID over time.

2.2.2 Method of Dimson (1979)

Dimson makes the following additional assumptions to the market model (1). R_t and M_t are serially uncorrelated and the error term ϵ_t has mean zero and is uncorrelated with M_t . Estimators $\hat{\alpha}$ and $\hat{\beta}$ for the constituents of (1) are found by running a linear regression on observed returns

$$R_t^o = \hat{\alpha} + \hat{\beta}M_t^o + \nu_t. \quad (17)$$

For the upcoming derivations it will prove helpful to introduce the following parameters. Let θ_i represent the probability that a security was most recently traded in time period $t-i$. Similarly, let ϕ_i denote the proportion of the market index which was last traded at $t-i$. We assume that a trade takes place at least every n periods. From this it follows that

$$\sum_{i=0}^n \theta_i = \sum_{i=0}^n \phi_i = 1. \quad (18)$$

Due to nonsynchronous trading an observed return R_t^o could correspond either to the true return in that very period, R_t , or it could correspond to the return at the last time of transaction, R_{t-i} . Therefore it is sensible to express the observed return as a combination of true returns

$$R_t^o = \sum_{i=0}^n \theta_i R_{t-i} + u_{Rt}. \quad (19)$$

A similar argument holds for the market index

$$M_t^o = \sum_{i=0}^n \phi_i M_{t-i} + u_{Mt}. \quad (20)$$

Here we introduced the errors u_{Rt} and u_{Mt} which have zero mean and no correlation with the other constituents of their respective equations.

Having set the mathematical framework, it is possible to derive the main result referred to as the *Aggregated Coefficients Method*. The starting point is a linear regression where the observed returns are regressed on previous, current and subsequent returns of the market

$$R_t^o = \hat{\alpha} + \sum_{k=-n}^n \hat{\beta}_k M_{t+k}^o + w_t. \quad (21)$$

Now the observed returns in (21) can be replaced by true returns by using the above relationships. Beginning with the left-hand side of (21), we note that using (19), (1) and (18) one can write

$$\begin{aligned} R_t^o &= \sum_{i=0}^n \theta_i R_{t-i} + u_{Rt} \\ &= \sum_{i=0}^n \theta_i (\alpha + \beta M_{t-i} + \epsilon_{t-i}) + u_{Rt} \\ &= \alpha + \sum_{i=0}^n \theta_i \beta M_{t-i} + \sum_{i=0}^n \theta_i \epsilon_{t-i} + u_{Rt} \end{aligned} \quad (22)$$

As for the right-hand side of (21), it is simplified using (20)

$$\begin{aligned} R_t^o &= \hat{\alpha} + \sum_{k=-n}^n \hat{\beta}_k \left(\sum_{i=0}^n \phi_i M_{t+k-i} + u_{Mt+k} \right) + w_t \\ &= \hat{\alpha} + \sum_{k=-2n}^n M_{t+k} \sum_{i=0}^n \phi_i \hat{\beta}_{k+i} + \sum_{k=-n}^n \hat{\beta}_k u_{Mt+k} + w_t. \end{aligned} \quad (23)$$

Equations (22) and (23) express the same thing. Specifically, we can set the coefficients of the market return M_{t+k} equal and get

$$\sum_{i=0}^n \phi_i \hat{\beta}_{k+i} = \begin{cases} \theta_{-k} \beta, & \text{if } -n \leq k \leq 0. \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

In the final step we replace the true β in (24) by the sampled $\hat{\beta}$. Summing over k one ends up with

$$\begin{aligned} \sum_{k=0}^n \theta_k \hat{\beta} &= \sum_{k=-2n}^n \sum_{i=0}^n \phi_i \hat{\beta}_{k+i} \\ \hat{\beta} \sum_{k=0}^n \theta_k &= \sum_{k=-n}^n \hat{\beta}_k \sum_{i=0}^n \phi_i \\ \hat{\beta} &= \sum_{k=-n}^n \hat{\beta}_k \end{aligned} \quad (25)$$

This implies that a consistent estimate of β is obtained by adding the coefficients from the regression equation (21).

This method only provides clear theory on how to compute adjusted values of the alpha and the beta. In the relation (19) between observed and true returns there are no values specified for the θ_i , nor is the error term u_{Rt} assigned a variance. Hence, there is no way of finding an adjusted value of the volatility and in turn an adjusted Sharpe ratio. One could, however, assume that the observed return is randomly selected from the last n true returns. With this logic one does not need any additional information. On the other hand, there will be no change in variance if observed returns are randomly chosen from true returns. Either way it is not possible to derive adjusted volatility from the theory provided.

2.3 Smoothed returns

This is an approach where nonsynchronous trading is still treated, but rather than being considered the only reason for the biases it is seen as a special case of the more general issue of illiquidity and smoothed returns. The assumption is that even if prices are quoted synchronously there might still be serial correlation in the returns if the assets are traded infrequently. This

assumption is supported for instance by Lo and MacKinlay (1990) and Kadlec and Patterson (1999) who find it unlikely for the effects of nonsynchronous trading alone to account for more than 15% serial correlation in weekly equity returns. Although the market has evolved since then, and the data treated here is more sparse, those conclusions ought to still have some bearing on the issue.

