Hedge Fund Performance: A Quantitative Survey

Fan Yang[1], Tomas Havranek[1 2], Zuzana Irsova[1], and Jiri Novak[*1]

[1]Charles University, Prague
[2]Centre for Economic Policy Research, London

June 29, 2022

Abstract

We provide the first quantitative survey of the empirical literature on hedge fund performance. We examine the impact of potential biases on the reported results. Empirical analysis in prior studies has been plagued by fragmentation of underlying data and by limited consensus on how hedge fund performance should be measured. Using a sample of 1,019 intercept terms from regressions of hedge fund returns on risk factors (the "alphas") collected from 74 studies published between 2001 and 2021 we show that inferences about hedge fund returns are not significantly contaminated by publication selection bias. Most of our monthly alpha estimates adjusted for the (small) bias fall within a relatively narrow range of 30 to 40 basis points. Considering several partitions of our sample, we document a modest publication bias only for estimates based on instrumental variables (IV), for which relatively large standard errors are common and that tend to be less precise. In contrast, studies that explicitly control for the potential biases in the underlying data (e.g. the backfilling bias and the survivorship bias) report lower alphas. Our results demonstrate that despite the prevalence of the publication selection bias in numerous other research settings, publication may not be selective when there is no strong *a priori* theoretical prediction about the sign of estimated coefficients, which may induce greater readiness to publish statistically insignificant results.

Keywords: Hedge funds, meta-analysis, publication bias

JEL Codes: J23, J24, J31

* Corresponding author: Jiri Novak, jiri.novak@fsv.cuni.cz. This paper is part of a project GEOCEP that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 870245. This work was supported by the Cooperatio Program at Charles University, research area Economics. Novak acknowledges support from the Czech Science Foundation (GA CR, project 21-09231S). Data and code are available in an online appendix at meta-analysis.cz/hedge.

1 Introduction

> *"I can't figure out why anyone invests in active management, so asking me about hedge funds is just an extreme version of the same question. Since I think everything is appropriately priced, my advice would be to avoid high fees. So you can forget about hedge funds."*

Eugene F. Fama [1]

Over the past three decades hedge funds have experienced a spectacular increase in popularity. The value of assets under management (AUM) increased about 100 times between 1990 and 2020 (Stulz, 2007; Barth *et al.*, 2020). This trend is difficult to reconcile with the idea of efficient financial markets. As aptly expressed in the opening quote by Eugene F. Fama, the 2013 Nobel laureate in economics, hedge funds represent an extreme form of active investment management. They mostly trade on competitive markets where it should be challenging to earn more than the "normal" rate of return, which represents a fair compensation for a given level of systematic risk that the investment involves. Furthermore, hedge funds charge investors high management and performance fees. Ben-David *et al.* (2020) estimate that for every dollar of gross excess return earned by a hedge fund, on average, 64 cents are paid in management and incentive fees and only 36 cents are collected by the investors. Given the magnitude of these fees, it is surprising that hedge funds keep attracting growing amounts of capital.

A potential explanation for this puzzling trend is that investors have distorted views of the value hedge funds actually generate. Hedge funds are relatively lightly regulated and so they remain rather opaque in terms of their investment strategies, asset holdings, and realized returns. The paucity of information constrains systematic analysis of hedge fund performance and it may bias investors' expectations about hedge funds' value generating potential. Hedge funds are not obliged to periodically publish information on their performance. Some but not all voluntarily report their performance data to commercial data providers. This implies that data available for research is fragmented and it may suffer from numerous biases. Furthermore, hedge funds tend to engage in a wide range of unconventional investment strategies and so it is not trivial to adequately adjust for the risks they bear. It is not clear to what extent these

[1] Source: https://www.azquotes.com/quotes/topics/hedge-fund.html.

constraints bias reported performance estimates. Prior empirical literature includes numerous conflicting results, which make it difficult to draw clear conclusions. Research literature thus lacks a study synthesizing this pool of diverse empirical results.

In this paper, we perform the first quantitative survey of research literature on hedge fund performance. Our aim is to review and integrate published empirical findings and to examine how they are affected by publication selection and data biases. Brodeur *et al.* (2020b) argue that research methods that offer researchers more degrees of freedom are more likely to suffer from selective publication as researchers may choose research designs and data sample to generate results that are attractive for publication. We argue that the fragmentation of hedge fund performance data and the wide range of alternative approaches to controlling for risk give researchers considerable discretion in research design. Various data sources and estimation techniques may produce different results, some of which may be more attractive for publication than others. This potentially creates opportunities for selective publication that could bias the pool of coefficients reported in research articles. Hence, we believe that research literature on hedge fund performance merits a systematic investigation of the prevalence of this potential bias and of its potential impact on the reported estimates. Nevertheless, to the best of our knowledge, no prior study estimates and corrects for publication bias in this stream of literature. We provide the first quantification of the impact of potential selective publication and data biases on the hedge fund performance estimates.

We review empirical results in 74 studies on hedge fund performance published between 2001 and 2021. Our analysis is based on a sample of 1,019 estimates of intercept terms (i.e. the "alphas") from regressions of hedge fund returns on risk factors. The risk factors on the right-hand side of the regression equation represent various risk dimensions, to which hedge fund investments may be exposed to. The slope coefficients (i.e. the "betas") capture hedge funds' exposure to the individual risk dimensions. The intercept terms, the alphas, represent the portion of realized returns that is not attributable to the fund's exposure to the systematic risk factors. In other words, the alphas represent the abnormal return earned by the hedge fund, which represents the difference between the actual realized return and the return that would be expected as fair compensation for the investment's systematic risk.

We examine the extent to which the alpha estimates are affected by selective publication and data biases. A publication selection bias represents a tendency to publish empirical results that are consistent with the underlying theoretical predictions or with prior empirical findings. Selective publication may result from both conscious and subconscious decisions made by authors, editors, and referees who discard results that look implausible in the light of their *a priori* expectations (Ioannidis *et al.*, 2017). Publication selection bias and its implications are extensively discussed in prior literature including Stanley (2001, 2005); Stanley *et al.* (2010); Havranek (2015); Brodeur *et al.* (2016); Bruns & Ioannidis (2016); Stanley & Doucouliagos (2017); Christensen & Miguel (2018); Brodeur *et al.* (2020a); Blanco-Perez & Brodeur (2020); Zigraiova *et al.* (2021). These studies document that publication selection bias is indeed widespread in a wide range of economic settings and it has a substantial impact on the mean value of reported estimates. Given the discretion in research design due to data fragmentation and the variety of risk-adjustment methods, it is worthwhile to examine whether a similar bias is present also in the empirical literature on hedge fund performance.

We use several approaches to test for the publication selection bias. First, we exploit the property that tests of statistical significance typically assume that publication bias is a linear function of the standard error. Hence, documenting a correlation between the two can be used as evidence on biased reporting of results in primary studies (Egger *et al.*, 1997). We complement this conventional approach with several other recently developed techniques that use different combinations of fixed effects and weighting, that relax the assumptions about the underlying distribution of the estimated coefficients, and exploit discontinuities in these distributions (Stanley *et al.*, 2010; Ioannidis *et al.*, 2017; Bom & Rachinger, 2019; Furukawa, 2020; Andrews & Kasy, 2019; van Aert & van Assen, 2020). Using these modern techniques allows us to evaluate the robustness of our findings to assumptions that underlie various methodological approaches.

We find that despite the multitude of data sources and methodological approaches, empirical research on hedge fund performance is not substantially contaminated by the publication selection bias. In our regressions, most of the slope coefficients that capture the impact of publication bias are statistically insignificant. Furthermore, even in the few specifications where the coefficients capturing publication bias are significant or marginally significant, they do not dramatically change the magnitude of the estimated mean alpha coefficient corrected for the publication bias. Most of our monthly alpha estimates adjusted for this (small) bias fall within a relatively narrow range of 30 to 40 basis points, which is very close to our unconditional sample mean of reported monthly alpha estimates of 36 basis points (i.e. 0.36%).

To allow for a potentially more complex relationship between the coefficients and their standard errors reported in the primary studies, we use several nonlinear techniques designed for publication bias correction and also methods based on a selection of estimates based on their precision. We also use two recently developed approaches – the selection model by Andrews & Kasy (2019) and the p-uniform* by van Aert & van Assen (2020). For the most part, the results of these tests are consistent with the main findings based on the conventional methodology. Furthermore, using the methodology recently developed by Elliott *et al.* (2022) we do not find evidence that would suggest "p-hacking", i.e. a higher propensity to publish results that just surpass cutoffs that are conventionally used to evaluate statistical significance. These findings suggest that our results are remarkably robust to the choice of methodology to detect selective publication of results.

In addition, our results also mostly hold when we consider more homogeneous sub-samples of alpha estimates that either adjust or do not adjust for survivorship and backfilling biases and sub-samples that use a specific asset pricing model to compute the alphas. A notable exception is the group of empirical estimates based on instrumental variables (IV). Such a conclusion is consistent with a recent paper by Brodeur *et al.* (2020a) who find that IV-based estimates are more likely to suffer from publication bias than estimates based on other techniques. Furthermore, we observe little publication bias in data sub-samples partitioned based on whether a given estimate explicitly adjusts for potential biases in the underlying data (e.g. the backfilling bias and the survivorship bias). However, consistent with the expected relevance of these biases discussed in hedge fund performance literature, we observe somewhat lower alphas in studies that adjust for these data biases.

We make several important contributions to the literature. First, using several recently developed methodological approaches we aggregate and synthesize fragmented empirical evidence on hedge fund performance. Prior research has long acknowledged that the absence of a comprehensive database may lead to distortion and misleading findings (Fung & Hsieh, 2004a; Agarwal *et al.*, 2009b). Fung & Hsieh (2004a) point out that differences in coverage across various hedge fund data providers may lead to rather different average returns for a given hedge fund type. To illustrate this observation, they state that two data providers specified two substantially different estimates for equity market-neutral hedge funds for the month of January 2001 (-1.57% vs. 2.13%). Such discrepancies across data sources imply that the choice of the database may have a substantial impact on the estimated hedge fund performance. In their recent working paper, Joenväärä *et al.* (2019) underscore the importance of combining data from various databases and they propose a new way of doing so. We offer an alternative approach to overcome the data fragmentation problem. Our approach is based on aggregating the alpha coefficients estimated in prior studies that are themselves based on various data sources. Relative to Joenväärä *et al.* (2019) the advantage of our approach is that it allows us to include even estimates based on private or hand-collected data and to simultaneously control for potential data and publication biases.

Second, we provide a comprehensive battery of tests to evaluate the impact of the publication selection and data biases on hedge fund performance estimates. This paper is the first study that systematically analyzes the impact of selective publication on the reported hedge fund performance results. Recent research suggests that research settings that offer researchers considerable discretion are particularly prone to suffer from selective publication (Brodeur *et al.*, 2020a). We evaluate this effect in a research field that is characterized by fragmented data and a plurality of methodological approaches to estimate abnormal returns. Furthermore, hedge fund literature frequently mentions a concern that survivorship and backfilling biases may distort estimates of hedge fund performance (Fung & Hsieh, 2004a). Prior studies typically address this issue by measuring the returns of funds of funds as their reported performance is less likely to be affected by backfilling historical information for successfully incubated funds and by omitting data for dead funds. However, relying on the data on funds of funds has several shortcomings. First, the inclusion of a hedge fund in a fund of funds is in itself an endogenous decision that has an impact on the reported returns. There is no guarantee that the funds that are actually included in the fund of funds' portfolios are representative of the entire hedge fund population and that the individual funds are treated in these portfolios with appropriate weights relative to the general population. Second, funds of funds charge investors an additional layer of management and performance fees (Stulz, 2007), which may distort the quantification of the abnormal return generated by individual hedge funds (Amin & Kat, 2003a). Due to these additional costs, funds of funds may not represent an attractive investment opportunity for many hedge fund investors. We offer a different approach to adjust for these issues that is based on the aggregation of estimates reported in prior studies.

Third, by documenting a publication selection bias for the subset of estimates based on IV our study provides out-of-sample evidence in support of the recent finding by Brodeur *et al.* (2020a) who argue that IV-based estimates are more likely to suffer from publication bias than estimates based on other techniques. When exploring the potential underlying reasons for this finding, Brodeur *et al.* (2020a) suggest that it may arise due to the considerable discretion IV estimation gives researchers in designing their empirical tests. In line with this conjecture the authors observe that when the instruments are relatively weak, the second stage results are likely to be close to the conventional thresholds for statistical significance. Our evidence is consistent with this proposed explanation. We observe that IV-based estimates in our sample seem to be more likely to suffer from selective publication.

Our analysis is relevant to investors who consider investing in hedge funds, to regulators who seek the optimal design of the regulatory framework, as well as to researchers in economics and finance. Our evidence on the absence of a significant publication bias and the fairly narrow range of 30 to 40 basis points that we document for the corrected monthly alpha estimates allow investors to calibrate their expectations of hedge fund performance. Our study also informs regulators that even though hedge funds are not obliged to systematically publish their performance and the data is fragmented in numerous private databases, prior empirical research does not suffer from selectivity in reporting hedge fund performance. Finally, our study demonstrates that despite the prevalence of the publication selection bias in numerous other research settings in economics and finance, publication tends not to be selective when there is no strong *a priori* theoretical prediction about the sign of the estimated coefficient and when journals may be more open to publishing statistically insignificant estimates. This may help researchers identify areas where the publication bias can be expected and where, in contrast, it is less likely.

The remainder of the paper is organized as follows. In Section 2, we provide an overview of the institutional background and we discuss the arguments why hedge funds may or may not be expected to generate abnormal returns. In Section 3, we review relevant research literature. In Section 4, we present our main empirical results. In Section 5, we report and discuss our robustness checks. Section 6 concludes.

2 Background

2.1 Hedge Fund Characteristics

It is commonly believed that the first hedge fund was created in 1949 by a former Fortune magazine writer Alfred Winslow Jones (Connor & Woo, 2004; Stulz, 2007). Even though the financial industry has undergone a dramatic development over the seven decades that have passed since then many of the features of this first hedge fund resemble common hedge fund characteristics today. First, A. Jones structured the fund to be exempt from the Investment Company Act of 1940, which was the main pillar of the Security and Exchange Commission (SEC) regulations of investment entities at the time (Connor & Woo, 2004). This exemption gave the fund greater flexibility in the use of investment techniques. Second, the fund made a relatively concentrated (rather than well-diversified) investment in a limited number of stocks that it considered undervalued and it hedged some of its risks by short selling other stocks. The long-short equity strategy still remains one of the most popular hedge fund strategies. It is also a strategy that gave "hedge" funds their name. Third, to build investors' confidence A. Jones co-financed a substantial portion of the fund's assets (40%) with his own money (Stulz, 2007). Fourth, A. Jones used financial leverage to increase risk and simultaneously enhance the fund's ability to earn a higher return on the base capital. Fifth, A. Jones charged the investors a performance fee of 20% of returns earned (Connor & Woo, 2004). All of these features are quite common in hedge funds even nowadays.

In the 1960s, news about the high and relatively stable returns earned by A. Jones' hedge fund inspired imitation and many new hedge funds arose. Many of these new funds modified the original investment strategy. First, due to the hedged long-short strategy hedge funds missed out on the strong bull market of the late 1960s. That prompted many hedge funds to abandon hedging against market downturns and to pursue a leveraged long-bias strategy that keeps the fund exposed to overall market movements (Connor & Woo, 2004). In the 1980s, new global macro funds started to appear, e.g. Julian Robertson's Tiger Fund, George Soros'

Quantum Fund (Connor & Woo, 2004). In contrast to the original hedge fund that aimed at limiting its exposure to overall market conditions, these funds aimed to exploit the impact of general macroeconomic conditions typically in foreign exchange markets. The global macro funds made highly leveraged bets on the appreciation or depreciation of specific currencies. When successful (e.g. the Tiger Fund's bet on U.S. dollar appreciation, the Quantum Fund's bet on U.K. pound depreciation) these strategies generated spectacular returns, which attracted further investors (Connor & Woo, 2004). However, betting against currencies also earned hedge funds a reputation as a destabilizing force that profits from financial market turmoil.

Naturally, not all hedge fund bets turned out successful. Especially, the events of the late 1990s with the dot-com equity market bubble and the Russian debt crisis uncovered many vulnerabilities in hedge fund investment strategies. Both the Tiger Fund and the Quantum Fund lost billions on bets against the new economy that they were not able to sustain. The late 1990s also witnessed perhaps the most infamous hedge fund collapse of the Long-Term Capital Management (LTCM). The fund was started in 1993 by John Meriwether (a renowned Wall Street trader) and Myron Scholes and Robert Merton (Nobel Prize-winning economists). Between 1994 and 1998 it was very successful in pursuing the fixed-income arbitrage strategy that exploits small interest rate spreads between various debt securities. Pricing discrepancies in fairly efficient bond markets tend to be relatively small. Thus, the LTCM used very high leverage to earn an acceptable return on the capital provided by investors. This leverage became unsustainable during the Russian debt crisis when debt markets exhibited temporary anomalies. Some large investors "flew to safety" and closed their positions in riskier debt securities (Connor & Woo, 2004), which prompted the LTCM's collapse. To avoid wider contagion in financial markets the Federal Reserve Bank (FED) organized a bailout. The cost of this bailout led to further discussions about the potentially destabilizing macroeconomic impact of hedge funds. It became widely acknowledged that notwithstanding their prominent role in promoting financial markets' efficiency hedge funds may also play a more detrimental role. This understanding provided a strong motivation for systematic research in hedge funds.

Despite their growth there is, in fact, no universally accepted definition of a hedge fund (Brav *et al.*, 2008). However, hedge funds share several characteristics that distinguish them from other investment facilities. First, hedge funds are structured to take advantage of exceptions from regulatory requirements and to benefit from a favorable tax treatment (Connor & Woo, 2004). The legal framework that regulates investment entities, such as the Securities Act of 1933 and the Investment Company Act of 1940, typically allows funds with a number of investors below some threshold (often 100) to be exempted from regulatory requirements that commonly apply to mutual funds (Connor & Woo, 2004; Stulz, 2007). To qualify for such exceptions hedge funds target a limited number of high-net-worth individuals and institutional investors. From the regulatory perspective, these investors may be considered sufficiently competent to make investment decisions and sufficiently wealthy to sustain potential losses. Hence, regulators may consider it unnecessary to protect these investors from potentially adverse consequences of their investment decisions (Stulz, 2007). Furthermore, hedge funds tend to be organized as limited partnerships to benefit from pass-through tax treatment where the returns are only taxed at the individual investors' level rather than at the level of the hedge fund (Connor & Woo, 2004).

