Recent trend of Size Premium: A Meta Analysis

Introduction

Banz (1981) first documented the presence of size effect in the US stock returns. Banz reported that the stocks in the quantiles comprised of smallest market capitals earned higher risk-adjusted return than the firms with high market capitalisation. The market value of equity (the number of outstanding shares times the share price) is a common proxy for size; academics typically log-transform this metric because it is right-skewed and bounded by zero. Multiple studies use a cross-sectional regression of (excess) individual or portfolio-sorted stock returns on a size to calculate the size effect.

Many authors delved into the sources of such proxy risk, but the results were mostly not conclusive (Crain, 2011). Some concluded that size premium is not linear across all firms. The effect is more visible in micro-cap stocks, but not a general phenomenon across all small-cap stocks. Horowitz et al. (2000a) found that the size effect is concentrated only in the smallest firms with market capitalisation of below $5 million. The size effect is five times larger in the 20th percentile of NYSE-listed firms and only marginal for larger firms (Fama & French, 2008). In UK, Michou et al. (20210), showed that smaller firms had higher returns only in the three smallest deciles but not in the larger size groups.

Amihud & Mendelson (1986) argued that small firms have higher returns because they are less liquid, and the size effect is linked to the liquidity risk. Liu (2006) demonstrated that the higher returns of smaller firms are primarily driven by their lower liquidity rather than the size effect itself. When explaining returns, Pastor & Stambaugh (2003) found that market-wide liquidity is more of a crucial factor than size. Hence, the size effect may not be related to risk, but rather an outcome of trading friction and tend to disappear when market liquidity improves.

Numerous researchers have observed that the size effect exhibits seasonality, with its impact being most pronounced in January. Keim (1983), Reinganum (1983) and Roll (1983) they all found size effect appears predominantly in January. Horowitz et al. (2000a) showed that in non-January months, small-cap stocks do not surpass larger stocks. Easterday et al. (2009) further confirmed that the size effect exists in January but disappears for the rest of the year. The January phenomenon raises concerns about the validity of size effect as a systematic risk factors.

In the recent years, many authors concluded that size premium has significantly declined, some even believe it has disappeared since the 1980s in various markets. Fama and French (2011) could not demonstrate size premium in four global regions over 20- years horizon starting in 1990. It was observed by Dimson et al. (2011) that the higher return for the smaller firms is patchy and non-existent for extended periods. Horowitz et al (2000a) concluded that from 1963 to 1981 smaller firms outperformed larger firms but from 1982 to 1997 larger firms outperformed smaller firms. Barry et al. (2002) found no evidence of a size premium in 35 emerging markets. If the size premium no longer exists or varies significantly over time, it may not be a fundamental risk factor. The concern is that firm size may no longer be correlated with an underlying risk factor, making it a poor explanatory variable for stock returns.

It is suggested that the diminishing of size premium occurred when the investors bid up small-cap stock prices due to their widely availability in 1980s, thus reducing the future returns (Horowitz et al., 2000a). Dimson & Marsh (1999) found that after investors became aware of the size effect, they exploited it, leading to its decline. Schwert (2002) argued that once an anomaly is well-documented, it tends to disappear. One of the reasons for the size premium to vanish is the investors recognition of the effect and readjusting their strategies.

The issues outlined above raise questions about the rationality of the size effect as a proxy for risk in asset pricing models in recent times. Thus, it is of paramount importance to understand the plethora of study on size premium to address the criticism and to comprehend whether the strength of the effect is only a statistical conclusion of methodological issues and data snooping biases. It is argued that size effect is partially an artifact of empirical methods (Lo & Mackinlay, 1990 and Black, 1992).

Literature Review

The size premium is one of the most widely documented anomalies in empirical asset pricing literature. The anomaly refers to the tendency of small-capitalisation firms to earn higher average returns than large-capitalisation firms. Early evidence by Rolf W. Banz (1981) first documented the negative relationship between firm size and stock returns in the US market, challenging the traditional Capital Asset Pricing Model (CAPM), which predicts that beta should be the sole determinant of expected returns. Subsequent studies by David B. Reinganum (1983) further confirmed that small firms outperform large firms even after adjusting for market beta, suggesting that the CAPM fails to fully explain the cross-section of stock returns.