Aside from nonsynchronous trading, Getmansky et al. (2004) suggest a few different ways connected to illiquidity in which serial correlation might be induced into return series. The major reason given is associated with linear extrapolation of market values. One common way to value an illiquid asset, they argue, is to simply make a linear extrapolation based on the last market value. This is obviously an example of return smoothing and it will lead to a volatility which is biased downwards as well as to inflated serial correlation. Another possible source of return smoothing is the one induced deliberately by fund managers by holding out on positive returns to be able to "cancel out" negative future returns. This is a complicated issue surrounded by industry practice and legislation. What is clear, however, is that such deliberate smoothing will lower the volatility and in turn inflate risk-adjusted performance. The method presented by Okunev and White (2003) is based on arguments closely akin to deliberate smoothing.

It is hard to know how common such behaviour is, but regardless it is relevant to bring up methods providing a broader perspective as compared to the ones offered in the literature on nonsynchronous trading. The models and results are however, as we will see, quite similar.

2.3.1 Method of Getmansky et al. (2004)

The authors assume that in the market model (1) we have that $E[M_t] = E[\epsilon_t] = 0$ and that both M_t and ϵ_t are IID. In addition, the variance of the true return is $\text{Var}(R_t) \equiv \sigma^2$ and it has expected value $E[R_t] \equiv \alpha$. The model (1) describes the true returns which are unobservable. In practice we measure the observed returns R_t^o which are assumed to be given by a weighted average of true returns as

$$R_t^o = \theta_0 R_t + \theta_1 R_{t-1} + \dots + \theta_k R_{t-k} \quad (26)$$

where

$$\theta_j \in [0, 1] \quad \text{for } j = 0, 1, \dots, k \quad (27)$$

and

$$\sum_{j=0}^k \theta_j = 1. \quad (28)$$

In other words, the observed returns are assumed to be a weighted average of the last $k + 1$ time periods. There is no theoretical practice regarding the choice of k here, it is found empirically. A rule of thumb is that the more illiquid the asset, the higher the k . The intuition behind (28) is that eventually there will always be a trade and at that time all the relevant information will be included in the price.

This smoothing process alters the performance statistics. It is easy to see that the expected value of observed returns coincides with that of true returns

$$\begin{aligned} E[R_t^o] &= E[\theta_0 R_t + \dots + \theta_k R_{t-k}] = \theta_0 E[R_t] + \dots + \theta_k E[R_{t-k}] = \\ &\theta_0 \alpha + \dots + \theta_k \alpha = \alpha \sum_{j=0}^k \theta_j = \alpha = E[R_t]. \end{aligned} \quad (29)$$

For the variance we have the following

$$\begin{aligned} \text{Var}(R_t^o) &= \text{Var}(\theta_0 R_t + \dots + \theta_k R_{t-k}) = \theta_0^2 \text{Var}(R_t) + \dots + \theta_k^2 \text{Var}(R_{t-k}) = \\ &(\theta_0^2 + \dots + \theta_k^2) \sigma^2 \equiv c_\sigma^2 \sigma^2 \leq \sigma^2 = \text{Var}(R_t) \end{aligned} \quad (30)$$

where we introduced $c_\sigma^2 \equiv \theta_0^2 + \dots + \theta_k^2$. This shows that the variance of observed returns will always be less than or equal to the variance of true returns. In a similar manner the observed Sharpe ratio (SR) will be inflated since

$$SR^o = \frac{E[R_t^o]}{\sqrt{\text{Var}(R_t^o)}} = \frac{E[R_t]}{\sqrt{c_\sigma^2 \text{Var}(R_t)}} = \frac{1}{c_\sigma} \frac{E[R_t]}{\sqrt{\text{Var}(R_t)}} = \frac{1}{c_\sigma} SR. \quad (31)$$

Further, the observed beta value for each period is also erroneous. In fact, it will be biased towards zero. Due to the weighted average structure in (26) each beta value, i.e the coefficient of the market index, will be multiplied by the corresponding weight θ_j yielding

$$\beta_j^o = \begin{cases} \theta_j \beta, & \text{if } 0 \leq j \leq k. \\ 0, & \text{otherwise.} \end{cases} \quad (32)$$

Here we let β denote the true correlation with the market index as in (1) while the corresponding observed quantity in time period j is denoted β_j^o .

Another use of the factor c_σ^2 is as a measure of the level of smoothing. Often denoted

$$\xi = \sum_{j=0}^k \theta_j^2 \quad (33)$$

and referred to as the *Herfindahl index*, it can be used to capture the level of smoothing. The closer to zero, the more severe the smoothing. If $\xi = 1$ there is no smoothing present.