Second, the exemptions from regulatory oversight allow hedge funds to implement unorthodox and often dynamic investment strategies that may exploit a wide range of diverse investment opportunities. Furthermore, hedge funds typically use limited amounts of base capital and they use substantial leverage to increase the return earned on their investment strategies. Leverage makes hedge fund investments substantially riskier than what is common for mutual funds. Nevertheless, hedge funds frequently engage in short selling and they make a complex use of financial derivatives (Aragon & Spencer Martin, 2012) to concentrate their exposure to the idiosyncratic risk components that are inherent to the information trading they perform (Brown *et al.*, 2018). Besides the investment strategies already discussed above (long-short equity, global macro, and fixed-income arbitrage) hedge funds also engage in event-driven strategies that are based on investing in anticipation of major corporate events, e.g. mergers and acquisitions (M&As), spin-offs, reorganizations, and bankruptcies (Stulz, 2007). The success of event-driven strategies crucially depends on fund managers' ability to predict the outcome and the price impact of these events and on identifying the optimal time to make the investment.

Third, hedge funds often require their investors to commit their investment for a fairly long time (Teo, 2009). The "lockup periods" may last for several years. Even after the expiration of the lockup periods investors may be obliged to notify managers several months in advance when they want to redeem their investment (Aragon, 2007). These withdrawal restrictions give managers more flexibility in investing in illiquid assets, the value of which may remain depressed for some time. Hedge funds may also exploit opportunities that arise when more conservative investment entities such as pension funds are obliged to divest distressed securities (Connor & Woo, 2004). Holding distressed securities is typically associated with higher liquidity risk. Hence, hedge funds may have substantial exposures to macroeconomic liquidity shocks (Boyson *et al.*, 2006; Sadka, 2010). The lockup period and redemption notice period thus limit the likelihood that hedge funds will be forced to quickly liquidate these assets under unfavorable conditions.

Fourth, being exempted from many regulatory requirements allows hedge funds to remain rather opaque, which helps them protect their proprietary trading strategies from imitation by competitors. Hence, investors can typically barely learn about the rough contours of investment strategies that a given fund aims to pursue. Furthermore, unlike mutual funds, most hedge funds are not obliged to periodically report audited financial statements to regulators. Nevertheless, some funds may provide information on their performance on a voluntary basis (Stulz, 2007). Hedge funds are not allowed to publicly advertise and so having their performance record included in commercial databases may help them attract investors (Fung & Hsieh, 2004a; Baquero *et al.*, 2005). This discretion was constrained by the Dodd-Frank Act of 2010 which mandates investment funds domiciled in the U.S. that manage more than $150 million in aggregate assets to register with the SEC and to provide basic periodic disclosures on asset values, returns, borrowings, strategy classifications, investor composition, and their largest counterparties (Barth *et al.*, 2020). The asset value cutoff implies that the regulation applies only to the large hedge funds that may be systemically important.

Fifth, hedge funds typically charge their investors substantial management and performance fees (Malkiel & Saha, 2005). A common arrangement consists of a flat management fee of 1% to 2% of AUM and a variable performance fee usually 20% of realized returns above the risk-free rate (Fung & Hsieh, 1999; Connor & Woo, 2004; Stulz, 2007; Kouwenberg & Ziemba, 2007; Getmansky *et al.*, 2015). The performance fee is usually paid only after reaching the so-called "high water mark", i.e. the minimum level of absolute performance over the entire investment lifetime (Asness *et al.*, 2001; Goetzmann *et al.*, 2003; Lim *et al.*, 2016; Stulz, 2007). In other words, in a given year fund managers receive the performance fee only after having recovered any losses incurred in previous years. However, effectively the performance fees constitute even a larger portion of realized returns because investors cannot offset gains and losses across funds, they tend to withdraw capital after a poor past performance, and managers sometimes terminate hedge funds after large losses, which renders the high water mark provision irrelevant. (Ben-David *et al.*, 2020) find that due to these three reasons the effective performance fees approach one-half of the aggregate gross profits. The high level of hedge fund managers' participation in realized returns strongly incentivizes them to perform and it allows successful managers to earn compensation similar to what they would earn in mutual funds 10 times their hedge fund size (Connor & Woo, 2004; Jobman, 2002). Furthermore, unlike in mutual funds, the performance fee in hedge funds makes the compensation structure highly asymmetric. Hedge fund managers are compensated for gains, but they are not equivalently penalized for commensurate losses. These option-like payoffs strongly motivate them to take risk. The high-water mark provisions are likely to only partially moderate these risk-taking incentives because managers of unsuccessful hedge funds may opt to close the fund down and open a new one (Stulz, 2007). (Getmansky *et al.*, 2015) report that only about one-half of hedge funds survive for more than five years. Hence, hedge funds are likely to take substantial risks, which should be taken into consideration when measuring their performance.

2.2 Hedge Fund Performance

A priori, it is not quite obvious whether hedge funds should be expected to outperform other types of investments. As the opening quote suggests hedge funds typically make their investments in financial markets that are rather competitive and where investors have strong incentives to quickly eliminate any mispricing. In efficient markets, any quest for mispriced assets that subsequently earn abnormal returns may be elusive. In the past, many famous hedge fund successes were followed by spectacular failures, which suggests that extraordinary performance may be temporary and driven by chance. For example, the once-lauded and abundantly financed investment strategy of the LTCM later failed and necessitated a massive bailout (Stulz, 2007). Furthermore, competition is intensive even within the hedge fund industry. Light regulation implies relatively low barriers to entry. Any profitable strategies discovered by hedge funds may invite imitation by competitors and their ability to generate abnormal returns may quickly disappear.

Furthermore, the generous and convex "option-like" compensation packages that reward success but do not commensurately penalize failure may encourage excessive risk-taking (Cao *et al.*, 2016). Hedge fund managers may thus take aggressive positions that expose investors to substantial risks. Stulz (2007) argues that hedge fund risk profiles may resemble those of firms selling earthquake insurance. They may exhibit stable profitability for a long time but incur catastrophic losses at rare events when a disaster strikes. The LTCM's arbitrage strategy was ex-post likened to "*picking up pennies in front of a steamroller*" (Stulz, 2007). Since most hedge funds are not obliged to systematically report their performance some of these failures may be kept off the radar. If successful hedge funds are more likely to be included in the private databases and become better known to investors than the failed ones (Posthuma & Van der Sluis, 2003), investors' view of overall hedge fund performance may be distorted.

In addition, the light regulatory oversight and limited reporting requirements may impair managerial accountability and complicate monitoring by investors. Information on the portfolio composition and periodic performance may not be independently audited and so its reliability may be in question. Hedge fund managers may thus be able to camouflage inferior performance for some time, which may prevent investors from taking timely corrective action. When investors are kept in the dark they may find it difficult to base their investment decisions on a pragmatic economic calculus. Rather, they may fall prey to hedge fund managers' personal charm and keep trusting them for longer than appropriate. The Bernard L. Madoff Investment Securities investors mention the founder's personality as one of the reasons why they remained confident in the fund for so long.[2]

Finally, hedge funds charge very substantial management and performance fees. Thus, it is also conceivable that hedge funds actually beat the benchmark but the return they generate is not sufficient to cover these high fees. Paying these fees may thus leave the investors worse off than they would be by simply tracking the market index at a modest cost.

On the other hand, the flexibility resulting from the regulatory status puts hedge funds in a strong position to exploit opportunities that others cannot. It allows them to adopt a wide range of rather unorthodox investment styles that cannot be pursued by more tightly regulated mutual funds and pension funds. The light regulation allows hedge funds to remain

[2] Source: https://https://en.wikipedia.org/wiki/Madoff_investment_scandal.

secretive about the nature of their strategies, their holdings, and annual performance, which may allow them to protect their proprietary trading strategies and keep exploiting them longer than conventional mutual funds could. Hedge funds can thus act as investment strategy innovators and benefit from their first-mover advantage. They can also benefit from being a counterparty to transactions when more conventional investment entities are obliged due to regulation to divest distressed assets. Hedge funds may also benefit from introducing competition into previously oligopolistic market segments such as fixed-income arbitrage that used to be the domain of investment banks (Connor & Woo, 2004; Schneeweis, 1998).

Furthermore, investors typically agree to forgo some of the diversification benefits, which allows hedge funds to keep asset holdings relatively concentrated and to specialize in a fairly narrowly defined niche. Investment concentration may allow hedge funds to realize some gains from their high degree of investment specialization. The lack of aspiration to hold well-diversified portfolios may also give hedge funds an opportunity to act more aggressively in acquiring substantial stakes in firms and to become "activist", i.e. they can actively use their ownership rights to alter how the companies are run. Hedge fund activism can make the companies more valuable by rectifying some of the agency conflicts between the owners and managers, by adopting more suitable business strategies, and by reducing inefficiencies in operations. Besides hedge funds' stock picking and market timing skills, their activism may be another source of generating value for investors.

Hedge funds may also benefit from their flexibility in designing specific contractual arrangements with their investors and their managers. The lockup periods and the withdrawal restrictions may relieve hedge fund managers from potentially myopic short-term performance pressures and allow them to pursue strategies that may temporarily underperform but that are profitable in the long run avoiding the risk of being forced to liquidate some of their assets at temporarily depressed prices. Hedge funds may also benefit from their flexibility in designing managerial compensation contracts (Agarwal *et al.*, 2009b; Cao *et al.*, 2016). Mutual fund regulation obliges incentive compensation to be symmetric, i.e. gains and losses of equal size must have an identical opposite effect on managerial compensation (Elton *et al.*, 2003; Stulz, 2007). Most mutual funds thus make limited use of performance-based compensation and they mostly remunerate managers based on the value of assets under their management (Elton *et al.*, 2003;

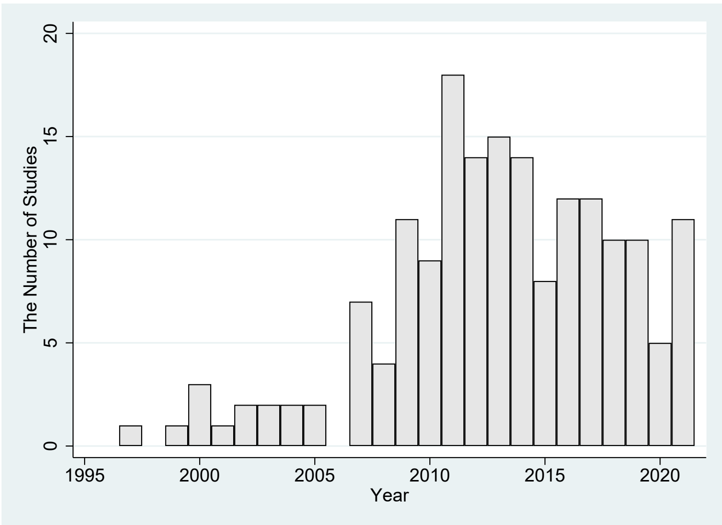
Stulz, 2007). In contrast, hedge funds are unconstrained in the design of their compensation packages. They may offer managers highly asymmetric (i.e. "option-like") compensation contracts, which helps them attract managerial talent and keep managers incentivized to perform and to take risks. Exceptionally talented and strongly motivated managers are likely to have a positive impact on hedge fund performance.

Hedge funds may also be more efficient on the cost side. They target a limited number of accredited investors, which allows them to save on marketing and communication costs. Their opacity also implies that they avoid the disclosure and attestation (audit) costs.

Thus, whether hedge funds generate abnormal returns to investors is ultimately an empirical question. To address this question, we perform a systematic quantitative survey of the empirical evidence on hedge fund performance.

3 Literature

Figure 1: Academic Articles on Hedge Funds



Notes: The figure shows the number of hedge fund-related articles in top finance journals (JF, JFE, RFS, JFQA, and RF) published in a given year excluding articles that are only published online without a print version.

The increasing prominence of hedge funds as an investment device and the increasing role they play in the economy prompted extensive empirical research aimed at evaluating how well they perform. Over the past decade numerous studies on hedge fund performance have been published. Figure 1 shows the surging number of studies on hedge funds published in leading finance journals – the *Journal of Finance*, the *Journal of Financial Economics*, the *Review of Financial Studies*, the *Journal of Financial and Quantitative Analysis*, and the *Review of Finance*.

3.1 Methodological Approaches

A standard challenge addressed in empirical research analyzing the performance of investment strategies (including those followed by hedge funds) is to properly adjust for the systematic risk these strategies involve. Two broad categories of methodological approaches are commonly used to address this issue: (i) the reward-to-risk ratios, e.g. the Sharpe (1966) ratio, and (ii) the intercept terms from regressions of realized returns on risk factors, e.g. the Jensen (1968) alpha.

The Sharpe ratio is defined as the mean divided by the standard deviation of a portfolio's excess returns, as specified in Equation 1.

$$S_p = \frac{R_p - R_f}{\sigma(R_p - R_f)}, \tag{1}$$

where S_p stands for the Sharpe ratio of portfolio p, R_p is the realized return on portfolio p, R_f denotes the risk-free rate of return, and σ denotes the standard deviation. Sharpe ratios are simple to compute as they solely depend on hedge fund return data. However, they also have several shortcomings. First, they measure risk as a simple variance, which does not accurately capture the systematic risk that should be relevant for performance evaluation. Diversified investors not only care about the volatility of returns but also about how given returns correlate with returns on other investments and with the state of the economy in general (Eling & Schuhmacher, 2007). An investment that generates high payoffs when financial market returns are low and when the economy is in decline (and the marginal utility of wealth is high) is likely more attractive than an investment that generates equivalent but differently timed payoffs.

Second, since the risk characteristics of hedge funds' unorthodox investment strategies may not be comparable to equity or debt risk profiles, the conventional variances-covariance-based risk proxies may be inadequate (Eling & Schuhmacher, 2007; Géhin, 2004). The standard deviation characterizes well the risk of returns when these are normally distributed (Eling & Schuhmacher, 2007). However, it is well-established that the distribution of hedge fund returns exhibits significant deviations from normality (Brooks & Kat, 2002; Mahdavi, 2004; Sharma, 2004; Agarwal & Naik, 2004; Malkiel & Saha, 2005). Hence, it is questionable how suitable the Sharpe ratios are for measuring hedge fund performance (Eling & Schuhmacher, 2007; Zakamouline & Koekebakker, 2009).

Third, the Sharpe ratio can be manipulated with non-linear option-based strategies that increase the ratio but do not add value to investors (Henriksson & Merton, 1981; Dybvig & Ingersoll Jr, 1982; Bernardo & Ledoit, 2000; Spurgin, 2001; Goetzmann *et al.*, 2002). For example, managers can increase the Sharpe ratio by trading in financial derivatives that sell-off very high realizations of returns (Spurgin, 2001). Furthermore, strategies that combine fairly stable modest returns with occasional large negative returns are likely to have relatively high Sharpe ratios (Goetzmann *et al.*, 2002). These considerations are likely relevant for measuring hedge fund performance because hedge funds frequently make extensive use of financial derivatives and for some of them a combination of relatively stable performance and infrequent crashes seems to be rather characteristic of the realized performance pattern (Agarwal & Naik, 2004).

The other measures commonly used in evaluating hedge fund performance are based on the intercept terms (the "alphas") from regressions of hedge fund returns on various combinations of risk factors, as shown in Equation 2.

$$(R_p - R_f) = \alpha_p + \sum_{n=1}^{N} \beta_{n,p} \cdot F_n + \epsilon_p \qquad (2)$$

where R_p denotes the realized return on portfolio p, R_f denotes the risk-free rate of return, α_p represents the intercept term, F_n represents the n-th risk factor, $\beta_{n,p}$ denotes the sensitivity of portfolio p to the n-th risk factor, and ϵ_p represents the error term. Loadings on the risk factors (the "betas") represent a "normal" compensation for the risk that the investment entails the alphas capture the portion of realized returns unexplained by the set of risk factors. The alphas can thus be interpreted as "abnormal" returns that the fund generates for the investors over and above (or below) what would be expected for a given level of risk. This approach explicitly models an investment's exposure to various risk dimensions. However, the set of relevant risk dimensions is open to question. Thus prior literature provides estimates based on various risk models.

The Jensen (1968) alpha is the simplest of the intercept-based approaches. It was initially designed to measure the investment performance of mutual funds. Returns are measured relative to a benchmark that is relevant for a well-diversified investor. Building on the portfolio theory (Markowitz, 1952) and the Capital Asset Pricing Model (CAPM) (Sharpe, 1966; Lintner, 1965; Mossin, 1966; Black, 1972) this approach uses the equity market excess return $(R_m - R_f)$ as the sole risk factor. It maintains that well-diversified investors only require compensation for an investment's contribution to the volatility of returns on the market portfolio, i.e. for an investment's systematic risk, which is in turn determined by its returns' sensitivity to the variation in market returns. The slope coefficient beta in a regression of an investment's excess return on the market portfolio excess returns captures this sensitivity. In contrast, the intercept term alpha represents the portion of realized excess return that cannot be explained by an investment's contribution to the portfolio risk, i.e. the value generated for investors.

The simplicity in modeling systematic risk with the use of a single risk dimension constitutes a limitation that may be particularly relevant for hedge funds that engage in complex and dynamic investment strategies that are likely to exhibit various forms of exposure to systematic risk. Due to this complexity, prior research develops risk models that are specifically designed to capture risk dimensions relevant to hedge fund strategies. Most prominently, Fung & Hsieh (2004a); Fung *et al.* (2008) propose a seven-factor model that comprises risk factors that mirror various risk exposures common in popular hedge fund investment strategies. Specifically, the model comprises the following risk factors: (i) the stock market excess return, (ii) the spread between the small-capitalization and large-capitalization stock returns, the excess return pairs of look-back call and put options (iii) on currency futures, (iv) on commodity futures, and (v) on bond futures, (vi) the duration-adjusted change in the yield spread of the U.S. 10-year Treasury bond over the 3-month T-bill, and (vii) the duration-adjusted change in the credit spread of Moody's BAA bond over the 10-year Treasury bond. These risk factors are intended to capture risk exposures of a broad set of hedge fund types ranging from equity long-short funds to managed futures funds.