The importance of the size effect was later formalised by Eugene F. Fama and Kenneth R. French (1992, 1993), who introduced the SMB (Small Minus Big) factor within the Fama–French three-factor model. Their findings demonstrate that firm size and book-to-market equity jointly explain average stock returns more effectively than beta alone. The SMB factor became one of the most influential empirical risk factors in modern finance literature and significantly shaped subsequent multifactor asset pricing models.

Several studies argue that the size premium reflects compensation for systematic risk associated with small firms. Roger G. Ibbotson and Peng Chen (2003) suggest that firm size represents a fundamental determinant of expected returns because smaller firms are generally more financially constrained, less diversified, and more vulnerable to economic shocks. Similarly, Narasimhan Jegadeesh and Sheridan Titman (1993) argue that small firms face greater distress risk and earnings uncertainty, which justifies higher expected returns.

The literature also suggests that the size effect is not evenly distributed across all firms. Horowitz et al. (2000a) show that the premium is concentrated primarily among micro-cap stocks, while Eugene F. Fama and Kenneth R. French (2008) report that the size effect is strongest among the smallest 20% of NYSE firms. These findings imply that the anomaly weakens considerably outside extremely small firms, raising questions regarding the economic significance and persistence of the premium across broader equity universes.

Liquidity-based explanations became increasingly important in explaining the size effect. Yakow Amihud and Haim Mendelson (1986) argue that illiquid securities require higher expected returns because investors must be compensated for transaction costs, trading delays, and price-impact risk. Since smaller firms are generally less liquid, the observed size premium may partly reflect a liquidity premium rather than a distinct size effect. This argument is reinforced by Lubos Pastor and Robert F. Stambaugh (2003), who find that market-wide liquidity risk explains stock returns more effectively than firm size itself. Similarly, Weimin Liu (2006) demonstrates that once liquidity is adequately controlled for, the independent explanatory power of size becomes substantially weaker, suggesting that size may proxy for liquidity-related risks rather than capturing a unique systematic factor.

Behavioural finance literature provides an alternative explanation for the anomaly. Werner F. M. De Bondt and Richard Thaler (1985) argue that investors systematically overreact to information, causing small and neglected firms to become temporarily undervalued. Similarly, Josef Lakonishok, Andrei Shleifer, and Robert Vishny (1994) suggest that many anomalies, including the size effect, may reflect investor behavioural biases rather than rational compensation for risk. Under this perspective, the abnormal returns associated with small firms arise because investors neglect or misprice less visible companies.

International evidence on the size premium remains mixed. Christopher B. Barry et al. (2002), studying 35 emerging markets, find that small firms generally outperform large firms, although the size effect weakens considerably after removing extreme returns. Their results indicate that the premium is sensitive to outliers and methodological choices. Similarly, André Luis Leite et al. (2018) document clear evidence of size effects in emerging-market excess returns, although value and profitability effects appear less consistent.

Research on emerging and frontier markets also highlights the role of market frictions. Judith Lischewski and Svitlana Voronkova (2012) find that market, size, and value factors significantly explain returns on the Polish stock market, although liquidity itself is not strongly priced. Similarly, Geert Bekaert et al. (2007) argue that liquidity risk is particularly important in emerging markets because these markets are characterised by lower investor participation, thin trading, and higher transaction costs. In such markets, size effects are often intertwined with market segmentation and liquidity constraints.