Having established how the different performance statistics are being skewed by smoothing, what remains is to calculate the magnitude of the deviations and make corrections. This is done by estimating the set of weights θ_j , which can be done using linear regression. We can substitute the market model (1) into the observed returns given in (26) in order to get

$$R_t^o = \alpha + \beta(\theta_0 M_t + \dots + \theta_k M_{t-k}) + w_t \quad (34)$$

where the error term $w_t = \theta_0 \epsilon_t + \dots + \theta_k \epsilon_{t-k}$. If we run a linear regression of observed returns on present and lagged market index returns we obtain

$$R_t^o = \alpha + \gamma_0 M_t + \dots + \gamma_k M_{t-k} + w_t. \quad (35)$$

Equating the expressions (34) and (35) and applying the normalization in (28) it turns out that the sought estimators are given as

$$\hat{\beta} = \hat{\gamma}_0 + \hat{\gamma}_1 + \dots + \hat{\gamma}_k \quad (36)$$

$$\hat{\theta}_j = \frac{\hat{\gamma}_j}{\hat{\beta}}. \quad (37)$$

2.3.2 Method of Okunev and White (2003)

This method is based on the theory introduced by Geltner (1993) to find the underlying market values from appraised values on the real estate market. Since then, variations of the method have been applied to other markets. Besides Okunev and White, there has been implementations on hedge funds by e.g. Brooks and Kat (2001). The specific method applied by Okunev and White offers means to remove any order of serial correlation in the return series. However, considering the scope of this paper, it will suffice to present the method for removing first order serial correlation.

The assumption is that observed returns can be represented as a weighted average of the true return and lagged instances of observed returns. Note the distinction between this view and the one presented in Getmansky et al. (2004) which assume the weighted average to contain only lagged and current values of the *true* returns. Thus, the process can be written as

$$R_t^o = (1 - c_1)R_t + c_1R_{t-1}^o. \quad (38)$$

Solving for the true return we find that

$$R_t = \frac{R_t^o - c_1R_{t-1}^o}{1 - c_1}. \quad (39)$$

The challenge is hence to find an appropriate value for the parameter c_1 . This can be done either analytically or by approximation. Starting with the analytical derivation, we denote the k th order of corrected or true autocorrelation a_k and the corresponding observed quantity a_k^o . We seek an expression relating the corrected and observed autocorrelations. First we simply state the definition

$$a_1 \equiv \text{Corr}(R_t, R_{t-1}) = \frac{\text{Cov}(R_t, R_{t-1})}{\sqrt{\text{Var}(R_t)\text{Var}(R_{t-1})}}. \quad (40)$$

Making the assumption that observed returns have variance equal to one it is possible to express this quantity in terms of observed autocovariances. Beginning with the numerator in (40) we use (39) and the expansion rule for covariances to get

$$\begin{aligned} \text{Cov}(R_t, R_{t-1}) &= \text{Cov}\left(\frac{R_t^o - c_1R_{t-1}^o}{1 - c_1}, \frac{R_{t-1}^o - c_1R_{t-2}^o}{1 - c_1}\right) = \\ &= \frac{1}{(1 - c_1)^2} \text{Cov}\left(R_t^o - c_1R_{t-1}^o, R_{t-1}^o - c_1R_{t-2}^o\right) = \\ &= \frac{1}{(1 - c_1)^2} \left[\text{Cov}(R_t^o, R_{t-1}^o) - c_1\text{Cov}(R_t^o, R_{t-2}^o) - c_1\text{Cov}(R_{t-1}^o, R_{t-1}^o) + c_1^2\text{Cov}(R_{t-1}^o, R_{t-2}^o) \right] = \\ &= \frac{1}{(1 - c_1)^2} \left[a_1^o - c_1a_2^o - c_1 + c_1^2a_1^o \right] = \frac{1}{(1 - c_1)^2} \left[(1 + c_1^2)a_1^o - (1 + a_2^o)c_1 \right]. \end{aligned} \quad (41)$$

Next we make the calculation for the denominator in (40). Again, using (39) and the expansion

rule for variances one gets

$$\begin{aligned} \text{Var}(R_t) &= \text{Var}\left(\frac{R_t^o - c_1 R_{t-1}^o}{1 - c_1}\right) = \frac{1}{(1 - c_1)^2} \text{Var}\left(R_t^o - c_1 R_{t-1}^o\right) = \\ &= \frac{1}{(1 - c_1)^2} \left[\text{Var}(R_t^o) + c_1^2 \text{Var}(R_{t-1}^o) - 2c_1 \text{Cov}(R_t^o, R_{t-1}^o) \right] = \\ &= \frac{1}{(1 - c_1)^2} \left[1 + c_1^2 - 2c_1 a_1^o \right]. \end{aligned} \quad (42)$$

Naturally, the same result holds for $\text{Var}(R_{t-1})$. Inserting the obtained expressions into (40) we note that the pre-factors cancel out and

$$a_1 = \frac{(1 + c_1^2)a_1^o - (1 + a_2^o)c_1}{1 + c_1^2 - 2c_1 a_1^o}. \quad (43)$$

From the relationship (43) it is possible to solve for c_1 in terms of the observed autocorrelations. Setting the corrected first order autocorrelation a_1 equal to zero we get the following second order equation in c_1

$$c_1^2 - \frac{1 + a_2^o}{a_1^o} c_1 + 1 = 0. \quad (44)$$

This equation has the solution

$$c_1 = \frac{1 + a_2^o \pm \sqrt{(1 + a_2^o)^2 - 4(a_1^o)^2}}{2a_1^o}. \quad (45)$$

This solution holds as long as the root is non-negative. If this is not the case the analytical solution is unhelpful. Instead, one may approximate the entire process $\{R_t^o\}$ as an AR(1). Accepting this assumption, first order autocorrelation is removed by simply setting c_1 equal to the observed autocorrelation

$$c_1 = a_1^o. \quad (46)$$

Having found the value of c_1 the entire return series can be reconstructed using the relationship (39).