The use of various pricing models has some advantages and disadvantages. The Jensen alpha is well-rooted in financial theory and universally applicable in a wide range of research settings. Hence, empirical results based on the Jensen alpha are easily comparable across various research strategies and data samples. On the other hand, the seven-factor model is specifically designed for research in hedge fund performance. Thus, it is likely more effective in filtering away various risk exposures relevant to complex investment strategies followed by hedge funds. Prior research offers alpha based on various risk models. In our robustness checks, we evaluate whether our findings are sensitive to limiting our analysis to a sub-sample of alpha coefficients that are based on the one-factor and seven-factor models.

3.2 Data Fragmentation

Besides the uncertainty about the appropriate risk model, hedge fund performance research faces the challenge of data fragmentation. Due to the relatively light regulatory oversight, hedge funds are mostly not obliged to periodically disclose audited financial information on their performance. Hence, there is no comprehensive central depository of hedge fund data. Only a subset of funds self-select to voluntarily report information on their performance to private data providers. Prior research thus mostly relies on data sets obtained either from commercial databases or hand-collected. The data sets used in prior research may not be comprehensive and so they may not be fully representative of the entire hedge fund population (Aggarwal & Jorion, 2010; Liang, 2000; Posthuma & Van der Sluis, 2003). This may complicate the interpretation of these findings and raise questions about the generalizability of these results to the universe of hedge funds.

Fung *et al.* (2006) discuss the level of overlap in hedge fund coverage between various databases. Liang (2000) and Agarwal *et al.* (2009b) show that the information provided is not always consistent across all the databases, which implies that the results reported in prior research may be sensitive to the choice of the source database. Similarly, in a recent working paper Joenväärä *et al.* (2019) propose a new way of combining data from various databases and they conclude that using this combined database matters for conclusion about hedge fund performance. They argue that based on this combined database hedge fund performance appears to be lower but more persistent. These findings underscore the importance of aggregating results based on different segments of hedge fund data.

Hedge funds are not obliged to independently verify reported data by auditors or established data providers. Liang (2003) finds that surviving funds are more likely to be effectively audited and funds with more reputable auditors report more consistent data. Patton *et al.* (2015) find that data on hedge fund returns change depending on when the database is accessed. They also observe that underperforming funds are more likely to alter their performance histories. Data on hedge fund returns may be unreliable because valuation of illiquid holdings may be imprecise (Cassar & Gerakos, 2011) or because the highly incentivized managers may tamper with the reported information to give an impression of better and more stable performance (Bollen & Pool, 2009). These complications may contaminate the results of hedge fund research and affect inferences about overall hedge fund performance.

3.3 Empirical Findings

Given the multitude of data sources and the range of methodological approaches used to estimate hedge fund performance, it is not surprising that prior research amassed an extensive body of sometimes conflicting empirical findings. Several studies indicate that hedge funds generate value for investors. Brown *et al.* (1999) document superior risk-adjusted returns in offshore hedge funds, but they find little support for performance persistence. Ackermann *et al.* (1999) and Liang (1999) observe that hedge funds earn higher risk-adjusted returns than mutual funds even though they have a higher overall risk due to which hedge funds do not outperform general stock market indices. Agarwal & Naik (2000) find that combining investments in hedge funds with passive investing generates better reward-risk combinations than a passive investment in various asset classes. Fung & Hsieh (2004a) propose seven risk factors relevant to hedge fund research and they find that jointly these factors explain about 80% of hedge fund returns. Nevertheless, they also find that even after considering these risk factors hedge funds generate positive alphas for the full sample period. Kosowski *et al.* (2007) use bootstrapping and Bayesian approaches to address some of the limitations common in hedge fund research. They document significant alphas and also substantial persistence in alphas in hedge funds, which suggests that the superior performance of hedge funds cannot be solely attributed to luck. Similarly, Ibbotson *et al.* (2011) conclude that alphas earned by hedge funds are positive and remarkably stable over time even during a financial crisis.

In contrast, Malkiel & Saha (2005) and Getmansky *et al.* (2015) argue that after adjusting for database biases hedge funds on average underperform their benchmarks. Fung *et al.* (2008) observe a positive and statistical alpha only for an 18-month long subperiod out of the sample covering 120 months. Billio *et al.* (2014) conclude that the alphas generated by hedge funds change dramatically over time and across categories. Capocci & Hübner (2004) observe positive excess return for 10 out of 13 investment strategies that they analyze, but only for one quarter of individual hedge funds. They also show that best-performing funds follow momentum strategies and have limited holdings of emerging market bonds. Also Ding & Shawky (2007) suggest that the evaluation of hedge fund performance relative to market indices depends on the level of aggregation of hedge fund data and on the adjustments for skewness in hedge fund returns distribution. They conclude that even though all hedge fund categories outperform the general market index less than half of the individual hedge funds beat it. Griffin & Xu (2009) find limited evidence of superior skills of hedge fund managers in timing the market and in picking individual stocks. The alphas they observe are small on a value-weighted basis and insignificant on an equal-weighted basis.

Some of the variation in the published results likely arises due to different methodological approaches used in various studies. A commonly voiced concern related to the measurement of hedge fund performance is related to the deviations from normality in the distribution of hedge fund returns (Malkiel & Saha, 2005). Several studies explicitly address this issue. Agarwal & Naik (2004) document a significant left-tail risk in a wide range of hedge fund strategies. To account for this left-tail risk they develop a conditional value-at-risk framework, which shows that the conventional mean-variance measures may underestimate expected left-tail losses by more than half. Amin & Kat (2003a) use an approach that does not require specific characteristics of the underlying returns distribution and they conclude that the vast majority of individual hedge funds and hedge fund indices are inefficient relative to the general market index. Also Bali *et al.* (2013) use an approach that accommodates the non-normality in returns distribution. Out of eleven hedge fund indices they consider, they find outperformance only for two of them – the long-short equity and emerging markets hedge fund indices. In a similar vein, Agarwal *et al.* (2009a) document that hedge funds are exposed to the risks associated with the higher moments of their returns distribution and that adjusting for this exposure substantially reduces the observed abnormal performance, especially for equity-based hedge fund strategies.

Another stream of research investigates the dependence of hedge fund performance on macroeconomic conditions. Bali *et al.* (2011) report that hedge funds with higher exposure to default risk premium and lower exposure to inflation earn higher returns. Avramov *et al.* (2013) consider four variables related to the macroeconomic conditions: the default spread, the dividend yield, the volatility index (VIX), and the aggregate fund flows into hedge funds, and they show that they predict future hedge fund returns. Similarly, Agarwal *et al.* (2017b) measure hedge funds' exposure to uncertainty about aggregate volatility and they show that funds with low exposure to this uncertainty outperform those with high exposure. Building on these findings that underscore the relevance of macroeconomic conditions for hedge fund performance Bali *et al.* (2014) include measures of macroeconomic uncertainty directly in the risk model used to measure hedge fund performance and they demonstrate the relevance of most of the macroeconomic factors in this setting.

Related to the macroeconomic conditions other papers also examine how hedge fund performance depends on conditions in financial markets. Hedge funds sometimes aspire to be "market neutral", i.e. to be able to generate fairly stable returns regardless of the general market conditions. Market neutrality should be valued by investors because robust returns during market downturns help investors diversify away risk. Nevertheless, empirical research does not provide strong support for hedge funds' market neutrality. Capocci *et al.* (2005) examine hedge fund performance in bull and bear markets and they conclude that hedge fund outperformance is concentrated in periods of rising markets. Patton (2009)considers five different ways of measuring market neutrality and he concludes that hedge fund returns tend to be positively correlated with market returns. The author also finds that about one-quarter of funds classified in the market-neutral style exhibit substantial exposure to market risk.

Hedge fund research also analyzes the impact of biases that may arise due to voluntary reporting of hedge fund performance in hedge fund databases. Fung & Hsieh (2000) and Fung & Hsieh (2002) and Fung *et al.* (2008) argue that the impact of these biases may be mitigated by using data on the funds of hedge funds (FoFs). FoFs' returns should not be affected by backfilled returns and they should appropriately reflect returns of hedge funds that decide not to report returns in commercial databases and that cease to exist (Posthuma & Van der Sluis, 2003). However, using FoF returns generates new problems. FoFs endogenously decide on what hedge funds to include in their holdings, which may not be representative of the overall hedge fund population. Furthermore, FoFs charge investors an additional layer of management and performance fees (Stulz, 2007) that reduce the realized return, which may distort the quantification of the abnormal return generated by individual hedge funds (Amin & Kat, 2003a). Brown *et al.* (2005) find that due to the extra layer of fees individual funds actually dominate FoFs in terms of net-of-fee returns, which make FoFs unattractive to investors. Getmansky *et al.* (2015) observe a decline in the number of FoFs over time, which the authors ascribe to their fee structure, competition from multi-strategy funds, and their limited ability to protect investors from losses during financial downturns. With fewer available their ability to represent the hedge fund universe in their holdings likely also declines. Thus, examining the performance of FoFs to evaluate the impact of data biases is problematic and so research literature exploits also other approaches. Below we provide an overview of this literature.

A self-selection bias arises when successful hedge funds are more likely to report their performance to commercial databases. However, it is not obvious that better-performing funds are always more inclined to report their performance to commercial databases. Some very successful hedge funds may avoid reporting to the databases to prevent disclosing clues about their proprietary trading strategies. Furthermore, well-performing hedge funds may reach their capacity limits and they may not seek any additional capital inflows. Such hedge funds may stop reporting performance to databases because they no longer have incentives to advertise themselves among investors (Ackermann *et al.*, 1999). Jorion & Schwarz (2014) indeed find that investment companies act strategically and they list in multiple commercial databases their small, best-performing funds, which helps them raise awareness about the funds and attract new investments (Fung & Hsieh, 1997, 2000). Agarwal *et al.* (2013) examine the impact of self-selection bias by comparing data in five commercial databases with information in Form 13F that are reported quarterly by advisors (rather than funds) with the Securities and Exchange Commission (SEC). They find that even though reporting initiation is more likely after a superior performance it subsequently declines. They conclude that the differences in performance between the reporting and non-reporting funds are small. Similarly, Edelman *et al.* (2013) combine previously unexplored data sources with manual data collection to construct a comprehensive dataset of returns earned by large hedge fund management companies. Based on the sample covering more than half of the industry's AUM they observe little differences between the reporting and non-reporting firms. In contrast, Aiken *et al.* (2013) use the mandatory regulatory filings by registered funds-of-funds (FoFs) that are obliged to report their holdings in individual hedge funds. They observe that only about one-half of these fund-level returns are reported to one of the five major hedge funds databases. Comparing the two subsamples they observe that non-database funds significantly underperform funds that report their performance to one of the databases. The result seems to be driven by the left tail of the returns distribution, that is by funds in decline that quit reporting to databases before their performance further deteriorates.

The backfilling bias or the "instant-history bias" arises when hedge funds are included in databases together with their performance history only after succeeding during an "incubation period" intended to accumulate a performance track record before offering the fund to investors. Recording performance histories of only the successful funds introduces a positive bias into the database (Fung & Hsieh, 2000; Posthuma & Van der Sluis, 2003). To quantify its effect prior research compares returns generated in the first years of hedge fund existence in the database with other years. Estimates based on this approach range between 1.0% and 1.5% per annum (Fung & Hsieh, 2000; Edwards & Caglayan, 2001). Posthuma & Van der Sluis (2003) access additional information on the length of the incubation period in the TASS database and they find the bias to be more prevalent and significant. They observe that a typical incubation period lasts for about 3 years. They also find that more than half of the recorded returns are backfilled, which results in a bias of about 4% per annum. To mitigate the effect prior research sometimes eliminates the first year of data that are most likely to be affected by the backfilling bias (Kosowski *et al.*, 2007; Teo, 2009; Avramov *et al.*, 2011). Nevertheless, Fung & Hsieh (2009) argue that this approach is problematic. The length of the incubation period may differ greatly and the information on funds' inception dates may be unreliable or missing in the databases. Some hedge funds may also enter the sample due to database mergers. Hence, removing the first year of observations is a rather blunt instrument that also results in a substantial loss of data and impairs the power and generalizability of empirical tests. Similarly, Jorion & Schwarz (2019) suggest that truncating early returns does not resolve the backfilling bias and it can lead to misleading conclusions. They recommend removing returns prior to the listing date and they propose an approach of inferring these dates when they are missing in the database.

The survivorship bias may arise when commercial databases terminate coverage of previously included funds. Providers may wish to purge the database of funds that no longer operate because they are not relevant to their clients anymore. Hodder *et al.* (2014) report that on average 15% of hedge funds exit the database every year. A bias arises when the funds that exit the database on average underperform the "surviving" funds. Edelman *et al.* (2013) and Getmansky *et al.* (2015) argue that two types of hedge funds are likely to stop reporting their performance to databases: those that are no longer attractive to investors and those that do not seek to attract new investors anymore. Funds that approach liquidation after having incurred substantial losses and experiencing an outflow of funds by investors lack the incentive to continue reporting their performance because they are no longer attractive to investors. On the other hand, well-performing funds that approach their capacity limit and no longer seek additional capital inflows also have incentives to quit reporting their performance to databases. Hence, the impact of survivorship bias in the context of hedge funds is not *a priori* quite obvious.

Prior research suggests that database exits due to poor performance tend to be more common. Fung & Hsieh (2000) observe that 60% of defunct funds are liquidated whereas 28% are removed from the database because the managers stopped reporting return information. To estimate the performance of successful funds that may exit the database due to capacity constraints Edelman *et al.* (2013) compare performance of large non-reporting funds identified through an industry survey with funds of comparable size that do report their performance to one of the commercial databases. They observe fairly similar performance for both groups. These findings suggest that databases likely overstate true hedge fund performance. Brown *et al.* (1999) examine survivorship bias in a database of active and defunct offshore funds and observe positive risk-adjusted returns even after adjusting for the bias. Liang (2000) observes that poor performance is the main reason for a fund's disappearance from the databases and finds that the survivorship bias exceeds 2% per annum and it varies with investment styles. Edwards & Caglayan (2001) compare the performance of defunct funds with those that are still in operation and they estimate the impact of the bias at 1.85% per annum. Similarly, Amin & Kat (2003b) estimate the impact of the survivorship bias to be around 2.0% per annum on average, but substantially higher for small, young, and leveraged funds (between 4.0% and 5.0%). Fung *et al.* (2006) estimate the impact of the survivorship bias at 1.8% and 2.4% per annum. In comparison, Agarwal *et al.* (2015) propose a range between 2.0% and 3.6% per annum. They also state that the bias varies across databases, sample periods, and fund characteristics.

The survivorship bias may be expected to decrease over time as databases improve the consistency of their coverage and retain historical data. However, even databases that retain the data for defunct funds may be contaminated by the delisting bias or liquidation bias. Aiken *et al.* (2013) find that about half of the hedge funds continue to operate two years after the delisting date and their returns are 1.8% lower than returns of funds that continue reporting their performance to the database. Edelman *et al.* (2013) argue that the reliability and consistency of performance data provided by hedge funds approaching liquidation often deteriorates, which may prompt data vendors not to record them due to questionable reliability. This implies that even databases that include records for the "dead" funds may miss some of the last performance data that tend to be rather poor. Hodder *et al.* (2014) use estimated portfolio holdings for funds-of-funds and they estimate the average delisting return for all hedge funds of -1.61%. They also find that the negative delisting return is substantially larger for funds with poor prior performance and with no clearly stated delisting reason. Other studies estimate the impact of missing delisting returns on estimates of average hedge fund performance. Edelman *et al.* (2013) estimate the magnitude of the delisting bias at a modest 0.02% per annum. Jorion & Schwarz (2013) exploit the differences in the timing of hedge fund delisting from various databases and estimate the impact of the bias to be at least 0.35% per annum. They suggest that hedge fund indices should be adjusted downward by 0.5% per annum to adjust for the effect.

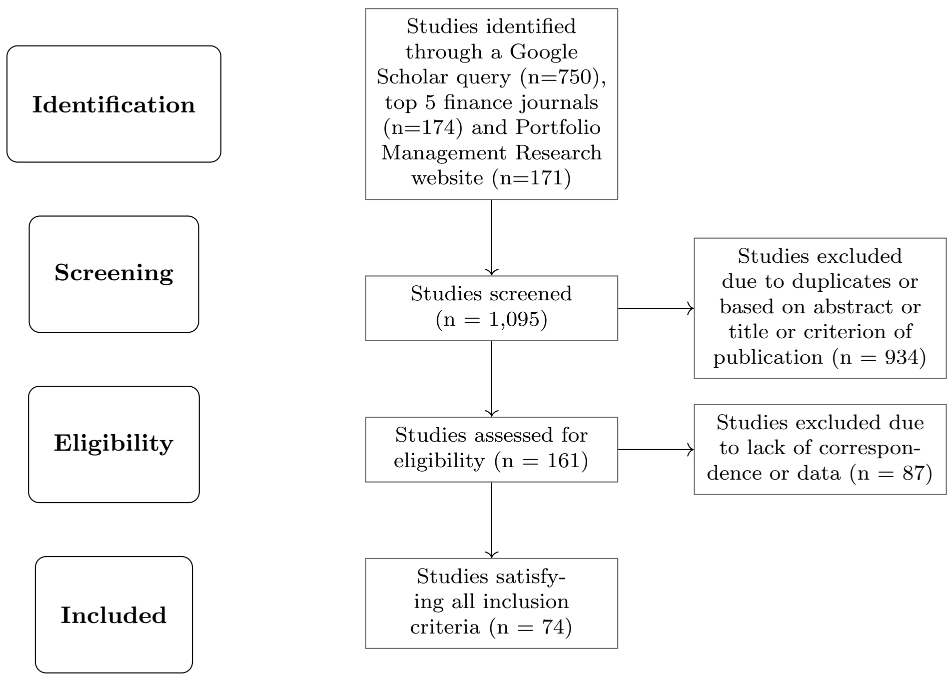
The variability of prior empirical results and the potential impact of various data biases complicate the interpretation of this stream of research. Thus, we consider it worthwhile to conduct a quantitative survey to synthesize this pool of diverse empirical results and to examine how they are affected by publication selection and data biases.