Despite the substantial supportive evidence, the size premium has faced increasing criticism in recent decades. One major criticism concerns seasonality. Easterday et al. (2009) and Horowitz et al. (2000a) show that much of the abnormal return associated with small firms occurs during January, commonly referred to as the "January effect." This raises doubts regarding whether the size premium represents a persistent anomaly throughout the year or merely a seasonal pattern. Another criticism relates to the declining magnitude of the anomaly over time. Elroy Dimson and Paul Marsh (1999), along with G. William Schwert (2002, 2003), argue that many market anomalies weaken after becoming widely known to investors. Increased investor awareness and arbitrage activity reduce the profitability of anomaly-based trading strategies, implying that the size premium may partly reflect historical data mining rather than a stable structural phenomenon. Schwert (2003) particularly notes that anomalies often decline significantly after publication, which raises concerns regarding publication bias and selective reporting in empirical finance research.

Methodological criticisms also challenge the robustness of the size premium. Tyler Shumway (1997) and Tyler Shumwaywith Vincent A. Warther (1999) demonstrate that survivorship and delisting bias significantly inflate the apparent size effect, particularly among NASDAQ stocks. Once delisting returns are properly included, the premium weakens substantially or disappears entirely. This finding suggests that earlier evidence may have overstated the profitability of small-stock strategies due to database construction issues.

The interpretation of size as a distinct risk factor has also been questioned theoretically. Jonathan Berk (1995) argues that firm size is itself an equilibrium market outcome and therefore should not be interpreted as an independent anomaly. According to Berk, the negative relationship between size and expected returns may simply reflect underlying risk characteristics already embedded in market valuations.

Recent multifactor models further challenge the independent role of size. Eugene F. Fama and Kenneth R. French (2015) show that the traditional SMB factor weakens considerably once profitability and investment factors are included in the five-factor model. Similarly, Clifford Asness et al. (2015) argue that the size premium only becomes economically meaningful after controlling for firm quality. Their findings suggest that many small firms are low-quality firms with weak profitability and high distress risk, implying that the raw size effect may simply capture compensation for poor firm quality.

Practical implementation issues also limit the economic significance of the anomaly. Richard Roll (1981) highlights that small-cap strategies are associated with substantial transaction costs, bid–ask spreads, and market impact costs. Because many micro-cap firms are thinly traded, the abnormal returns reported in academic studies may not be fully attainable after accounting for real-world trading frictions.

Overall, the literature demonstrates that the size premium remains one of the most important yet controversial anomalies in asset pricing research. While early studies interpreted the premium as compensation for systematic risk, later research increasingly attributes the anomaly to liquidity risk, behavioural biases, market frictions, survivorship bias, or methodological limitations. International evidence, particularly from emerging markets, generally supports the existence of size-related return patterns, although the magnitude and persistence of the effect vary substantially across markets, sample periods, portfolio construction methods, and empirical methodologies. Consequently, the size premium should not be interpreted as a universally stable phenomenon, but rather as a complex and evolving anomaly shaped by economic, behavioural, and methodological factors.

Given the mixed and often conflicting evidence surrounding the size premium, the present study makes several important contributions to the existing literature. First, while prior research provides extensive evidence both supporting and criticising the size effect, there remains considerable uncertainty regarding whether the anomaly represents a genuine risk factor or merely reflects seasonal patterns, survivorship bias, or methodological inconsistencies. Existing studies also differ substantially in terms of sample periods, portfolio construction methods, index selection, market coverage, and estimation techniques, making direct comparison difficult. Second, much of the literature focuses either on developed markets or on individual emerging-market studies, with relatively limited attention given to systematically synthesising the broader empirical evidence across studies and regions. Third, despite increasing concerns regarding publication bias and post-publication anomaly decay, relatively few studies formally investigate whether the reported size premium itself is influenced by selective reporting and methodological design choices.

Therefore, this research addresses an important gap by conducting a comprehensive meta-analysis of the size premium literature, integrating evidence across different markets, time periods, and empirical methodologies. By examining the impact of moderators such as sample period, publication period, index specification, portfolio breakpoints, survivorship bias, and geographical region, this study aims to distinguish between genuine economic effects and variations driven by research design. In doing so, the study contributes not only to the asset pricing literature but also to the broader debate regarding the reliability, persistence, and economic significance of financial market anomalies.