3 Data

All data used in this paper is on index form, as opposed to data on individual funds. Both formats are represented in the referenced papers and the choice of indexed data was simply

due to availability. The return series used were collected from reports by Cambridge Associates (CA), which is a recognized actor within the field of private funds performance data. Updated data is compiled on a quarterly basis and it is based on financial statements provided by the fund managers. All return series are net of fees, expenses and carried interest. For statistical reasons, only funds with complete return history are included in the datasets. This effectively eliminates all funds lacking any return since the fund inception up until the current reporting date. This leads to the usual complication of survivorship bias, which is an issue connected to the assumption that funds which stop reporting their returns are performing worse than average. If this was the case, the aggregated return would be inflated. However, the issue of survivorship bias in the CA datasets seems to be almost entirely absent. During the period 2009-2017 the number of funds which ceased to report returns was on average 0,7% per year. In addition, these funds covered all quartiles in terms of performance.

The asset classes investigated in this paper are private equity, real estate and infrastructure. These are all examples of illiquid assets which are becoming increasingly popular. Where possible, time series spanning the last 20 years were used. In order to carry out the mathematical analysis one also needs corresponding market indexes and risk-free assets. These were obtained from Thomson Reuters. Each component of data is explained in more detail below.

3.1 Private equity

The data on private equity was taken from the CA report *US Private Equity Index and Selected Benchmark Statistics* as of the fourth quarter of 2017. The sample contains 1 455 US private equity funds including buyout, growth equity, private equity energy and subordinated capital funds, as well as liquidated partnerships. Since only American funds are included it is clear that an American market index should be used. The return history reaches back to the second quarter of 1986. Data covering the period Q4 1998 - Q4 2017 was used.

3.2 Real estate

The data on real estate was taken from the CA report *Real Estate Index and Selected Benchmark Statistics* as of the fourth quarter of 2017. The dataset consists of 1 001 funds of diverse

nationalities. Opportunistic and value-added funds are included, as well as liquidated partnerships. The geographic spread of the funds suggests the use of a global market index. There is return history since the first quarter of 1986. Data covering the period Q4 1998 - Q4 2017 was used.

3.3 Infrastructure

The data on infrastructure was taken from the CA report *Real Assets Impact Investing Indexes and Benchmark Statistics* as of the fourth quarter of 2017. Note that this data only includes impact investments, defined as investments which aside from financial returns also have a positive social and environmental impact. This obviously shrinks the fund universe. However, the 20 funds included in the sample are all officially aiming for risk adjusted market rate returns and hence the focus on environmental and social impact should not be too heavily emphasized. There is a geographic spread of the funds, but 70 % of the market cap is in American funds. Hence, it seems reasonable to use an American market index. There is return history since the third quarter of 2006. Data covering the period Q3 2006 - Q4 2017 was used.

3.4 Market Index

There is a variety of indexes to choose from when deciding on the market index. There are several customized indexes designed to fit specific assets, and there are the more general and broad ones. This paper sets out to prove a general point concerning illiquid assets, and hence the goal is to keep the implementation as general as possible. Therefore, the indexes used will be broad and well known ones. For the American market the index S&P 500 was used and for the global market the index MSCI World was used.

3.5 Risk-free asset

In order to compute Sharpe ratios and perform a portfolio optimization the return of a risk-free asset is required. For the American market the three-month US Treasury Bill is a widely accepted proxy for a risk-free asset. On a global scale it is harder to find a risk-free asset since

interest rates differ between countries. In this paper the three-month US Treasury Bill was used as the risk-free asset for both American and global markets.

4 Results

The first section presents the application of the presented theory to the sparse return series and its end result in terms of adjusted statistics. In the second section the minimum variance portfolio is found for a number of different scenarios and evaluated on a test set of the data.

4.1 Adjusted performance statistics

The methods outlined in Section 2 were applied to the data presented in Section 3 using the program R. In all cases where linear regression was used the significance level was set to 5%. For each of the three datasets the root expression for c_1 in (45) in the Okunev & White method was non-negative and the solution with the minus sign was used. As is apparent from the theory section, the four methods provide a different range of adjusted statistics. The method by Dimson only targets the alpha and the beta, whereas the method by Okunev & White reconstructs the entire return series and hence provides an adjusted measure for all statistics. The methods by Getmansky et al. and Scholes & Williams are somewhere in between, offering adjusted volatility and Sharpe ratios in addition to the alpha and the beta. For simplicity all methods are presented in a common table for each dataset, hence leaving some of the boxes intentionally blank.