4 Dataset

In order to perform a comprehensive analysis of how published evidence on hedge fund performance is affected by publication selection and data biases we collect a large dataset of alpha estimates from primary studies. Our data collection process follows the guidelines proposed by Havranek *et al.* (2020). To cover the entire time period over which hedge fund research surged, we start our search on January 1, 2001 and we end it on September 1, 2021. The long time span exceeding 20 years ensures that our sample of alpha estimates is representative of the accumulated pool of evidence in this stream of research literature. We restrict our analysis to estimates published in peer-reviewed research journals. The peer-review process constitutes an important quality assurance mechanism. Using only estimates that underwent the peer-review process increases the likelihood that the collected alpha coefficients are estimated using established methodological approaches and they are free of error. Furthermore, we expect most researchers and practitioners to form their subjective understanding of typical alpha estimates predominantly based on published articles. Our sample thus likely mirrors the set of studies that shape people's views of hedge fund performance.

Figure 2 provides an overview of the individual steps of our data collection process. First, we built a preliminary list of studies based on references included in the sections on hedge fund performance in two comprehensive review articles: Connor & Woo (2004) and Agarwal *et al.* (2015). Second, we perform a systematic Google Scholar search using the following

Figure 2: PRISMA Flow Diagram



Notes: Our baseline search query is *("hedge fund") AND ("returns" OR "performance")* in Google Scholar and *("hedge") AND ("fund" OR "funds")* in top 5 finance journals and Portfolio Management Research website. We collect the first 750 studies returned by the search in Google Scholar and check the relevant 174 results in top 5 finance journals and 171 results on the Portfolio Management Research website. We are left with 161 studies after the screening. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based set of items for reporting in systematic reviews and meta-analyses. More details on PRISMA and reporting standard of meta-analysis in general are provided by Havranek *et al.* (2020).

combinations of keywords: "hedge fund returns" OR "hedge fund performance". We search for alpha estimates in the articles as ordered by Google Scholar. We terminate this phase of data collection after having covered the first 750 articles in the Google Scholar list. We observe that after having reached this position at Google Scholar list the articles become less relevant and the likelihood of identifying additional articles with usable alpha estimates drops dramatically. Third, to make sure that our search does not miss any important articles we perform a slightly broader Google Scholar search using less restrictive keywords: "hedge fund" OR "hedge funds" in the following finance journals: the *Journal of Finance*, the *Journal of Financial Economics*, the *Review of Financial Studies*, the *Journal of Financial and Quantitative Analysis*, and the *Review of Finance*. Fourth, to ensure a comprehensive coverage of articles in journals aimed primarily at investment professionals and which may not be as highly cited and ranked by Google Scholar we perform a similar search based on the combination of keywords: "hedge fund" OR "hedge funds" in the journals listed on the Portfolio Management Research website[3]: the *Journal of Portfolio Management*, the *Journal of Financial Data Science*, the *Journal of Impact and ESG Investing*, and the *Journal of Fixed Income*.

To be included in the dataset, a given alpha estimate must be accompanied by a measure of statistical significance, i.e. a t-statistic, a standard error (SE), and/or a p-value. We use these measures to compute the precision of individual alpha estimates. We use the precision variable in our tests of selective publication as well as for our data verification. Before constructing our final sample we attempt to identify alpha coefficients that may have resulted from human error in data hand-collection. To do so we first convert all the measures of statistical significance to a common metric, i.e. t-statistic. Whenever available we collect corresponding t-statistics from primary studies. If the authors report standard errors instead, we compute the implied t-statistic as the ratio of the alpha coefficient and the corresponding standard error. In studies using the Bayesian approach, we divide the alpha coefficient by the reported standard deviation. If the authors report p-values we check whether they explicitly state that these are based on a one-tailed or a two-tailed tests. If the type of the test is not explicitly stated in the article we try to infer it from the discussion of the level of statistical significance of results tabulated in the primary studies. If the type of the test cannot be ascertained from the interpretation of the results we assume a two-tailed test (1 study). We then use the inverse t-distribution to convert

[3]Source: https://www.pm-research.com/).

the reported p-value to a t-statistic. If the authors report the total number of observations based on which a given alpha coefficient is estimated we use that number for the degrees of freedom. If the authors report both the number of observations in the cross-section and in the time-series we use the product of the two numbers. If the information on the number of observations is only provided for the cross-section or the time-series, we use that number instead. If none of the above is provided we assume 168 observations, which is equal to the sample median for the sub-sample where the number of observations is explicitly stated. We then check all observations with the implied t-statistic greater than 10 for potential errors in hand-collecting the data. We ensure that such results are presented as highly significant in the main text of the primary study. We discard the 1 observation where the authors report a t-statistic greater than 50.

Table 1: List of 74 primary studies.

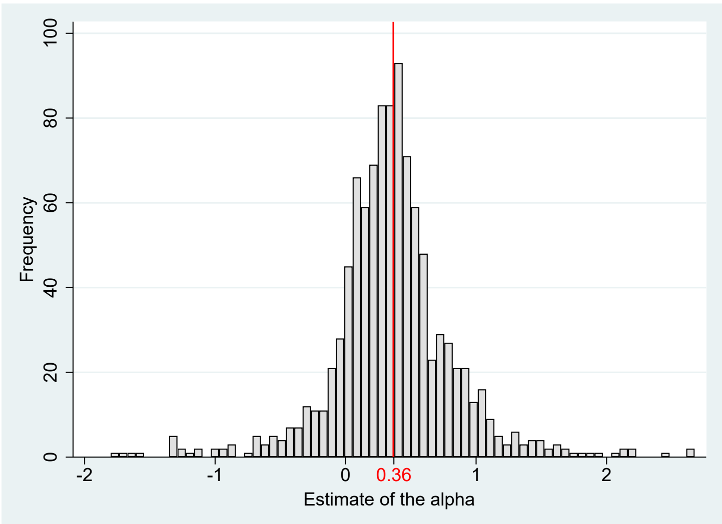
Agarwal *et al.* (2017a)	Edelman *et al.* (2013)	Malladi (2020)
Ahoniemi & Jylha (2014)	Edwards & Caglayan (2001)	Meligkotsidou & Vrontos (2008)
Aiken *et al.* (2013)	Eling & Faust (2010)	Mitchell & Pulvino (2001)
Ammann & Moerth (2005)	Frydenberg *et al.* (2017)	Mladina (2015)
Ammann & Moerth (2008a)	Fung & Hsieh (2004a)	Molyboga & L'Ahelec (2016)
Ammann & Moerth (2008b)	Fung & Hsieh (2004b)	Mozes (2013)
Aragon (2007)	Fung *et al.* (2002)	Patton & Ramadorai (2013)
Asness *et al.* (2001)	Fung *et al.* (2008)	Racicot & Theoret (2009)
Bali *et al.* (2013)	Gupta *et al.* (2003)	Racicot & Theoret (2013)
Bhardwaj *et al.* (2014)	Hong (2014)	Racicot & Theoret (2014)
Blitz (2018)	Huang *et al.* (2017)	Ranaldo & Favre (2005)
Bollen & Whaley (2009)	Ibbotson *et al.* (2011)	Diez De Los Rios & Garcia (2011)
Brown (2012)	Jame (2018)	Rzakhanov & Jetley (2019)
Buraschi *et al.* (2014)	Joenvaara & Kosowski (2021)	Sabbaghi (2012)
Cao *et al.* (2016)	Joenvaara *et al.* (2019)	Sadka (2010)
Chen & Liang (2007)	Jordan & Simlai (2011)	Sadka (2012)
Chen *et al.* (2017)	Jylha *et al.* (2014)	Sandvik *et al.* (2011)
Chincarini & Nakao (2011)	Kanuri (2020)	Stafylas *et al.* (2018)
Clark & Winkelmann (2004)	Klein *et al.* (2015)	Stafylas & Andrikopoulos (2020)
Dichev & Yu (2011)	Kooli & Stetsyuk (2021)	Stoforos *et al.* (2017)
Ding & Shawky (2007)	Kosowski *et al.* (2007)	Sullivan (2021)
Ding *et al.* (2009)	Kotkatvuori-Ornberg *et al.* (2011)	Sun *et al.* (2012)
Do *et al.* (2005)	Liang (2004)	Teo (2009)
Duarte *et al.* (2007)	Ling *et al.* (2015)	Vrontos *et al.* (2008)
Edelman *et al.* (2012)	Lo (2001)	

Notes: The the list of primary studies, from which we have collected at least one alpha estimate.

Our data collection procedure yields 1019 alpha estimates obtained from 74 primary studies. Table 1 shows the list of primary studies, from which we have collected at least one alpha estimate. The data sample size makes our study one of the largest quantitative surveys of prior studies in financial economics. Alpha estimates represent abnormal returns adjusted for exposures to risk factors. Individual alpha coefficients reported in primary studies thus aim to capture the same underlying concept of value generated by hedge funds for investors. All the collected alpha coefficients are measured in the same unit (i.e. percentage) and they are normalized to monthly frequency. Hence, they are directly comparable, which makes them suitable for aggregation in a quantitative survey.

Figure 3 shows the histogram of the alpha estimates in our sample. The figure suggests that the distribution is fairly normal and quite symmetric. Furthermore, we do not observe any significant kinks in the distribution, which indicates that no levels of alpha estimates are significantly underrepresented or over-represented. Figure 3 thus offers some preliminary indication that the distribution of our dataset has the expected characteristics and it is free from dramatic discontinuities.

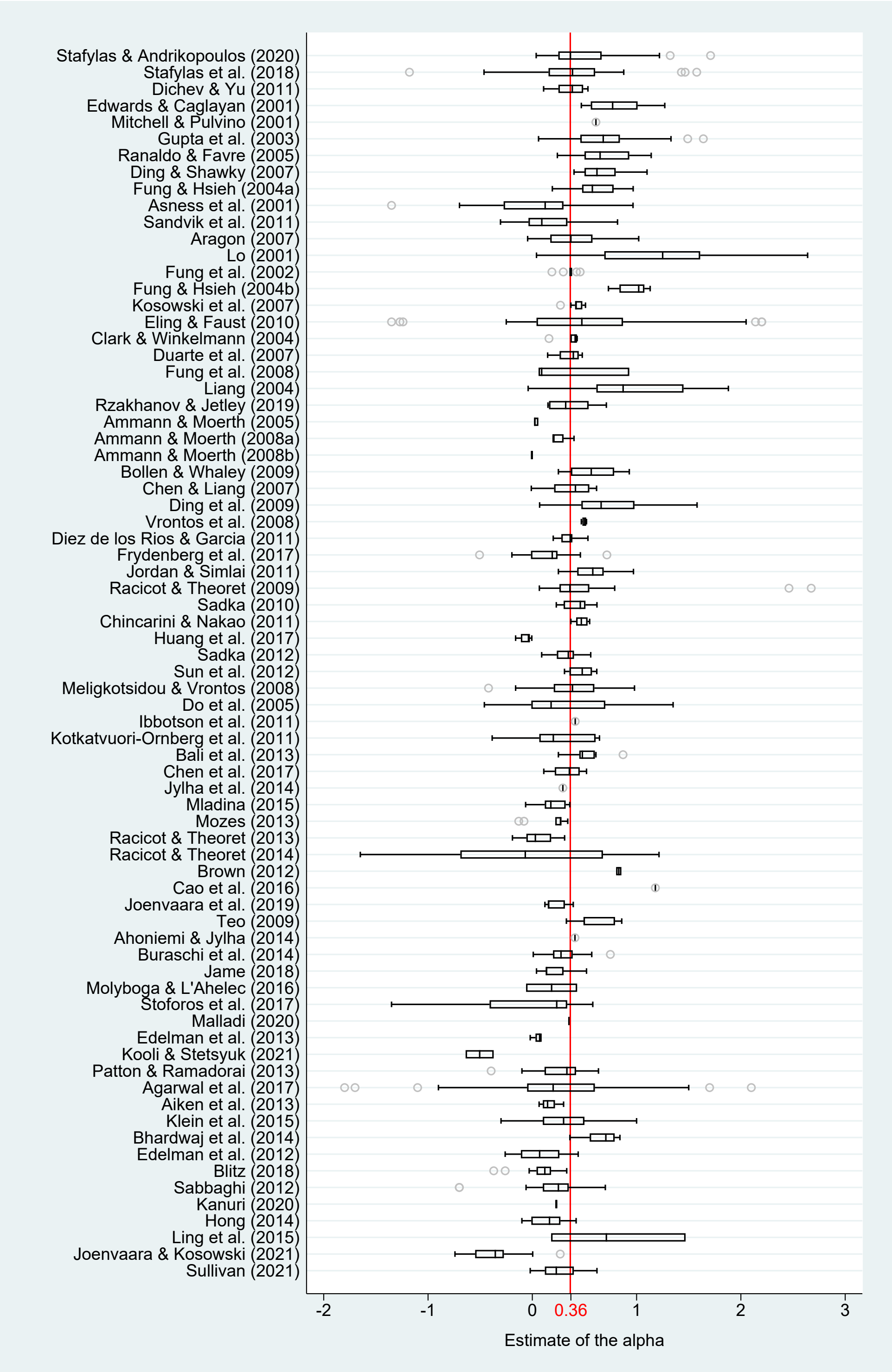
Figure 3: Distribution of the reported estimates



Notes: The figure depicts a histogram of the alphas reported by individual studies. The vertical lines denotes sample mean.

The vertical line in Figure 3 denotes the unconditional sample mean of monthly alphas of 0.36%, which corresponds to annual abnormal return of 4.32%. This result is broadly consistent with values proposed in prominent studies on hedge fund performance. For example, Getmansky *et al.* (2015) report monthly alphas based on the Fung & Hsieh (2001) seven-factor model for various hedge fund strategies between 0.18% and 0.56%. This suggests that our dataset does not dramatically differ from what would be expected based on prior literature. At the same time the histogram shows that the individual alpha estimates are relatively dispersed. This suggests that there are substantial differences across various studies and estimation approaches.

Figure 4: Estimates of the alpha vary both within and across studies



Notes: The studies are sorted by the age of the data they use from oldest to youngest. The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. The vertical line denotes sample mean. For ease of exposition, outliers are excluded from the figure but included in all statistical tests.

Figure 4 visualizes the distribution of alpha estimates reported in the individual primary studies. The boxes represent the interquartile range between percentile 25 and percentile 75 and the vertical line inside each box denotes the median value for a given study. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. Consistent with the preliminary conclusion based on the histogram shown in Figure 3, also the pattern depicted in Figure 4 shows that the individual studies differ greatly in the dispersion of reported alpha coefficients. For most studies, the interquartile ranges cross the vertical line representing the unconditional sample mean of 0.36%. However, there are some studies with interquartile ranges not overrlapping with the unconditional sample mean. In fact, some of them are fully below zero. Furthermore, we observe that some studies exhibit rather wide interquartile ranges exceeding one percentage point. This suggests that even within some studies the reported coefficients vary greatly. The substantial heterogeneity of alpha estimates reported in primary studies further underscores the importance of conducting a quantitative survey that aggregates these diverse results and corrects them for potential biases.

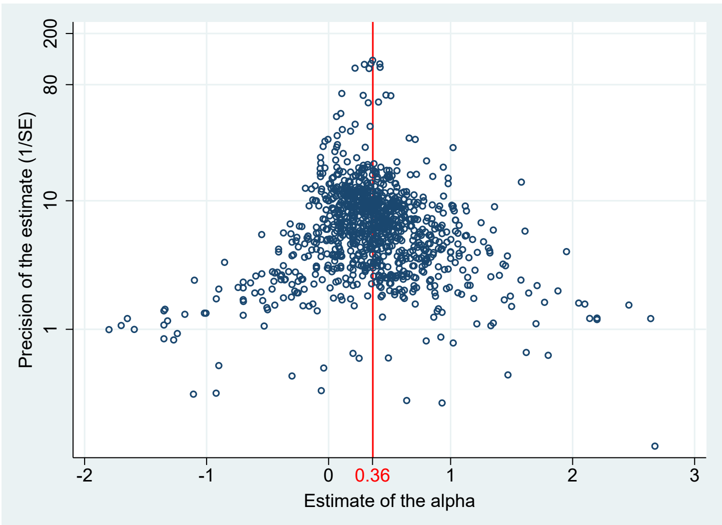
5 Main Results

5.1 Funnel Plot

Having observed heterogeneity in reported alpha coefficients across various studies we now analyze whether these estimates are affected by publication selection bias. We start this analysis by visualizing how the alpha estimates reported in primary studies depend on their precision, which is defined as one over the estimate's standard error. Tests of statistical significance based on the t-distribution assume that the estimated coefficients and their standard errors are not correlated. Hence, in the absence of publication bias there should be no systematic relationship between the alpha coefficient and its standard error. In contrast, detecting a positive or a negative association between the coefficients and standard errors suggests selective publication. The authors of primary studies usually report t-statistics for their estimates, which implies that they assume that the estimates and their standard errors are statistically independent and the ratio of the estimates to their standard errors has a t-distribution. The association between the coefficient and its standard error can thus be used to detect selective publication.

In our setting the association can be depicted with a funnel plot with the alpha coefficients on the x-axis and their precision (i.e. $1/SE$) on the y-axis. We show such a funnel plot in Figure 5. In a bias-free world, the graph should resemble a symmetrical inverted funnel. The funnel shape arises because the most precise estimates tend to be concentrated around the underlying mean value, whereas less precise estimates with larger standard errors are more dispersed around the mean. The funnel plot shall be symmetric if for any given level of estimate precision both high and low estimates are equally likely to be published. In contrast, if imprecise estimates that are high trend to be reported, while equally imprecise estimates that are low get discarded then the funnel plot shall miss some observations in the left part and consequently it shall be asymmetric. An asymmetric funnel plot indicates that estimates are reported selectively in primary studies, which implies that their mean value provides a biased estimate of the underlying mean value in the population.

Figure 5: The funnel plot of alpha estimates



Notes: When there is no publication bias, estimates should be symmetrically distributed around the mean (denoted by the vertical line). Outliers are excluded from the figure for ease of exposition but included in all statistical tests.