The results for the private equity data are displayed in Table 1. When regressing the return series on leading and lagging market returns (i.e. market returns corresponding to time periods subsequent to and prior to the current one, respectively) only the contemporaneous and the first two lagged market returns proved significant at a 5% level. Since no leading terms were significant, the Dimson method reduces to the Getmansky et al. method and hence their alphas and betas coincide in this case. As for the alphas and betas in general, we note that in all cases the alphas are adjusted downwards while the betas are adjusted upwards. The average adjusted alpha is 22% lower than the unadjusted value. Further, the average adjusted beta

is 40% higher than the corresponding unadjusted value. The Okunev & White, Scholes & Williams and Getmansky et al. methods show an average increase in volatility of 39 %, and the Okunev & White method reduces first order autocorrelation to nearly zero. The average adjusted Sharpe ratio is 28% lower than the original value.

| Method | Alpha | Beta | Volatility | Sharpe ratio | 1st order autocorr. |
|--------------------|--------------|--------------|-------------------|---------------------|----------------------------|
| <i>Unadjusted</i> | <i>2.452</i> | <i>0.501</i> | <i>5.339</i> | <i>0.245</i> | <i>0.406</i> |
| Scholes & Williams | 2.229 | 0.652 | 7.209 | 0.182 | |
| Dimson | 2.159 | 0.720 | | | |
| Getmansky et al. | 2.159 | 0.720 | 7.511 | 0.174 | |
| Okunev & White | 2.100 | 0.718 | 7.499 | 0.175 | 0.008 |

Table 1: Results for the private equity data. The unadjusted statistics are shown in italics, followed by adjusted statistics for each of the four methods.

The results for the real estate data are shown in Table 2. When running a regression of the real estate return series on leading and lagging market index returns it turned out that only the first lagged market index along with the contemporaneous market index were significant. This means that the beta and alpha of the Dimson and Getmansky et al. methods will, again, coincide. Also note that for all four methods the beta is increased and the alpha is decreased after corrections. The average adjusted alpha is 8% lower than its unadjusted counterpart and the adjusted beta is 68% higher than the original value. The Sharpe ratio is adjusted downwards, with an average 34% lower than the unadjusted value. Finally we note that the Okunev & White method is successful in removing first order autocorrelation, and that there on average is a 53% increase of volatility.

| Method | Alpha | Beta | Volatility | Sharpe ratio | 1st order autocorr. |
|--------------------|--------------|--------------|-------------------|---------------------|----------------------------|
| <i>Unadjusted</i> | <i>2.025</i> | <i>0.236</i> | <i>5.038</i> | <i>0.104</i> | <i>0.667</i> |
| Scholes & Williams | 1.868 | 0.396 | 7.727 | 0.068 | |
| Dimson | 1.878 | 0.381 | | | |
| Getmansky et al. | 1.878 | 0.381 | 7.081 | 0.074 | |
| Okunev & White | 1.857 | 0.426 | 8.318 | 0.063 | 0.001 |

Table 2: Results for the real estate data. The unadjusted statistics are shown in italics, followed by adjusted statistics for each of the four methods.

The results for the infrastructure data are presented in Table 3. After running the linear regressions it was evident that only the contemporaneous market return was significant, hence coinciding with the original market model (1). This implies that the methods based on linear regression, i.e Dimson and Getmansky et al., provide no adjustments in this case. Therefore, only the results from the Scholes & Williams and Okunev & White methods will be addressed here. Beginning with the Okunev & White method it once again proves successful in removing first order serial correlation. The average volatility provides an increase of 14% to the original value. However, as opposed to the results on the previous datasets, there is an increase of the Sharpe ratio. This is due to the ratio being negative on this dataset. The alphas are adjusted downwards in both cases, on average by 26%. The betas on this dataset provide the only seemingly unexpected result. The beta is adjusted upwards, as in all previous cases, with the Okunev & White method but it is adjusted downwards when using the Scholes & Williams method. However, this is not totally unexpected since only the contemporaneous market return was significant on the dataset. This means that both β_n^{o-} and β_n^{o+} are small and the relation (14) becomes $\beta_n^o \approx \beta_n + 2\beta_n\rho_M^o$ which implies that $\beta_n^o > \beta_n$. On average, however, the adjusted beta is 2 % higher than the unadjusted value.

| Method | Alpha | Beta | Volatility | Sharpe ratio | 1st order autocorr. |
|--------------------|---------------|--------------|--------------|---------------|---------------------|
| <i>Unadjusted</i> | <i>-0.143</i> | <i>0.122</i> | <i>2.628</i> | <i>-0.222</i> | <i>0.109</i> |
| Scholes & Williams | -0.210 | 0.112 | 2.906 | -0.200 | |
| Dimson | - | - | | | |
| Getmansky et al. | - | - | - | - | |
| Okunev & White | -0.151 | 0.138 | 3.036 | -0.192 | 0.005 |

Table 3: Results for the infrastructure data. The unadjusted statistics are shown in italics, followed by adjusted statistics for each of the four methods.