Figure 5 exhibits no obvious asymmetry, which is consistent with little or no publication bias. For any given level of precision both high and low estimates seem to be represented in the plot. The funnel plot thus provides initial suggestive evidence indicating that hedge fund alpha estimates reported in primary studies are not significantly contaminated by publication selection bias. Furthermore, a simple visual examination of Figure 5 suggests that the funnel plot is slightly “hollow”, which might suggest that insignificant estimates (low precision alpha close to zero) are less likely to be published. Below we formally test for the significance of these observed patterns.

5.2 Formal Tests

Having provided preliminary evidence about the likelihood of a publication selection bias in hedge fund performance literature we proceed with using several approaches to formally test for it. The first set of tests exploits the above mentioned association between the alpha coefficients reported in primary studies and their standard errors. Since we use the term “alpha” to refer to the intercept term in the regression of returns on risk factors reported in primary studies, we use “kappa” to denote the constant (i.e. the intercept term) in our regressions of alpha coefficients on their standard errors. Furthermore, we use “lambda” to refer to the slope coefficient at the explanatory variable of SE. We estimate the following equation:

$$\alpha_{ij} = \kappa + \lambda \cdot SE(\alpha_{ij}) + \epsilon_{ij}, \tag{3}$$

where α_{ij} stands for the i-th estimate of hedge fund alpha reported in the j-th study, $SE(\alpha_{ij})$ denotes its standard error, and ϵ_{ij} is the error term.

In the absence of any publication bias the slope coefficient λ is expected to be zero, which implies no association between the alpha estimates (α_{ij}) and their standard errors ($SE(\alpha_{ij})$). In contrast, if publication of estimated alpha coefficients is selective and low alpha estimates are more likely to remain unreported in primary studies then imprecise estimates (i.e. those with a large SE) should be more likely to be high rather than low leading to a positive λ coefficient. Conversely, a tendency to discard high rather than low alpha coefficients would lead to a negative λ coefficient. Hence, the slope coefficient λ reflects the effect of publication selection bias and the intercept term κ captures as the true mean alpha estimate corrected for the bias.

Panel A of Table 2 shows the results for several alternative ways of estimating Equation 3. In the first column we report the conventional ordinary least squares (OLS) estimate. As discussed above the OLS estimate represents the most straightforward way of testing for selective publication that is commonly used in prior research. However, it could yield spurious results in case unobserved features of the primary study design are correlated with the reported alphas.

Table 2: Full Sample Results

Panel A: Linear models						
	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.0152	-0.0265	0.0602	0.178	0.324	0.0497
	(0.188)	(0.215)	(0.131)	(0.353)	(0.320)	(0.127)
	[-0.534, 0.455]			[-0.526, 0.971]	[-0.415, 1.12]	[-0.348, 0.457]
				{-0.626, 0.983}		
Effect beyond bias (κ)	0.366***	0.369***	0.350***	0.316***	0.301***	0.353***
	(0.0426)	(0.0540)	(0.0474)	(0.0854)	(0.0440)	(0.0380)
	[0.277, 0.458]			[0.157, 0.475]	[0.186, 0.412]	[0.270, 0.436]
First-stage robust F-stat				12.71		
Observations	1,019	1,019	1,019	979	1,019	1,019
Panel B: Non-linear models						
	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.183*	P = 0.631	L = 0.364
				(0.106)	(0.092)	(*p = 0.834*)
Effect beyond bias	0.310***	0.325***	0.355***	0.320***	0.274***	0.386***
	(0.026)	(0.009)	(0.093)	(0.008)	(0.03)	(0.045)
Observations	1,019	1,019	1,019	1,019	1,019	1,019

Notes: The first two panels report the results of the regression $\alpha_{ij} = \kappa + \lambda \cdot SE(\alpha_{ij}) + \epsilon_{ij}$, where α_{ij} denotes the i-th alpha coefficient estimated in the j-th study, and $SE(\alpha_{ij})$ denotes its standard error. FE: study-level fixed effects, BE: study-level between effects, IV: the inverse of the square root of the number of observations is used as an instrument for the standard error, WLS: model is weighted by the inverse of the standard error of an estimate, wNOBS: model is weighted by the inverse of the number of estimates per study. In Panel B, Top10 is model by Stanley *et al.* (2010), WAAP stands for Weighted Average of the Adequately Powered model by Ioannidis *et al.* (2017), Kinked-meta is endogenous kink model by Bom & Rachinger (2019), Stem model is by Furukawa (2020), selection is model by Andrews & Kasy (2019) using clustered SEs, P denotes the probability that estimates insignificant at the 5% level are published relative to the probability that significant estimates are published (normalized at 1), p-uni* is by van Aert & van Assen (2020), L denotes test statistic of p-uniform's publication bias test. Standard errors, clustered at the study level, are reported in parentheses. 95% confidence intervals from wild bootstrap in square brackets (Roodman *et al.*, 2018). In curly brackets we show the two-step weak-instrument-robust 95% confidence interval based on Andrews (2018) and Sun (2018).* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To address this potential problem we complement the OLS estimates with several alternative estimation techniques. The results reported in the second column of Table 2 are based on an estimation that includes study-level fixed effects (FE). Including study-level fixed effects filters out idiosyncratic study-level variation. Hence, as long as alpha estimates in a given primary study are estimated using similar methodologies including study-level fixed effects removes potential confounding effect of these methodological choices on the reported alpha estimates. Identification of the fixed-effect estimator rests on studies that report more than one estimate. Thus we complement the analysis with study-level between-effect estimation (BE) that accounts for the differences in study size. We report these results in the third column of Table 2.

To further address the issue of potential endogeneity in the method choices and reported standard errors in the primary studies we follow Stanley (2005) and Havranek (2015) and we use the inverse of the square root of the number of observations in primary studies as an instrumental variable (IV) for the standard error. This measure has the desirable characteristics for a valid instrument. By construction, the number of observations is correlated with the standard error. At the same time, it is plausibly unrelated to the chosen estimation technique. Furthermore, it seems reasonable to assume that the number of observations is quasi-randomly distributed among the primary studies. The results based on this instrument thus constitute an important robustness check. We report these results in the fourth column in Table 2.

In the last two columns of Panel A of Table 2 we report our weighted least squares estimates of Equation 3. In the fifth column we weigh the observations by the inverse of their standard error (WLS). This approach gives less weight to less precise estimates, which helps adjusting for potential heteroskedasticity in our observations. The sixth column shows our results from estimation when the observations are weighted by the inverse of the number of estimates reported in a given study (wNOBS). This approach provides a more comparable basis for larger and smaller studies.

Considering the above-discussed results reported in Panel A of Table 2 we find little evidence of publication selection bias for the alpha estimates reported in the primary studies. The λ coefficients that capture the effect of selective publication are all small in magnitude and statistically insignificant. These findings are remarkably consistent across the alternative ways of estimating Equation 3. Thus, consistent with our preliminary findings based on the funnel plot in Figure 5, our formal tests provide evidence consistent with non-selective publication of estimated monthly alphas in the primary studies in our sample.

Panel A of Table 2 also shows the κ estimates that reflect the estimated magnitude of the monthly alphas adjusted for the publication selection bias. We observe that these estimates range between 0.301 and 0.369 and they are strongly statistically significant at a better than 1 percent level in all specifications. It is noteworthy that the unconditional mean of monthly alpha estimates of 0.36 highlighted in the histogram in Figure 3 falls within this range of estimates corrected for the publication selection bias. Considering perhaps the most conservative estimate reported in Panel A of Table 2 we observe that the lower bound of the bootstrapped confidence interval of the IV specification is 0.157. This suggests that the true hedge fund alpha is unlikely to be below 1.9% per annum (0.157*12). These results further strengthen our earlier conclusion that the alpha estimates in our sample are not contaminated by selective publication and hedge funds do earn positive alphas for their investors.

Our estimates of Equation 3 reported in Panel A of Table 2 are subject to several limitations. First, these tests of selective publication are based on an assumption of a linear relationship between the estimate and its standard error. In reality, this association may not be linear. For example, it may exhibit discontinuities around conventional levels of statistical significance, i.e. when the t-statistics approaches 1.96. Second, the IV specification may not fully remedy the endogeneity problem because it may arise for reasons other than the bias due to omitted variable related to the research design in primary studies. Gechert *et al.* (2022) point out that endogeneity may arise even when deliberately reporting spuriously precise estimates, for example due to reverse causality. Furthermore, since the standard error is itself an estimate endogeneity can also manifest itself through the measurement error. We address these shortcomings in Panel B of Table 2. To address the first issue, we use nonlinear techniques for publication bias correction. To address the second limitation, we use the p-uniform* approach recently developed by van Aert & van Assen (2020) that does not rely on the assumption of exogeneity.

In the first column of Panel B of Table 2 we report results based on the Top10 method proposed by Stanley *et al.* (2010). The method is based on a simple proposition that the bias arising from aggregating potentially selectively reported coefficients can be addressed by simply considering only the 10% most precise estimates. The second column of Panel B of Table 2 shows results based on the Weighted Average of the Adequately Powered (WAAP) model proposed by Ioannidis *et al.* (2017). Similarly to Top10 also WAAP is based on averaging only a subset of published coefficients. Ioannidis *et al.* (2017) examine statistical power of the results published in the field of economics and they propose dropping all estimates with statistical power lower than 80% and weighting the remaining estimates by the inverse of their variance. In the third column of Panel B of Table 2 we report the results from the Stem-based method recently developed by Furukawa (2020). The stem-based method builds on Stanley *et al.* (2010) but it aims at limiting the loss of sample variation that results from discarding 90% of the less precise estimates. Furukawa (2020) optimizes the trade-off between the bias and variance, discards only estimates that do not add value in the light of this trade-off and uses the remaining estimates to compute the average value. The first three columns in Panel B of Table 2 show estimates ranging from 0.310 to 0.355, which falls within the range documented earlier for the linear methods reported in Panel A (0.301 to 0.369). Thus, even based on these alternative methods we reach a similar conclusion on the limited impact of publication bias and on the values of alpha estimates corrected for a potential publication bias.

In the fourth column of Panel B of Table 2 we report results based on the endogenous kink model (Kinked-meta) proposed by Bom & Rachinger (2019). The model is based on the assumption that the relationship between an estimate and its standard error is only linear to some point because for some levels of reported coefficients there is no reason to expect the presence of publication bias. Hence, there is an endogenously determined cut-off value (or a "kink") at which the relationship changes. The Kinked-meta model yields some weak evidence (significant at 10% level) on the presence of selective publication (λ 0.183, SE 0.106). Nevertheless, even this approach yields a κ coefficient of 0.320, which is very close to the uncorrected mean of 0.36 and comfortably within the interval of 0.301 and 0.369 shown in Panel A.

The fifth column of Panel B of Table 2 shows our results for the selection model recently developed by Andrews & Kasy (2019). The model is based on the assumption that the probability of publishing an estimate depends on its statistical significance. The model identifies how likely it is for an estimate to fall into different intervals determined by the critical values of t-statistics. The model gives more weight to intervals that are underrepresented. Our results from the selection model suggest that statistically insignificant estimates may be somewhat less likely to get published than statistically significant estimates (63% probability vs. 100% probability). However, the corrected mean of alpha estimates decreases only slightly to 0.274. Hence, even this methodological approach does not suggest that inferences about the magnitude of alpha coefficients are greatly affected by selective publication.

The Andrews & Kasy (2019) model relies on several assumptions. It requires the estimates and their standard errors are statistically independent. It also assumes that the probability of publication is the same for all estimates in a given interval. We test these assumptions in Table A8 using the Kranz & Putz (2022) framework. These tests suggest some of the underlying assumptions of the selection model (especially the independence assumption) may be violated in many of our samples. Therefore, as robustness checks, we also use models that do not rely on the underlying assumption of no correlation between the estimates and their standard errors in the absence of publication bias.

The last column in Panel B of Table 2 represents the results of the p-uniform* model recently proposed by van Aert & van Assen (2020). The model is intended to detect "p-hacking", i.e. a greater tendency to publish statistically significant rather than insignificant results. It is based on evaluating the distribution of p-values around the 5% cut-off level that is conventionally used to assess statistical significance. A tendency to publish statistically significant results implies over-representation of p-values just below the 5% cut-off and under-representation of p-values just above it. P-uniform* corrects for this potential bias by assigning different weights to estimates of various degrees of statistical significance based on the estimated publication probability. This selection model is robust to the assumption of zero correlation between estimates and standard errors in the absence of any publication bias. Our results based on this methodological approach are consistent with our previous findings. The test statistic for the publication bias (denoted "L") is statistically insignificant, which again suggests that publication of alpha estimates in primary studies is not selective. In fact, this method suggests a somewhat higher value of 0.386 for alpha estimates corrected for the publication bias. Thus, we reach similar conclusions about the absence of publication selection bias and a somewhat similar estimate of the true mean value of the alpha coefficient even when using the p-uniform* method, which does not require the exogeneity assumption for the standard errors to be satisfied.

In contrast to the more conventional linear approaches reported in Panel A of Table 2, the more sophisticated non-linear approaches shown in Panel B of Table 2 do not require the linearity and exogeneity assumptions to be met. Overall, these approaches lead to fairly similar conclusions about the limited impact of selective publication on the alpha estimates reported in primary studies. Only the Kinked-meta model shows some marginally significant evidence of publication selection bias. However, even this approach does not dramatically alter the estimated value of alpha coefficients corrected for the publication bias. Furthermore, the selection model suggests that statistically insignificant estimates may be somewhat less likely to get published. However, the estimate for the mean alpha coefficient does not dramatically change after correcting for this bias. The interval of corrected alpha estimates based on the more sophisticated approaches reported in Panel B is slightly wider and it ranges from 0.274 to

0.386. However, both the upper bound and the lower bound of this interval are fairly close to the unconditional mean of 0.36. These results thus provide further support for our conclusion that inferences about the magnitude of the alpha coefficient in the literature on hedge fund performance are not significantly affected by publication selection bias.

These results are remarkable especially when contrasted with the abundant empirical evidence on the prevalence of publication selection bias in a multitude of other settings in economics and finance, e.g. Stanley (2001, 2005); Stanley & Doucouliagos (2010); Havranek (2015); Brodeur *et al.* (2016); Bruns & Ioannidis (2016); Stanley & Doucouliagos (2017); Christensen & Miguel (2018); Brodeur *et al.* (2020a); Blanco-Perez & Brodeur (2020), and Zigraiova *et al.* (2021). Our main results indicate that publication may not be selective when there is no strong *a priori* theoretical prediction about the sign of estimated coefficients, which may induce greater readiness to publish statistically insignificant results.

To further strengthen our analysis, we perform several robustness checks intended to ensure that our results are not driven by the heterogeneity in the mix of various alpha coefficients estimated in the primary studies using a wide range of techniques. Heterogeneity in estimation may potentially lead to offsetting biases that would compromise our ability to detect selective publication in the full sample. For example, it is acknowledged that the p-uniform* method tends to overestimate the measured effect when large heterogeneity is present among the estimates collected from the primary studies Carter *et al.* (2019). To further strengthen the confidence in our findings and to rule out the possibility that our tests are adversely affected by the diversity of the techniques used in estimating the alpha coefficients in the primary studies we proceed by analyzing more homogeneous subsets of alpha estimates to determine whether selective publication can be observed in any of these sub-samples.

6 Robustness Checks

We consider several sub-samples of more homogenous alpha estimates. First, we partition our sample based on whether the survivorship and/or backfilling biases are adjusted for in a given primary study. Since these data biases may potentially have a significant impact on the documented returns estimating our regressions for the two sub-samples separately lets us draw stronger inferences from our results. Second, we consider alpha coefficients estimated using two commonly used risk models. Hedge funds exhibit unusual risk exposures to various risk dimensions and so the choice of a risk model may have an impact on the estimated abnormal return. Third, we recompute our results for the sub-sample of alpha coefficients estimated with the use of instrumental variables. Prior research shows that IV-based estimates are more likely to suffer from a publication bias because they tend to have larger standard errors (Brodeur *et al.*, 2020a). We examine whether we detect selective publication in this subset of estimates.

6.1 Survivorship and Backfilling Biases

Prior research has long argued that the survivorship and backfilling biases may have a substantial impact on hedge fund performance estimates (Fung & Hsieh, 2004a; Aggarwal & Jorion, 2010; Kosowski *et al.*, 2007). The survivorship bias arises when a data sample excludes performance results of funds that are no longer in existence. From the perspective of data providers, excluding these funds from their database is sensible because funds that no longer operate are not interesting for investors any more. Nevertheless, since performance of funds that stop reporting information on their performance to the database may systematically differ from performance of surviving funds purging this information biases the research results based on the database. The survivorship bias arises when funds undergo an "incubation period" intended to accumulate performance track record before they are offered to investors. If performance history is backfilled into the database only for those funds that succeed in the incubation period the database overstates performance of the entire hedge fund population in the early years of their existence.

In the following analysis we consider separately a sub-sample of alpha estimates that explicitly controls for the survivorship and/or backfilling biases. Then we consider only those alpha estimates that do not adjust for these biases. Given that the survivorship and backfilling biases may have a significant impact on the estimated alpha coefficients considering only one sub-sample at a time makes the individual alpha estimates more homogenous. We examine whether our main conclusions on the limited publication selection bias are robust to testing these relationships within the two sub-samples.

In Table 3 we report our result for the subset of alpha estimates that adjust for the survivorship and/or backfilling biases. The results based on the conventional approaches reported

Table 3: Survivorship and/or Backfiling Bias Treated

Panel A: linear models	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.00148	-0.0846	0.0795	0.493	0.472**	0.0814
	(0.180)	(0.0961)	(0.189)	(0.402)	(0.233)	(0.195)
	[-0.435, 0.568]			[-0.492, 2.534]	[-0.092, 0.948]	[-0.406, 0.561]
				{-0.342, 1.566}		
Effect beyond bias (κ)	0.329***	0.351***	0.300***	0.194*	0.241***	0.300***
	(0.0373)	(0.0252)	(0.0622)	(0.106)	(0.0337)	(0.0487)
	[0.243, 0.411]			[-0.104, 0.392]	[0.133, 0.377]	[0.183, 0.403]
First-stage robust F-stat				11.29		
Observations	605	605	605	565	605	605

Panel B: non-linear models	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.519***	P = 0.632	NA
				(0.125)	(0.115)	(NA)
Effect beyond bias	0.267***	0.248***	0.220***	0.234***	0.262***	0.325***
	(0.028)	(0.017)	(0.068)	(0.012)	(0.029)	(0.048)
Observations	605	605	605	605	605	605

Notes: Sample in which both biases are treated for (either the survivorship or the backfiling bias is treated or both biases are treated for). 50 studies used. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in Panel A convey a fairly similar albeit slightly weaker message as the results based on the full sample (reported in Table 2). In Panel A of Table 3 all but one λ coefficient are statistically insignificant. Only for the WLS model, which weighs the observations by the inverse of their standard error, we observe a significant λ coefficient of 0.472 (SE 0.233, significant at 5% level). This finding provides limited evidence on some publication bias within the subsets of alpha estimates adjusted for the survivorship and/or backfilling biases weighted by their precision. Furthermore, among the results based on the non-linear approaches reported in Panel B of Table 3 we observe one statistically significant λ for the Kinked-meta model of 0.519 (SE 0.234, significant at 1% level). Similarly, to the WLS model also Kinked-meta attributes different weights to alpha estimates based on their precision. However, overall, we find only limited evidence of selective publication of alpha coefficients within the sub-sample of estimates that adjust for the survivorship and/or backfilling biases.

Results presented in Table 3 also indicate that limiting the analysis to the subsample of estimates adjusted for the survivorship and/or backfilling biases does not dramatically affect the conclusions about the magnitude of the alpha coefficients. The κ coefficients reported in Table 3 reflect the average alpha coefficients adjusted for the publication selection bias range

Table 4: Survivorship and Backfiling Biases Untreated

Panel A: linear models						
	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.0168	0.0491	0.0602	-0.466	0.438	0.0337
	(0.363)	(0.486)	(0.170)	(0.725)	(0.492)	(0.140)
	[-0.941, 0.870]			[NA]	[-0.558, 1.749]	[-0.844, 0.671]
				{-5.850, NA}		
Effect beyond bias (κ)	0.416***	0.400***	0.436***	0.521***	0.334***	0.443***
	(0.0782)	(0.114)	(0.0686)	(0.163)	(0.0559)	(0.0563)
	[0.245, 0.615]			[NA]	[0.112, 0.482]	[0.326, 0.554]
First-stage robust F-stat				3.76		
Observations	414	414	414	414	414	414
Panel B: non-linear models						
	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.343*	P = 0.719	NA
				(0.191)	(0.124)	(NA)
Effect beyond bias	0.301***	0.359***	0.331***	0.351***	0.282***	0.507***
	(0.038)	(0.012)	(0.040)	(0.009)	(0.086)	(0.077)
Observations	414	414	414	414	414	414

Notes: Sample in which biases are not treated for (neither the survivorship nor the backfiling bias is treated for). 29 studies used. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

between 0.194 and 0.351. This range is only slightly lower than the corresponding interval for κ coefficients based on the full sample between 0.274 and 0.386 reported in Table 2. The most significant deviation from this pattern is the slightly lower and only marginally significant κ coefficient based on the IV estimate that uses the inverse of the square root of the number of observations as an instrument for the standard error. This κ of 0.194 (SE 0.106, significant at 10% level) is reported in the fourth column in Panel A of Table 3.

In Table 4 we report results based on the sub-sample of alpha estimates that do not explicitly control neither for the survivorship nor for the backfilling biases. The conclusion based on this sub-sample are very similar to the main results reported in Table 2. In line with the full sample results, the λ coefficients are statistically insignificant with the exception of the one based on the Kinked-meta model, which is equal to 0.343 and similarly to the full-sample result it is marginally significant at 10% (SE 0.191). Furthermore, the κ coefficients reported in Table 4 range between 0.282 and 0.521. Relative to the corresponding range for the κ coefficients based on the full sample this range is slightly wider. The difference is mainly driven by the higher upper bound, which is consistent with the proposition that studies that control for the survivorship and/or backfilling biases tend to report lower alpha estimates than those that do not. Overall, these findings suggest that the alpha coefficients that are not adjusted for backfilling and survivorship biases are not reported selectively.

6.2 Risk Models

One of the key methodological issues in hedge fund performance research concerns the choice of risk models used to adjust for the systematic risk that a given investment strategy involves. These models define the risk dimensions considered relevant for a given investment strategy. Prior hedge fund performance research uses several risk models. Models that feature fewer risk factors (e.g. the CAPM, the three-factor, and the four-factor model) are well-established in general asset pricing and investment research, which implies that the alpha coefficients based on these models are easily comparable with alpha coefficients estimated to evaluate performance of other types of investments, e.g. mutual funds. On the other hand, hedge funds commonly employ complex investment strategies that may exhibit unusual risk profiles and exposures to risk dimensions that are not essential for conventional asset classes. Thus, standard risk models may not fully capture the exposure of hedge funds' investment strategies to all relevant risk dimensions. More complex risk models featuring additional risk dimensions designed specifically to measure hedge fund performance may thus be more effective in capturing the plurality of risk exposures that hedge fund strategies involve. The choice of a risk model is thus one of the important drivers for the heterogeneity in the alpha coefficients that we collect from primary studies. In evaluating the robustness of our findings to various ways of reducing heterogeneity in our sample we re-estimate our regressions using two sub-samples of alpha coefficients estimated based on two frequently used risk models: (i) the one-factor model, and (ii) the seven-factor model.

Table 5 shows the results of our tests of selective publication for the subset of alpha coefficients based on the one-factor model. In these tests we include all the alpha estimates that use a single risk factor based on market portfolio returns, i.e. both the estimates that use raw market returns and those that use market returns in excess of the risk-free rate. These methodological modifications are relatively small and so we do not expect them to have a substantial impact on the reported alpha coefficients. In line with our earlier results, we do not find evidence of a significant publication bias for these narrowly defined sub-sample of alpha coefficients. The λ

Table 5: One-Factor Model

Panel A: linear models						
	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.456	-0.328	0.0326	-1.115	0.453	-0.338
	(0.488)	(0.657)	(0.494)	(0.819)	(0.602)	(0.249)
	[-1.649, 1.468]			[NA]	[-0.776, 2.269]	[-1.247, 0.909]
				{-3.629, 0.102}		
Effect beyond bias (κ)	0.562***	0.534***	0.411***	0.707***	0.404***	0.482***
	(0.0447)	(0.145)	(0.119)	(0.175)	(0.0883)	(0.0931)
	[0.465, 0.642]			[0.324, 1.411]	[0.044, 0.515]	[0.252, 0.702]
First-stage robust F-stat				14.47		
Observations	167	167	167	167	167	167
Panel B: non-linear models						
	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.450	P = 0.613	L = 0.188
				(0.347)	(0.177)	(p = *0.911*)
Effect beyond bias	0.446***	0.454***	0.349**	0.405***	0.426***	0.427***
	(0.072)	(0.030)	(0.163)	(0.028)	(0.088)	(0.103)
Observations	167	167	167	167	167	167

Notes: Sample in which the one-factor model or its modifications are used to estimate the alpha. 18 studies used. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

coefficients that capture the impact of a potential publication bias are all statistically insignificant. This suggests that the alpha estimates based on a one-factor model are not reported selectively in prior literature.

Table 5 also shows that the κ coefficients that reflect the estimated true magnitude of the alpha estimates corrected for the potential bias range from 0.349 to 0.707. This interval includes the unconditional mean of all the alpha estimates in our sample of 0.36. However, these estimates are somewhat higher than the ones we document for the full sample in Table 2. This may suggest that the single market-based risk factor does not fully control for the systematic risk hedge fund strategies involve and so the abnormal return based on the model is higher. Overall, these results provide additional support for the conclusion that the alpha estimates reported in prior literature are not subject to selective publication.

Prior literature acknowledges that the complexity and the dynamic nature of hedge funds' investment strategies may induce exposure to risk dimensions that are not included in conventional risk models. Prior research thus proposes alternative risk models designed specifically for investment strategies common in hedge funds. The most notable example of these models is the seven-factor model (Fung & Hsieh, 2004a; Fung *et al.*, 2008). The model comprises the following risk factors: (i) the stock market excess return, (ii) the spread between the small capitalization and large capitalization stock returns, the excess return pairs of look-back call and put options (iii) on currency futures, (iv) on commodity futures, and (v) on bond futures, (vi) the duration-adjusted change in the yield spread of the U.S. 10-year Treasury bond over the 3-month T-bill, and (vii) the duration-adjusted change in the credit spread of Moody's BAA bond over the 10-year Treasury bond. These risk factors are intended to capture risk exposures of a broad set of hedge fund types ranging from equity long-short funds to managed futures funds.

Table 6: Seven-Factor Model

Panel A: linear models	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.142	-0.0729	0.305*	0.624	0.0683	0.226
	(0.137)	(0.0547)	(0.155)	(0.557)	(0.296)	(0.265)
	[-0.666, 0.555]			[NA]	[-0.571, 0.948]	[-0.732, 0.644]
				{0.0184, NA}		
Effect beyond bias (κ)	0.326***	0.308***	0.200**	0.128	0.284***	0.222***
	(0.0392)	(0.0141)	(0.0730)	(0.150)	(0.0330)	(0.0641)
	[0.239, 0.413]			[NA]	[0.132, 0.361]	[0.073, 0.375]
First-stage robust F-stat				3.41		
Observations	298	298	298	298	298	298
Panel B: non-linear models	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.019	P = 0.900	L = 0.269
				(0.173)	(0.212)	(p = *0.874*)
Effect beyond bias	0.229***	0.297***	0.325***	0.298***	0.302***	0.305***
	(0.036)	(0.013)	(0.059)	(0.008)	(0.059)	(0.040)
Observations	298	298	298	298	298	298

Notes: Sample in which the seven-factor model or its modifications are used to estimate the alpha. 33 studies used. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 6, we report results of our tests of selective publication for the alpha coefficients based on the seven-factor model. Similarly to the results on the one-factor model reported in Table 5, the results in Table 6 show little evidence of publication selection bias. The λ coefficients are insignificant with the exception of the between effect estimation (BE) that produces marginally significant λ of 0.305 (SE 0.155, significant at 10% level). Furthermore, the κ coefficients that reflect the expected value alpha coefficients after adjusting for selective publication range from 0.128 to 0.326, which is lower than the corresponding range in Table 5. The range is also below the unconditional mean of all monthly alpha estimates in our sample of 0.36. This magnitude of the κ coefficients is unlikely to be driven by selective publication.

Taken together, our results do not suggest that the heterogeneity in methodological approaches used for estimating alpha coefficients reported in primary studies is the underlying reason for not detecting any publication bias. We do not observe a significant publication bias even when concentrating on fairly homogeneous sub-samples of alphas that are estimated using one of the common risk models.

6.3 Instrumental Variables

We further consider a sub-sample of alphas estimated based on IV, for which selective publication is particularly likely. Prior research shows that IV-based estimates tend to suffer from publication bias more frequently than estimates based on other techniques (Brodeur *et al.*, 2020a). The authors argue that research methods that offer researchers more degrees of freedom are more likely to suffer from selective publication as researchers may exercise discretion in choosing research designs that help them achieve results that may be viewed as more attractive for publication. The choice of an IV and the specific way of measuring it give researches considerable leeway. Researchers may choose to report IV-based estimates that are consistent with their prior beliefs or that are otherwise more attractive for publication. Brodeur *et al.* (2020a) show that when IVs are relatively weak, the second stage results are likely to be close to the conventional thresholds for statistical significance, which is consistent with selectivity in the process that determines what coefficients eventually get published.

Motivated by this argument recently proposes in research literature we test for selective publication within the sub-sample of IV-based alpha coefficients. Primary studies typically use higher moments of the distribution of returns, such as skewness and kurtosis, as IV for the excess returns of the mimicking portfolios. This approach follows earlier research that shows that higher moments of the returns distribution are valid instruments and they are effective in removing errors-in-variables problem (Durbin, 1954; Pal, 1980; Dagenais & Dagenais, 1997).

Our results reported in Table 7 are consistent with the proposition in prior literature that IV-based estimates tend to exhibit a greater publication selection bias. Five out of seven λ coefficients are positive and statistically significant at 5% level or better. The positive association between reported alphas and their standard errors indicates that highly positive alpha

Table 7: Methods Using Instrumental Variables

Panel A: linear models						
	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	1.378***	1.307**	3.514	2.459	2.418***	1.445***
	(0.221)	(0.283)	(6.083)	(2.111)	(0.299)	(0.178)
	[0.760, 1.634]			[NA]	[1.886, 2.945]	[0.776, 1.647]
				{-1.512, 6.430}		
Effect beyond bias (κ)	0.127	0.144	-0.411	-0.137	-0.0601	0.104
	(0.0851)	(0.0691)	(1.525)	(0.438)	(0.0875)	(0.102)
	[-0.102, 0.235]			[NA]	[-0.195, 0.260]	[-0.113, 0.267]
First-stage robust F-stat				155.41		
Observations	46	46	46	46	46	46
Panel B: non-linear models						
	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				2.418***	P = 0.341	L = 3.551
				(0.431)	(0.103)	(p = *0.169*)
Effect beyond bias	-0.036	0.018	0.078	-0.060	0.231***	0.298***
	(0.027)	(0.067)	(0.091)	(0.048)	(0.088)	(0.071)
Observations	46	46	46	46	46	46

Notes: Sample where the instrumental variable approach (including 2SLS, GMM, Hasuman) is used for estimation of the alpha. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

estimates tend to be reported when they are rather imprecise, i.e. they have a large standard error. Such a pattern is characteristic of selective publication. Furthermore, we also observe that for the sub-sample of IV-based estimates the magnitude of the κ coefficients that represent the expected value alpha estimates after adjusting for selective publication is substantially lower than in our main results. The κ coefficients reported in Table 7 range from -0.411 to 0.298, many of them are close to 0, and five out of twelve are actually negative. This suggests that after correcting the IV-based alpha estimates for selective publication there is only limited evidence that they actually are positive and statistically significant. In fact, in contrast to our previous results, only two out of twelve κ coefficients are statistically different from zero. Consistent with the *a priori* expectations this evidence suggests that the composition of the pool of published IV-based alpha estimates tends to be affected by selective publication. These findings thus provide one of the first pieces of out-of-sample evidence in support of the recent proposition that IV-based estimates are more likely to suffer from publication bias than estimates based on other techniques Brodeur *et al.* (2020a).

In contrast, our results based on the sub-sample of the remaining alpha coefficients that are not estimated with the use of IV reported in Table 8 are in line with our main results. All

Table 8: Methods Not Using Instrumental Variables

Panel A: linear models	OLS	FE	BE	IV	WLS	wNOBS
Publication bias (λ)	-0.0335	-0.0230	0.0147	0.203	0.284	0.0364
	(0.191)	(0.220)	(0.132)	(0.351)	(0.321)	(0.131)
	[-0.543, 0.442]			[-0.493, 0.999]	[-0.453, 1.080]	[-0.351, 0.458]
				{-0.597, 1.003}		
Effect beyond bias (κ)	0.366^{***}	0.363^{***}	0.358^{***}	0.305^{***}	0.308^{***}	0.353^{***}
	(0.0429)	(0.0553)	(0.0487)	(0.0849)	(0.0440)	(0.0382)
	[0.277, 0.459]			[0.138, 0.469]	[0.192, 0.424]	[0.271, 0.435]
First-stage robust F-stat				12.98		
Observations	973	973	973	933	973	973

Panel B: non-linear models	Top10	WAAP	Stem-based	Kinked-meta	Selection model	p-uniform*
Publication bias				0.166	P = 0.636	L = 0.378
				(0.108)	(0.100)	(p = *0.828*)
Effect beyond bias	0.333^{***}	0.329^{***}	0.355^{***}	0.324^{***}	0.276^{***}	0.388^{***}
	(0.026)	(0.009)	(0.092)	(0.008)	(0.030)	(0.045)
Observations	973	973	973	973	973	973

Notes: Sample where the instrumental variable approach is not used for estimation of the alpha (mostly ordinary least squares). 74 studies used. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

λ coefficients are statistically insignificant, which indicates that these alpha estimates are not substantially affected by the publication selection bias. In comparison to the full-sample results, within this sub-sample even the λ coefficient based on the Kinked-meta model is statistically insignificant. Furthermore, the κ coefficients fall within a fairly narrow range between 0.276 and 0.388, which is very similar to the full sample result. Taken together, the sub-sample of alpha coefficients that are not estimated based on IV do not seem to be affected by publication bias and their mean value corrected for any (small) biases are quite close to the unconditional sample mean of 0.36.

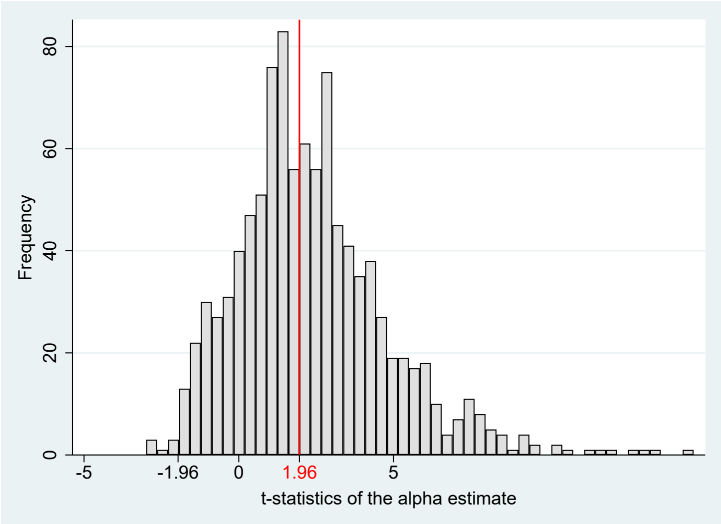
6.4 P-Hacking

Finally, to complement our analysis we follow recent advances in econometric techniques that propose new ways of detecting selective publication by examining evidence on so called "p-hacking". P-hacking occurs when researchers have a tendency to selectively report statistically significant estimates and discard those that narrowly miss the cutoffs for statistical significance. Publication selection bias and p-hacking are related but distinct concepts. They both distort the pool of estimates published in primary studies. However, they differ in the types of estimates that end up being suppressed and so they are tested for differently. P-hacking leads to discontinuities in the distribution of p-values or test statistics, which can be identified in the distribution of test statistics and formally tested for using dedicated statistical techniques. The p-hacking tests are conceptually different from the tests based on examining the relationship between the estimated coefficients and their standard errors. Thus they can be viewed as alternative ways of detecting distortions in publishing empirical results. Therefore we believe that an analysis of p-hacking complements well our preceding analysis of publication selection bias.