4.2 Portfolio optimization

In order to illustrate the results and provide a sense of how the adjusted performance statistics effect a portfolio, an optimization was performed. The specific method implemented was the minimization-of-variance problem. The theory is based on the contents of chapter 4 in (Hult et al. 2012). The setting is the following. The initial invested amount V_0 can either be placed in any of k risky assets or in a risk-free asset with return R_0 . The expected returns of the risky assets are held in the vector $\boldsymbol{\mu}$ and their corresponding weights in \boldsymbol{w} . The $k \times k$ matrix $\boldsymbol{\Sigma}$ is the covariance matrix of the risky assets. We let w_0 denote the weight placed in the risk-free asset. The minimization-of-variance problem is now to minimize the variance of the portfolio given a lower bound on the return. Denoting this bound by μ_0 the problem is

$$\begin{aligned}
& \text{Minimize } \frac{1}{2} \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \\
& \text{such that } w_0 R_0 + \boldsymbol{w}^T \boldsymbol{\mu} \geq \mu_0 V_0 \\
& \text{and } w_0 + \boldsymbol{w}^T \mathbf{1} \leq V_0.
\end{aligned} \tag{47}$$

It can be shown that the solution is given by the expression

$$\boldsymbol{w} = V_0 (\mu_0 - R_0) \frac{\boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_0 \mathbf{1})}{(\boldsymbol{\mu} - R_0 \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_0 \mathbf{1})}. \tag{48}$$

This optimization was implemented for the original return series as well as for the adjusted series given by the Scholes & Williams and Okunev & White methods. The data used spans

the period of Q3 2006 - Q4 2017, determined by the infrastructure data which has the shortest return history. The initial investment was set to be 1 monetary unit and the minimum variance portfolio was computed for the three different cases of the lower bound on the return being 1, 2 and 3 percent per quarter. The risky assets were numbered in the order infrastructure, private equity and real estate. This means that the first element of \mathbf{w} , denoted w_1 , corresponds to the infrastructure asset. Similarly w_2 corresponds to private equity and w_3 to real estate. A split was performed so that half of the available data was used for finding the optimal portfolio weights while the remaining half was used to evaluate the development of the optimal portfolios. Note that this implies only around 20 data points for estimation and another 20 for evaluation, leaving the results less reliable than if a larger dataset had been available.

When using $\mu_0 = 1\%$ the minimum variance portfolio suggests short positions in infrastructure and real estate, on behalf of a large position in the risk-free asset. As is clear from Table 4, the adjustments only provide slight changes from the original allocation, ending up with marginally more invested in the risk-free asset. It is evident that using this low threshold for the return it was favourable to invest more in the risk-free asset and private equity, as in the Scholes & Williams method. Figure 2 shows how the Scholes & Williams approach is superior in this case.

| Method | w_1 | w_2 | w_3 | w_0 |
|--------------------|--------|-------|--------|-------|
| <i>Unadjusted</i> | -0.048 | 0.136 | -0.099 | 1.012 |
| Scholes & Williams | -0.089 | 0.145 | -0.080 | 1.024 |
| Okunev & White | -0.053 | 0.129 | -0.091 | 1.015 |

Table 4: Minimum variance weights for risk-free and risky assets using $\mu_0 = 1\%$.

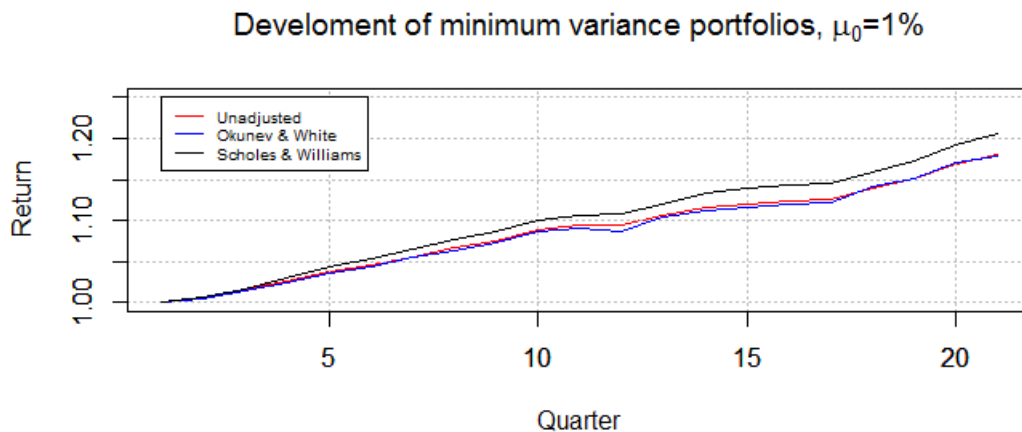


Figure 2: Performance of minimum variance portfolios using $\mu_0 = 1\%$.

When the threshold was increased to $\mu_0 = 2\%$ the same pattern continued. Both of the adjustment methods place more weight on the risk-free asset than what is done in the default portfolio (Table 5). As the demanded return increases, so does the long position in private equity and the short positions in infrastructure and real estate. The positions in the risk-free asset are also higher than in the previous case, but the change is relatively small. The Scholes & Williams approach takes the largest position in private equity, which again proves successful (Figure 3).

| Method | w_1 | w_2 | w_3 | w_0 |
|--------------------|--------|-------|--------|-------|
| <i>Unadjusted</i> | -0.179 | 0.506 | -0.371 | 1.043 |
| Scholes & Williams | -0.333 | 0.541 | -0.299 | 1.091 |
| Okunev & White | -0.196 | 0.480 | -0.338 | 1.055 |

Table 5: Minimum variance weights for risk-free and risky assets using $\mu_0 = 2\%$.