In Figure 6, we provide a visual representation of the distribution of t-statistics reported in the primary studies. In the presence of p-hacking we would expect t-statistics just above the conventional levels of statistical significance to be over-represented and t-statistics just below these levels to be underrepresented. The vertical red line in Figure 6 represents arguably the most important cutoff level for statistical significance of 1.96. In addition to this cutoff for statistical significance at the conventional 5% level, we also examine potential discontinuities around 0 as researchers may have a tendency to over-report positive or negative alpha estimates. A simple visual observation of the histogram in Figure 6 suggests some departures from normality in the distribution of reported t-statistics. The distribution is positively skewed. Furthermore, it is not quite smooth, especially in the interval between 1.5 and 2.5. However, discontinuities do not seem to be clustered around the typical critical values. The t-statistics both just above and just below 1.96 seem to be less frequent than expected. Crucially, we observe no asymmetry in the distribution of t-statistics around this critical value. Thus, consistent with our previous results this evidence indicates little p-hacking in research measuring hedge fund performance.

To formally test for the presence of p-hacking we follow a recent paper by Elliott *et al.* (2022) who theoretically derive a new restriction on the distribution of p-values under the condition of no p-hacking, that is when the results from regression specifications that yield statistically significant p-values are no more likely to be published than statistically insignificant results. The test is based on the conditional chi-squared tests by Cox & Shi (2022). The authors analytically characterize general sufficient conditions under which the distributions of p-values across studies is continuous and non-increasing in the absence of p-hacking. They demonstrate that absent p-hacking the distributions of p-values for t-tests is monotone and the magnitude of p-values and their derivatives are restricted by upper bounds. Testing this relationship requires

Figure 6: The distribution of t-statistics



Notes: The figure represents the distribution of t-statistics of the reported estimates of the alpha. Red lines represents critical value of 1.96 associated with significance at the 5% level and the value of 0 associated with changing the sign of the estimate. We exclude estimates with large t-statistics from the figure for ease of exposition but include them in statistical tests.

specifying the number of bins in the distribution, which we set to 10, 15, and 20.

Table 9 shows our results of the p-hacking tests based on Elliott *et al.* (2022). Consistent with our earlier results, we find no evidence on selective reporting of results. Both the test for non-increasingness of the distribution of p-values and the test for monotonicity and bounds show no evidence for any of the three bins. Hence, in line with our previous conclusions based on Figure 6 we conclude that the pool of hedge fund alpha estimates published in primary studies does not exhibit signs of p-hacking. This finding provides additional empirical support for our main conclusion that hedge fund performance results are not substantially contaminated by selective reporting.

Table 9: Tests of P-Hacking

	20 bins	15 bins	10 bins
Test for non-increasingness	0.469	0.179	0.403
Test for monotonicity and bounds	0.242	0.223	0.481
Observations ($p <= 0.15$)	663	663	663
Total observations	1,019	1,019	1,019

Notes: Results of p-hacking tests based on Elliott *et al.* (2022) based on the whole sample.

7 Conclusion

We perform a meta-analysis of prior empirical studies evaluating hedge fund performance. We examine whether published estimates of hedge fund alphas are affected by publication selection bias and by data biases. We observe that, prior research detects publication selection bias in a wide range of economic and finance settings, e.g. Stanley (2001, 2005); Stanley & Doucouliagos (2010); Havranek (2015); Brodeur *et al.* (2016); Bruns & Ioannidis (2016); Stanley & Doucouliagos (2017); Christensen & Miguel (2018); Brodeur *et al.* (2020a); Blanco-Perez & Brodeur (2020), and Zigraiova *et al.* (2021). In contrast to these findings, using a wide range of techniques and data partitions we do not detect selective publication in hedge fund performance literature with the exception of for estimates based on instrumental variables (IV). In contrast, we provide evidence that not controlling for the potential biases in the underlying data (e.g. the backfilling bias and the survivorship bias) affects reported alpha coefficients.

Fragmentation of hedge fund performance data and the wide range of alternative approaches for controlling for risk give researchers considerable discretion over the design of their research. This potentially creates opportunities for selective publication because the use of various estimation techniques based on different data sources may yield diverse results, some of which may be more attractive for publication than others. Our results demonstrate that despite the prevalence of the publication selection bias in numerous other research settings, publication may not to be selective when there is no strong *a priori* theoretical prediction about the sign of estimated coefficients, which may induce greater readiness to publish statistically insignificant results.

The heterogeneity in methodological approaches and data sources used in estimating hedge funds' alphas opens up additional research opportunities. Future research can examine whether and how the various aspects of methodological choices affect the magnitude of reported alpha coefficients. Our aim in this paper is to propose a representative alpha coefficient that is aggregated across the plurality of these approaches and corrected for the publication and data biases. Therefore, in this study we provide robustness checks based on sub-samples that narrow down the pool of collected alpha estimates to more homogeneous subsets but we do not explicitly exploit the full sample heterogeneity to analyze and draw conclusions about individual subsets or about the relative magnitude of alpha coefficients for the individual subsets. We leave the analysis of the impact of this heterogeneity on the reported alpha coefficients for future research that can examine the importance of various dimensions of methodological choices on the alpha coefficients reported in primary studies.

References

Ackermann, C., R. McEnally, & D. Ravenscraft (1999): "The performance of hedge funds: Risk, return, and incentives." *The journal of Finance* **54(3)**: pp. 833–874.

van Aert, R. C. & M. van Assen (2020): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Working paper*, Tilburg University & Utrecht University.

Agarwal, V., Y. Arisoy, & N. Y. Naik (2017a): "Volatility of aggregate volatility and hedge fund returns." *Journal of Financial Economics* **125(3)**: pp. 491–510.

Agarwal, V., Y. E. Arisoy, & N. Y. Naik (2017b): "Volatility of aggregate volatility and hedge fund returns." *Journal of Financial Economics* **125(3)**: pp. 491–510.

Agarwal, V., G. Bakshi, & J. Huij (2009a): "Do higher-moment equity risks explain hedge fund returns?" *Robert H. Smith School Research Paper No. RHS* pp. 06–153.

Agarwal, V., N. D. Daniel, & N. Y. Naik (2009b): "Role of managerial incentives and discretion in hedge fund performance." *The Journal of Finance* **64(5)**: pp. 2221–2256.

Agarwal, V., V. Fos, & W. Jiang (2013): "Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings." *Management Science* **59(6)**: pp. 1271–1289.

Agarwal, V., K. A. Mullally, N. Y. Naik *et al.* (2015): "The economics and finance of hedge funds: A review of the academic literature." *Foundations and Trends® in Finance* **10(1)**: pp. 1–111.

Agarwal, V. & N. Y. Naik (2000): "On taking the â€śalternativeâ€ť route: the risks, rewards, and performance persistence of hedge funds." *The Journal of Alternative Investments* **2(4)**: pp. 6–23.

Agarwal, V. & N. Y. Naik (2004): "Risks and portfolio decisions involving hedge funds." *The Review of Financial Studies* **17(1)**: pp. 63–98.

Aggarwal, R. K. & P. Jorion (2010): "Hidden survivorship in hedge fund returns." *Financial Analysts Journal* **66(2)**: pp. 69–74.

Ahoniemi, K. & P. Jylha (2014): "Flows, Price Pressure, and Hedge Fund Returns." *Financial Analysts Journal* **70(5)**: pp. 73–93.

Aiken, A., C. P. Clifford, & J. Ellis (2013): "Out of the Dark: Hedge Fund Reporting Biases and Commercial Databases." *Review of Financial Studies* **26(1)**: pp. 208–243.

Amin, G. S. & H. M. Kat (2003a): "Hedge Fund Performance 1990â€"2000: Do the â€śMoney Machinesâ€ť Really Add Value?" *Journal of Financial and Quantitative Analysis* **38(2)**: pp. 251–274.

Amin, G. S. & H. M. Kat (2003b): "Welcome to the dark side: Hedge fund attrition and survivorship bias over the period 1994–2001." *The Journal of Alternative Investments* **6(1)**: pp. 57–73.

Ammann, M. & P. Moerth (2005): "Impact of fund size on hedge fund performance." *Journal of Asset Management* **6(3)**: pp. 219–238.

Ammann, M. & P. Moerth (2008a): "Impact of Fund Size and Fund Flows on Hedge Fund Performance." *The Journal of Alternative Investments* **11(1)**: pp. 78–96.

Ammann, M. & P. Moerth (2008b): "Performance of Funds of Hedge Funds." *The Journal of Wealth Management* **11(1)**: pp. 46–63.

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM." *The Review of Economics and Statistics* **100(2)**: pp. 337–348.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–2794.

Aragon, G. O. (2007): "Share restrictions and asset pricing: Evidence from the hedge fund industry." *Journal of Financial Economics* **83(1)**: pp. 33–58.

Aragon, G. O. & J. Spencer Martin (2012): "A unique view of hedge fund derivatives usage: Safeguard or speculation?" *Journal of Financial Economics* **105(2)**: pp. 436–456.

Asness, C., R. Krail, & J. Liew (2001): "Do Hedge Funds Hedge?" *The Journal of Portfolio Management* **28(1)**: pp. 6–19.

Avramov, D., L. Barras, & R. Kosowski (2013): "Hedge fund return predictability under the magnifying glass." *Journal of Financial and Quantitative Analysis* **48(4)**: pp. 1057–1083.

Avramov, D., R. Kosowski, N. Y. Naik, & M. Teo (2011): "Hedge funds, managerial skill, and macroeconomic variables." *Journal of Financial Economics* **99(3)**: pp. 672–692.

Bali, T. G., S. Brown, & K. O. Demirtas (2013): "Do Hedge Funds Outperform Stocks and Bonds?" *Management Science* **59(8)**: pp. 1887–1903.

Bali, T. G., S. J. Brown, & M. O. Caglayan (2011): "Do hedge funds' exposures to risk factors predict their future returns?" *Journal of financial economics* **101(1)**: pp. 36–68.

Bali, T. G., S. J. Brown, & M. O. Caglayan (2014): "Macroeconomic risk and hedge fund returns." *Journal of Financial Economics* **114(1)**: pp. 1–19.

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

Barth, D., J. Joenväärä, M. Kauppila, & R. Wermers (2020): "The hedge fund industry is bigger (and has performed better) than you think." *OFR WP* pp. 20–01.

Ben-David, I., J. Birru, & A. Rossi (2020): "The performance of hedge fund performance fees." *Technical report*, National Bureau of Economic Research.

Bernardo, A. E. & O. Ledoit (2000): "Gain, loss, and asset pricing." *Journal of political economy* **108(1)**: pp. 144–172.

Bhardwaj, G., G. B. Gorton, & K. G. Rouwenhorst (2014): "Fooling Some of the People All of the Time: The Inefficient Performance and Persistence of Commodity Trading Advisors." *Review of Financial Studies* **27(11)**: pp. 3099–3132.

Billio, M., L. Frattarolo, L. Pelizzon *et al.* (2014): "A time-varying performance evaluation of hedge fund strategies through aggregation." *Bankers, Markets & Investors* **(129)**: pp. 40–58.

Black, F. (1972): "Capital market equilibrium with restricted borrowing." *The Journal of business* **45(3)**: pp. 444–455.

Blanco-Perez, C. & A. Brodeur (2020): "Publication Bias and Editorial Statement on Negative Findings." *The Economic Journal* **130(629)**: pp. 1226–1247.

Blitz, D. (2018): "Are Hedge Funds on the Other Side of the Low-Volatility Trade?" *The Journal of Alternative Investments* **21(1)**: pp. 17–26.

Bollen, N. P. & V. K. Pool (2009): "Do hedge fund managers misreport returns? evidence from the pooled distribution." *The Journal of Finance* **64(5)**: pp. 2257–2288.

Bollen, N. P. B. & R. E. Whaley (2009): "Hedge Fund Risk Dynamics: Implications for Performance Appraisal." *The Journal of Finance* **64(2)**: pp. 985–1035.

Bom, P. R. D. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

Boyson, N., C. W. Stahel, & R. M. Stulz (2006): "Is there hedge fund contagion?"

Brav, A., W. Jiang, F. Partnoy, & R. Thomas (2008): "Hedge Fund Activism, Corporate Governance, and Firm Performance." *The Journal of Finance* **63(4)**: pp. 1729–1775.

Brodeur, A., N. Cook, & A. Heyes (2020a): "Methods Matter: P-Hacking and Causal Inference in Economics." *American Economic Review* **110(11)**: pp. 3634–3660.

Brodeur, A., N. Cook, & A. Heyes (2020b): "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* **110(11)**: pp. 3634–3660.

Brodeur, A., M. Le, M. Sangnier, & Y. Zylberberg (2016): "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

Brooks, C. & H. M. Kat (2002): "The statistical properties of hedge fund index returns and their implications for investors." *The Journal of Alternative Investments* **5(2)**: pp. 26–44.

Brown, R. (2012): "Framework for Hedge Fund Return and Risk Attribution." *The Journal of Investing* **21(4)**: pp. 8–23.

Brown, S., Y. Lu, S. Ray, & M. Teo (2018): "Sensation Seeking and Hedge Funds." *The Journal of Finance* **73(6)**: pp. 2871–2914.

Brown, S. J., W. N. Goetzmann, R. G. Ibbotson *et al.* (1999): "Offshore hedge funds: Survival and performance, 1989-95." *The Journal of Business* **72(1)**: pp. 91–117.

Brown, S. J., W. N. Goetzmann, & B. Liang (2005): "Fees on fees in funds of funds." In "The World Of Hedge Funds: Characteristics and Analysis," pp. 141–160. World Scientific.

Bruns, S. B. & J. P. A. Ioannidis (2016): "p-Curve and p-Hacking in Observational Research." *PloS ONE* **11(2)**: p. e0149144.

Buraschi, A., R. Kosowski, & F. Trojani (2014): "When There Is No Place to Hide: Correlation Risk and the Cross-Section of Hedge Fund Returns." *The Review of Financial Studies* **27(2)**: pp. 581–616.

Cao, C., B. Goldie, B. Liang, & L. L. Petrasek (2016): "What Is the Nature of Hedge Fund Manager Skills? Evidence from the Risk-Arbitrage Strategy." *Journal of Financial and Quantitative Analysis* **51(3)**: pp. 929–957.

Capocci, D., A. Corhay, & G. Hübner (2005): "Hedge fund performance and persistence in bull and bear markets." *The European Journal of Finance* **11(5)**: pp. 361–392.

Capocci, D. & G. Hübner (2004): "Analysis of hedge fund performance." *Journal of Empirical Finance* **11(1)**: pp. 55–89.

Carter, E. C., F. D. Schönbrodt, W. M. Gervais, & J. Hilgard (2019): "Correcting for bias in psychology: A comparison of meta-analytic methods." *Advances in Methods and Practices in Psychological Science* **2(2)**: pp. 115–144.

Cassar, G. & J. Gerakos (2011): "Hedge funds: Pricing controls and the smoothing of self-reported returns." *The Review of Financial Studies* **24(5)**: pp. 1698–1734.

Chen, Y., M. Cliff, & H. Zhao (2017): "Hedge Funds: The Good, the Bad, and the Lucky." *Journal of Financial and Quantitative Analysis* **52(3)**: pp. 1081–1109.

Chen, Y. & B. Liang (2007): "Do Market Timing Hedge Funds Time the Market?" *Journal of Financial and Quantitative Analysis* **42(4)**: pp. 827–856.

Chincarini, L. & A. Nakao (2011): "Measuring Hedge Fund Timing Ability Across Factors." *The Journal of Investing* **20(4)**: pp. 50–70.

Christensen, G. & E. Miguel (2018): "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* **56(3)**: pp. 920–980.

Clark, K. A. & K. D. Winkelmann (2004): "Active Risk Budgeting in Action." *The Journal of Alternative Investments* **7(3)**: pp. 35–46.

Connor, G. & M. Woo (2004): "An introduction to hedge funds." .

Cox, G. & X. Shi (2022): "Simple Adaptive Size-Exact Testing for Full-vector and Subvector Inference in Moment Inequality Models." *Review of Economic Studies* **(forthcoming)**.

Dagenais, M. G. & D. L. Dagenais (1997): "Higher moment estimators for linear regression models with errors in the variables." *Journal of Econometrics* **76(1-2)**: pp. 193–221.

Dichev, I. D. & G. Yu (2011): "Higher risk, lower returns: What hedge fund investors really earn." *Journal of Financial Economics* **100(2)**: pp. 248–263.

Diez De Los Rios, A. & R. Garcia (2011): "Assessing and valuing the nonlinear structure of hedge fund returns." *Journal of Applied Econometrics* **26(2)**: pp. 193–212.

Ding, B. & H. A. Shawky (2007): "The Performance of Hedge Fund Strategies and the Asymmetry of Return Distributions." *European Financial Management* **13(2)**: pp. 309–331.

Ding, B., H. A. Shawky, & J. Tian (2009): "Liquidity shocks, size and the relative performance of hedge fund strategies." *Journal of Banking & Finance* **33(5)**: pp. 883–891.

Do, V., R. Faff, & J. Wickramanayake (2005): "An empirical analysis of hedge fund performance: The case of Australian hedge funds industry." *Journal of Multinational Financial Management* **15(4)**: pp. 377–393.

Duarte, J., F. A. Longstaff, & F. Yu (2007): "Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller?" *The Review of Financial Studies* **20(3)**: pp. 769–811.

Durbin, J. (1954): "Errors in variables." *Revue de l'institut International de Statistique* pp. 23–32.

Dybvig, P. H. & J. E. Ingersoll Jr (1982): "Mean-variance theory in complete markets." *Journal of Business* pp. 233–251.

Edelman, D., W. Fung, D. Hsieh, & N. N. Naik (2012): "Funds of hedge funds: Performance, risk and capital formation 2005 to 2010." *Financial Markets and Portfolio Management* **26(1)**: pp. 87–108.