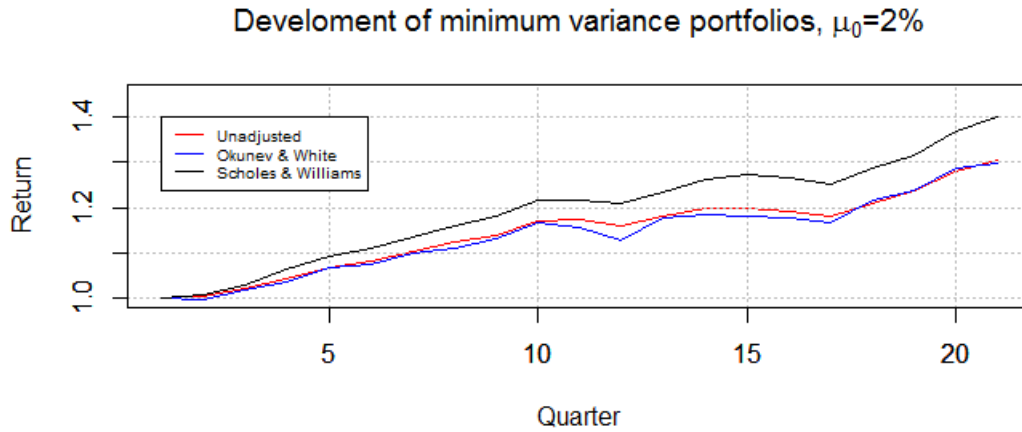


Figure 3: Performance of minimum variance portfolios using $\mu_0 = 2\%$.

For the final case of $\mu_0 = 3\%$ the allocation to private equity keeps increasing at the expense of the infrastructure and real estate positions. The risk-free weight also increases across all categories but the change is small in comparison. The largest position in private equity is again held by the Scholes & Williams method (Table 6). Figure 4 shows a development along the lines of the two previous cases.

| Method | w_1 | w_2 | w_3 | w_0 |
|--------------------|--------|-------|--------|-------|
| <i>Unadjusted</i> | -0.310 | 0.877 | -0.642 | 1.074 |
| Scholes & Williams | -0.577 | 0.937 | -0.519 | 1.158 |
| Okunev & White | -0.340 | 0.831 | -0.586 | 1.095 |

Table 6: Minimum variance weights for risk-free and risky assets using $\mu_0 = 3\%$.

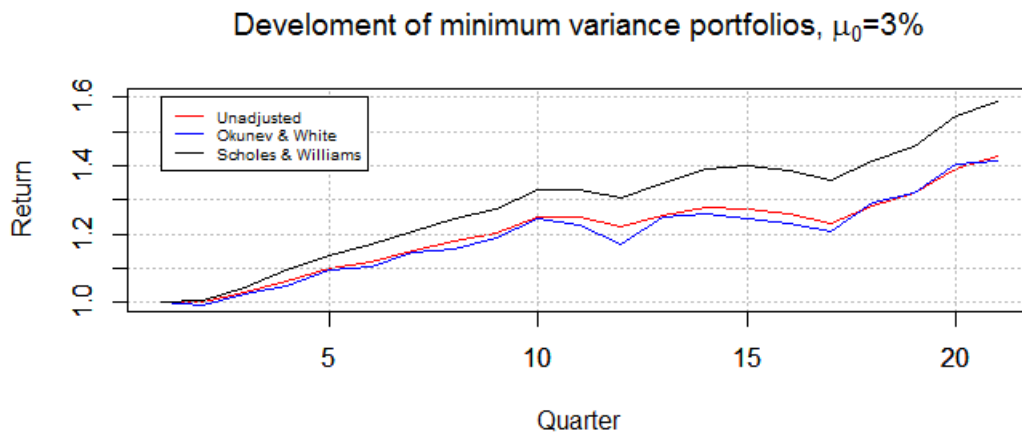


Figure 4: Performance of minimum variance portfolios using $\mu_0 = 3\%$.

5 Discussion

It is clear from the implementations that there were in fact notable adjustments made to the sparse return series investigated in this paper. After implementing the four different methods the return series statistics were in general adjusted in a way such that the volatility and the market beta were amplified. At the same time the alpha and the Sharpe ratio experienced a decrease. These results are in line with similar implementations made on hedge fund data. For instance, the work of Asness et al. (2001) established that the market exposures were understated in original return series for many hedge funds. Furthermore, the paper by Lo (2002) supports the phenomenon of overstated Sharpe ratios in original return series. In Pedersen et al. (2014) there is support presented for a substantial increase of volatility for a variety of illiquid assets, with real estate being one of them.

Regarding the results for the three different asset classes, emphasis ought to be put on the results for private equity and real estate. Both of these datasets are made up of a large pool of funds and more than 20 years of return history was available. The results for the infrastructure

data, on the other hand, should be taken with a grain of salt since the sample only contained 20 funds with a little over 10 years of return history. The fact that the two methods based on linear regression could not be successfully applied to the infrastructure dataset could be due to inadequate data.