Edelman, D., W. Fung, & D. A. Hsieh (2013): "Exploring uncharted territories of the hedge fund Industry: Empirical characteristics of mega hedge fund firms." *Journal of Financial Economics* **109(3)**: pp. 734–758.

Edwards, F. R. & M. O. Caglayan (2001): "Hedge Fund Performance and Manager Skill." *Journal of Futures Markets* **21(11)**: pp. 1003–1028.

Egger, M., G. D. Smith, M. Schneider, & C. Minder (1997): "Bias in meta-analysis detected by a simple, graphical test." *BMJ* **315(7109)**: pp. 629–634.

Eling, M. & R. Faust (2010): "The performance of hedge funds and mutual funds in emerging markets." *Journal of Banking & Finance* **34(8)**: pp. 1993–2009.

Eling, M. & F. Schuhmacher (2007): "Does the choice of performance measure influence the evaluation of hedge funds?" *Journal of Banking & Finance* **31(9)**: pp. 2632–2647.

Elliott, G., N. Kudrin, & K. Wuthrich (2022): "Detecting p-hacking." *Econometrica* **(forthcoming)**.

Elton, E. J., M. J. Gruber, & C. R. Blake (2003): "Marginal stockholder tax effects and ex-dividend day behavior-thirty-two years later." In "AFA 2004 San Diego Meetings," .

Frydenberg, S., K. Hrafnkelsson, V. S. Bromseth, & S. Westgaard (2017): "Hedge Fund Strategies and Time-Varying Alphas and Betas." *The Journal of Wealth Management* **19(4)**: pp. 44–60.

Fung, H.-G., X. E. Xu, & J. Yau (2002): "Global Hedge Funds: Risk, Return, and Market Timing." *Financial Analysts Journal* **58(6)**: pp. 19–30.

Fung, W. & D. A. Hsieh (1997): "Empirical characteristics of dynamic trading strategies: The case of hedge funds." *The review of financial studies* **10(2)**: pp. 275–302.

Fung, W. & D. A. Hsieh (1999): "A primer on hedge funds." *Journal of empirical finance* **6(3)**: pp. 309–331.

Fung, W. & D. A. Hsieh (2000): "Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases." *Journal of Financial and Quantitative analysis* **35(3)**: pp. 291–307.

Fung, W. & D. A. Hsieh (2001): "The risk in hedge fund strategies: Theory and evidence from trend followers." *The review of financial studies* **14(2)**: pp. 313–341.

Fung, W. & D. A. Hsieh (2002): "Hedge-fund benchmarks: Information content and biases." *Financial Analysts Journal* **58(1)**: pp. 22–34.

Fung, W. & D. A. Hsieh (2004a): "Hedge Fund Benchmarks: A Risk-Based Approach." *Financial Analysts Journal* **60(5)**: pp. 65–80.

Fung, W. & D. A. Hsieh (2004b): "Extracting Portable Alphas From Equity Long-Short Hedge Funds." *Journal of Investment Management* **2(4)**:

pp. 57–75.

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

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

Fung, W. K., D. A. Hsieh *et al.* (2006): "Hedge funds: An industry in its adolescence." *Economic Review-Federal Reserve Bank Of Atlanta* **91(4)**: p. 1.

Furukawa, C. (2020): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Working paper*, MIT.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2022): "Measuring capital-labor substitution: The importance of method choices and publication bias." *Review of Economic Dynamics* **45**: pp. 55–82.

Géhin, W. (2004): "A survey of the literature on hedge fund performance." *Available at SSRN 626441* .

Getmansky, M., P. A. Lee, & A. W. Lo (2015): "Hedge funds: A dynamic industry in transition." *Technical report*, National Bureau of Economic Research.

Goetzmann, W. N., J. E. Ingersoll Jr, & S. A. Ross (2003): "High-water marks and hedge fund management contracts." *The Journal of Finance* **58(4)**: pp. 1685–1718.

Goetzmann, W. N., J. E. Ingersoll Jr, M. Spiegel, & I. Welch (2002): "Sharpening sharpe ratios."

Griffin, J. M. & J. Xu (2009): "How smart are the smart guys? a unique view from hedge fund stock holdings." *The Review of Financial Studies* **22(7)**: pp. 2531–2570.

Gupta, B., B. Cerrahoglu, & A. Daglioglu (2003): "Evaluating Hedge Fund Performance: Traditional Versus Conditional Approaches." *The Journal of Alternative Investments* **6(3)**: pp. 7–24.

Havranek, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Havranek, T., T. D. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, & R. C. M. van Aert (2020): "Reporting Guidelines for Meta-Analysis in Economics." *Journal of Economic Surveys* **34(3)**: pp. 469–475.

Henriksson, R. D. & R. C. Merton (1981): "On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills." *Journal of business* pp. 513–533.

Hodder, J. E., J. C. Jackwerth, & O. Kolokolova (2014): "Recovering delisting returns of hedge funds." *Journal of Financial and Quantitative Analysis* **49(3)**: pp. 797–815.

Hong, X. (2014): "The dynamics of hedge fund share restrictions." *Journal of Banking & Finance* **49(C)**: pp. 82–99.

Huang, Y. S., C. R. Chen, & I. Kato (2017): "Different strokes by different folks: The dynamics of hedge fund systematic risk exposure and performance." *International Review of Economics & Finance* **48(C)**: pp. 367–388.

Ibbotson, R. G., P. Chen, & K. X. Zhu (2011): "The ABCs of Hedge Funds: Alphas, Betas, and Costs." *Financial Analysts Journal* **67(1)**: pp. 15–25.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The Power of Bias in Economics Research." *The Economic Journal* **127(605)**: pp. F236–F265.

Jame, R. (2018): "Liquidity Provision and the Cross Section of Hedge Fund Returns." *Management Science* **64(7)**: pp. 3288–3312.

Jensen, M. C. (1968): "The performance of mutual funds in the period 1945-1964." *The Journal of finance* **23(2)**: pp. 389–416.

Jobman, D. (2002): *The handbook of alternative investments*, volume 157. John Wiley & Sons.

Joenväärä, J., M. Kaupila, R. Kosowski, & P. Tolonen (2019): "Hedge fund performance: What do we know?" .

Joenvaara, J. & R. Kosowski (2021): "The Effect of Regulatory Constraints on Fund Performance: New Evidence from UCITS Hedge Funds." *Review of Finance* **25(1)**: pp. 189–233.

Joenvaara, J., R. Kosowski, & P. Tolonen (2019): "The Effect of Investment Constraints on Hedge Fund Investor Returns." *Journal of Financial and Quantitative Analysis* **54(4)**: pp. 1539–1571.

Jordan, A. & P. Simlai (2011): "Risk characterization, stale pricing and the attributes of hedge funds performance." *Journal of Derivatives & Hedge Funds* **17(C)**: pp. 16–33.

Jorion, P. & C. Schwarz (2013): "The delisting bias in hedge fund databases." *The Journal of Alternative Investments* **16(3)**: pp. 37–47.

Jorion, P. & C. Schwarz (2014): "The strategic listing decisions of hedge funds." *Journal of Financial and Quantitative Analysis* **49(3)**: pp. 773–796.

Jorion, P. & C. Schwarz (2019): "The fix is in: Properly backing out backfill bias." *The review of financial studies* **32(12)**: pp. 5048–5099.

Jylha, P., K. Rinne, & M. Suominen (2014): "Do Hedge Funds Supply or Demand Liquidity?" *Review of Finance* **18(4)**: pp. 1259–1298.

Kanuri, S. (2020): "Hedge Fund Performance in Japan." *Review of Pacific Basin Financial Markets and Policies* **23(03)**: p. 2050023.

Klein, P., D. Purdy, I. Schweigert, & A. Vedrashko (2015): "The Canadian Hedge Fund Industry: Performance and Market Timing." *International Review of Finance* **15(3)**: pp. 283–320.

Kooli, M. & I. Stetsyuk (2021): "Are hedge fund managers skilled?" *Global Finance Journal* **49(C)**: p. 100574.

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

Kotkatvuori-Ornberg, J., J. Nikkinen, & J. Peltomă¤ki (2011): "Geographical focus in emerging markets and hedge fund performance." *Emerging Markets Review* **12(4)**: pp. 309–320.

Kouwenberg, R. & W. T. Ziemba (2007): "Incentives and risk taking in hedge funds." *Journal of Banking & Finance* **31(11)**: pp. 3291–3310.

Kranz, S. & P. Putz (2022): "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics: Comment." *American Economic Review* **(forthcoming)**.

Liang, B. (1999): "On the performance of hedge funds." *Financial Analysts Journal* **55(4)**: pp. 72–85.

Liang, B. (2000): "Hedge funds: The living and the dead." *Journal of Financial and Quantitative analysis* **35(3)**: pp. 309–326.

Liang, B. (2003): "The accuracy of hedge fund returns." *The Journal of Portfolio Management* **29(3)**: pp. 111–122.

Liang, B. (2004): "Alternative investments: CTAs, hedge funds, and funds-of-funds." *Journal of Investment Management* **2(4)**: pp. 76–93.

Lim, J., B. A. Sensoy, & M. S. Weisbach (2016): "Indirect incentives of hedge fund managers." *The Journal of Finance* **71(2)**: pp. 871–918.

Ling, Y., J. Yao, & W. Liu (2015): "Chinese Hedge Funds - Performance and Risk Exposures." *The Chinese Economy* **48(5)**: pp. 330–350.

Lintner, J. (1965): "Security prices, risk, and maximal gains from diversification." *The journal of finance* **20(4)**: pp. 587–615.

Lo, A. W. (2001): "Risk Management for Hedge Funds: Introduction and Overview." *Financial Analysts Journal* **57(6)**: pp. 16–33.

Mahdavi, M. (2004): "Risk-adjusted return when returns are not normally distributed: Adjusted sharpe ratio." *The Journal of Alternative Investments* **6(4)**: pp. 47–57.

Malkiel, B. G. & A. Saha (2005): "Hedge funds: Risk and return." *Financial analysts journal* **61(6)**: pp. 80–88.

Malladi, R. (2020): "Luck versus Skill in Evaluating Hedge Fund ManagersÂ' Performance." *Journal of Business and Management* **26(1)**: pp. 22–39.

Markowitz, H. (1952): "Harry m. markowitz." *Portfolio selection, Journal of Finance* **7(1)**: pp. 77–91.

Meligkotsidou, L. & I. D. Vrontos (2008): "Detecting structural breaks and identifying risk factors in hedge fund returns: A Bayesian approach." *Journal of Banking & Finance* **32(11)**: pp. 2471–2481.

Mitchell, M. & T. Pulvino (2001): "Characteristics of Risk and Return in Risk Arbitrage." *The Journal of Finance* **56(6)**: pp. 2135–2175.

Mladina, P. (2015): "Illuminating Hedge Fund Returns to Improve Portfolio Construction." *The Journal of Portfolio Management* **41(3)**: pp. 127–139.

Molyboga, M. & C. L'Ahelec (2016): "A simulation-based methodology for evaluating hedge fund investments." *Journal of Asset Management* **17(6)**: pp. 434–452.

Mossin, J. (1966): "Equilibrium in a capital asset market." *Econometrica: Journal of the econometric society* pp. 768–783.

Mozes, H. A. (2013): "Decomposing Hedge Fund Returns: What Hedge Funds Got Right for the Past 20 Years." *The Journal of Investing* **22(3)**: pp. 9–20.

Pal, M. (1980): "Consistent moment estimators of regression coefficients in the presence of errors in variables." *Journal of Econometrics* **14(3)**: pp. 349–364.

Patton, A. J. (2009): "Are â€śmarket neutralâ€ť hedge funds really market neutral?" *The Review of Financial Studies* **22(7)**: pp. 2495–2530.

Patton, A. J. & T. Ramadorai (2013): "On the High-Frequency Dynamics of Hedge Fund Risk Exposures." *The Journal of Finance* **68(2)**: pp. 597–635.

Patton, A. J., T. Ramadorai, & M. Streatfield (2015): "Change you can believe in? hedge fund data revisions." *The Journal of Finance* **70(3)**: pp. 963–999.

Posthuma, N. & P. J. Van der Sluis (2003): "A reality check on hedge funds returns." *Available at SSRN 438840* .

Racicot, F.-E. & R. Theoret (2009): "Integrating volatility factors in the analysis of the hedge fund alpha puzzle." *Journal of Asset Management* **10(C)**: pp. 37–62.

Racicot, F.-E. & R. Theoret (2013): "The procyclicality of hedge fund alpha and beta." *Journal of Derivatives & Hedge Funds* **19(C)**: pp. 109–128.

Racicot, F.-E. & R. Theoret (2014): "Cumulant instrument estimators for hedge fund return models with errors in variables." *Applied Economics* **46(10)**: pp. 1134–1149.

Ranaldo, A. & L. Favre (2005): "Hedge Fund Performance and Higher-Moment Market Models." *The Journal of Alternative Investments* **8(3)**: pp. 37–51.

Roodman, D., J. G. MacKinnon, M. O. Nielsen, & M. D. Webb (2018): "Fast and wild: Bootstrap inference in Stata using boottest." *Queen's Economics Department Working Paper 1406*, Department of Economics, Queen's University, Canada: Kingston.

Rzakhanov, Z. & G. Jetley (2019): "Competition, scale and hedge fund performance: Evidence from merger arbitrage." *Journal of Economics and Business* **105(C)**: p. art. 105841.

Sabbaghi, O. (2012): "Hedge fund return volatility and comovement: recent evidence." *Managerial Finance* **38(1)**: pp. 101–119.

Sadka, R. (2010): "Liquidity risk and the cross-section of hedge-fund returns." *Journal of Financial Economics* **98(1)**: pp. 54–71.

Sadka, R. (2012): "Hedge-fund performance and liquidity risk." *Journal of Investment Management* **10(C)**: pp. 60–72.

Sandvik, S. H., S. Frydenberg, S. Westgaard, & R. K. Heitmann (2011): "Hedge Fund Performance in Bull and Bear Markets: Alpha Creation and Risk Exposure." *The Journal of Investing* **20(1)**: pp. 52–77.

Schneeweis, T. (1998): "Dealing with myths of hedge fund investment." *The Journal of Alternative Investments* **1(3)**: pp. 11–15.

Sharma, M. (2004): "Airap-alternative views on alternative investments." *Available at SSRN 497403* .

Sharpe, W. F. (1966): "Mutual fund performance." *The Journal of business* **39(1)**: pp. 119–138.

Spurgin, R. B. (2001): "How to game your sharpe ratio." *The Journal of Alternative Investments* **4(3)**: pp. 38–46.

Stafylas, D., K. Anderson, & M. Uddin (2018): "Hedge fund performance attribution under various market conditions." *International Review of Financial Analysis* **56(C)**: pp. 221–237.

Stafylas, D. & A. Andrikopoulos (2020): "Determinants of hedge fund performance during Â'goodÂ' and Â'badÂ' economic periods." *Research in International Business and Finance* **52(C)**: p. 101130.

Stanley, T. & H. Doucouliagos (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. & H. Doucouliagos (2017): "Neither fixed nor random: weighted least squares meta-regression." *Research Synthesis Methods* **8(1)**: pp. 19–42.

Stanley, T. D., S. B. Jarrell, & H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox." *The American Statistician* **64(1)**: pp. 70–77.

Stoforos, C. E., S. Degiannakis, & T. B. Palaskas (2017): "Hedge fund returns under crisis scenarios: A holistic approach." *Research in International Business and Finance* **42(C)**: pp. 1196–1207.

Stulz, R. M. (2007): "Hedge funds: Past, present, and future." *Journal of Economic Perspectives* **21(2)**: pp. 175–194.

Sullivan, R. N. (2021): "Hedge Fund Alpha: Cycle or Sunset?" *The Journal of Alternative Investments* **23(3)**: pp. 55–79.

Sun, L. (2018): "Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models." *Stata Journal* **18(4)**: pp. 803–825.

Sun, Z., A. Wang, & L. Zheng (2012): "The Road Less Traveled: Strategy Distinctiveness and Hedge Fund Performance." *The Review of Financial Studies* **25(1)**: pp. 96–143.

Teo, M. (2009): "The Geography of Hedge Funds." *The Review of Financial Studies* **22(9)**: pp. 3531–3561.

Vrontos, S. D., I. D. Vrontos, & D. Giamouridis (2008): "Hedge fund pricing and model uncertainty." *Journal of Banking & Finance* **32(5)**: pp. 741–753.

Zakamouline, V. & S. Koekebakker (2009): "Portfolio performance evaluation with generalized sharpe ratios: Beyond the mean and variance." *Journal of Banking & Finance* **33(7)**: pp. 1242–1254.

Zigraiova, D., T. Havranek, Z. Irsova, & J. Novak (2021): "How puzzling is the forward premium puzzle? A meta-analysis." *European Economic Review* **134**: p. 103714.

Appendix

Table A1: Specification Test of Andrews & Kasy (2019)

	All estimates	Bias treated	Bias not treated	IV estimates	non-IV estimates
Correlation	0.330	0.359	0.318	0.389	0.322
	[0.264, 0.392]	[0.275, 0.430]	[0.191, 0.431]	[-0.031, 0.769]	[0.243, 0.379]
Observations	1,019	605	414	46	973
	Strategy: equity hedge	Strategy: even driven	Strategy: relative value	Strategy: multi	
Correlation	0.359	0.499	-0.115	0.275	
	[0.207, 0.481]	[0.024, 0.738]	[-0.401, 0.249]	[-0.189, 0.630]	
Observations	229	113	94	40	
	1-factor model	3-factor model	4-factor model	7-factor model	Model uncertainty
Correlation	0.096	0.355	0.444	0.287	0.228
	[-0.091, 0.255]	[0.126, 0.538]	[0.178, 0.614]	[0.166, 0.407]	[0.042, 0.358]
Observations	167	71	205	298	142

Notes: The table shows the inverse publication-probability-weighted correlations between $log(\alpha)$ and $log(SE(\alpha))$, tests developed by Kranz & Putz (2022) for viability of Andrews & Kasy (2019) publication bias test. If all the assumptions of the selection model hold, the correlation should be zero. Bootstrapped standard errors in parentheses.