The four mathematical methods used differ in their approach and complexity. Maybe the most simple model is provided by Dimson, closely followed by Getmansky et al. Both methods are based on linear regression and their simplicity is an advantage when it comes to interpretation of the results. However, linear regression might not capture all of the relevant relations nor provide all requested statistics. Regardless, it is redundant to use both of these methods since they will nearly always coincide. This is due to the fact that leading market returns are rarely significant, at least not at a 5% significance level. Since the Getmansky et al. method provides means of finding adjusted volatility and Sharpe ratio in addition to the alpha and beta, it should be favored over the Dimson approach. The methods by Scholes & Williams and Okunev & White are more elaborate but also more difficult to interpret. Although they are both based on correlations, the methods do not really overlap and could successfully be used alongside each other.

The portfolio optimization shows that in this case the most profitable minimum variance portfolio was achieved with the Scholes & Williams adjustment method. Looking at Tables 4-6 it is clear that this method puts the largest weights on private equity and the risk-free asset. This turned out to be successful since the average return of private equity was considerably higher than for the other two asset classes. The relatively higher weight put on private equity by Scholes & Williams could be simply because the method did not increase the variance of the private equity data as much as for the other two asset classes. It is possible that another optimization, e.g. maximization of expectation, would have rendered another method the best one. Nonetheless, the outcome of the optimization underscores the point that the application of an adjustment method such as the one provided by Scholes & Williams might lead to a higher return over time. Just as pointed out for the infrastructure dataset, the data used for the optimization also only covers the last 10 years meaning that any interpretations should be made with caution.

This paper merely provides an overview and no more than limited conclusions can be made based on the results presented here. Anyhow, the results do support the claim that there are relevant adjustments to be made to the statistics of many illiquid asset classes. A natural next step to this work would be to pick one or two of the methods and apply to more complete datasets. This way one would obtain more reliable results and be able to determine the implications of adjustments of this sort in more detail.

6 Conclusion

Overall one can conclude that there were considerable adjustments made to the return series statistics of the illiquid assets investigated in this paper. The adjustments point out that the volatility and market exposure are actually higher than what is suggested by the original return series. For manager skill and risk-adjusted return the opposite is true. These results are in line with several other works on illiquid assets, which underscores the need for more thorough research on the subject.

7 Acknowledgements

I would like to thank my supervisor Anja Janssen at KTH for her suggestions and feedback provided during the process. This thesis was written on behalf of COIN Investment Consulting Group and I owe a thanks to Joakim, Carl and Richard at the firm for their support and ideas.

References

- [1] Asness, Clifford; Krail, Robert and Liew, John, 2001. "Do hedge funds hedge?" *The Journal of Portfolio Management*, vol.28, pp.6-19.
- [2] Brooks, Chris and Kat, Harry. 2001. "The statistical properties of hedge fund index returns and their implications for investors" *Journal of Alternative Investments*, vol.5, pp.26-44.
- [3] Dimson, Elroy. 1979. "Risk Measurement When Shares Are Subject to Infrequent Trading." *Journal of Financial Economics*, vol.7, pp.197-226.
- [4] Fisher, Lawrence, 1966. "Some new stock market indexes" *The Journal of Business*, vol.39, pp.191-225.
- [5] Geltner, David. 1993. "Estimating Market Values from Appraised Values without Assuming an Efficient Market" *Journal of Real Estate Research*, vol.8, pp.325-345.
- [6] Getmansky, Mila; Lo, Andrew W. and Makarov, Igor. 2004. "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns." *Journal of Financial Economics*, vol.74, pp.529-609.
- [7] Hult, Henrik; Lindskog, Filip; Hammarlid, Ola and Rehn, Carl Johan. 2012. "Risk and Portfolio Analysis - Principles and Methods" *Springer*.
- [8] Kadlec, Gregory and Patterson, Douglas. 1999. "A transactions data analysis of nonsynchronous trading" *Review of Financial Studies*, vol.12, pp.609-630.
- [9] Lo, Andrew. 2002. "The statistics of Sharpe ratios" *Financial Analysts Journal*, vol.58, pp.36-50.
- [10] Lo, Andrew and MacKinlay, Craig. 1990. "An econometric analysis of nonsynchronous trading" *Journal of Econometrics*, vol.45, pp.181-212.
- [11] McKinsey & Company. 2012. "The Mainstreaming of Alternative Investments - Fueling the Next Wave of Growth in Asset Management"

- [12] Pedersen, Niels; Page, Sebastien and He, Fei. 2014. "Asset Allocation: Risk Models for Alternative Investments" *Financial Analysts Journal*, vol.70.
- [13] Scholes, Myron and Williams, Joseph. 1977. "Estimating Betas from Nonsynchronous Data." *Journal of Financial Economics*, vol.5, pp.309-327.
- [14] Sharpe, W.F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk" *Journal of Finance*, vol.19, pp.425-442.
- [15] Stapleton, R.C. and Subrahmanyam, M.G. 1983. "The Market Model and Capital Asset Pricing Theory: A Note" *The Journal of Finance*, vol.38, pp.1637-1642.
- [16] Truong, Caitlyn; Drisko, Carl; Richards, Kimberly and Babarinde, Gbenga. 2015. "Alternative investments - It's time to pay attention"

TRITA -SCI-GRU 2018:344