More importantly, this study extends the existing literature by explicitly examining interaction effects between key moderators, an area that remains largely underexplored in prior size premium research. Specifically, the study introduces Hypothesis 1, which proposes that data construction choices interact to influence the magnitude of the reported size premium. In particular, the interaction between portfolio breakpoint methodology (e.g., Fama–French NYSE breakpoints versus alternative breakpoints) and index specification (whole-market versus specified indices) is investigated. This contribution is important because previous studies typically examine these methodological choices independently, despite the possibility that their combined use may systematically amplify or suppress reported size premiums. By modelling these interaction effects directly, the study provides deeper insight into how empirical design choices shape anomaly estimates.

In addition, the study introduces Hypothesis 2, which examines whether the impact of temporal trends on the size premium varies systematically across geographical regions. While previous literature documents that the size premium has weakened over time, little attention has been given to whether this decline occurs uniformly across different markets. By interacting sample end year with geographical classifications such as North America, Europe, Asia-Pacific, and emerging markets, this study investigates whether the evolution of the size effect differs across institutional and economic environments. This represents an important contribution because it helps determine whether the observed decline in the size premium reflects global market efficiency improvements or region-specific structural changes. Consequently, the interaction-based framework adopted in this study advances the literature beyond conventional single-moderator analyses and provides a more nuanced understanding of the conditional and context-dependent nature of the size premium.

Methodology

In 1981, Banz first documented the ideology of small firms outperforming large ones, even after adjusting their market risk (beta). In 1992, Fama and French found a negative relationship between firm size and return, further establishing the fact that small capital firm outperforms the large ones. This size effect, also known as the size premium, is a significant factor in explaining cross-sectional variation in stock returns.

Size is the total market value of a company’s outstanding shares, conventionally measured as market capitalisation, which is calculated by multiplying the current share price by the total number of shares issued (Fama and French, 1992). Market capitalisation, is inherently bounded at zero and right skewed. To address this, it is common to log-transform market capitalisation to reduce skewness and to obtain values of the firm size that are closer to a normal distribution, which can improve the efficiency and reliability of statistical analyses (Fama and French, 1992; Banz, 1981).

The Cross-Sectional Regression Approach developed by Fama and MacBeth (1973) is a seminal method for estimating the size effect in asset pricing, treating firm size as a direct explanatory variable for stock returns (Fama and French, 1992; Banz, 1981). This approach captures the consistent empirical finding that smaller firms tend to generate higher average returns than larger firms—a phenomenon originally documented by Banz (1981) and later incorporated into multi-factor models by Fama and French (1992). However, this method encounters several econometric challenges. In the short term, stock returns are highly volatile and may diverge from a firm's intrinsic value, resulting in large standard errors for the estimated coefficients (Merton, 1980). Moreover, the estimation of market beta—a central input—may introduce measurement error, particularly if the error correlates with the regression residuals, thus biasing the size coefficient (Blume, 1970). To mitigate these issues, the Fama-MacBeth procedure estimates cross-sectional regressions separately for each month using the following specification:

$$R_{it} = \gamma_{0t} + \gamma_{1t}\beta_i + \gamma_{2t}X_i + u_{it}$$

where R_{it} the return on stock *i* in month *t*, β_i is the estimated market beta from a preceding time-series regression (Sharpe, 1964; Lintner, 1965), X_i is a vector of additional explanatory variables (including firm size), and u_i is the error term. This two-step approach reduces the impact of short-term volatility and provides more stable long-term estimates by averaging the monthly slope coefficients over time (Fama and MacBeth, 1973). This technique has become standard in literature for its ability to control time-varying risk premia (Cochrane, 2001) and address the errors-in-variables problem (Shanken, 1992).

To enhance the signal-to-noise ratio, some researchers group firms into portfolios based on size before conducting regressions. This technique, referred to as portfolio sorting or grouping, helps reduce firm-specific noise and can lead to more stable coefficient estimates. However, it comes with limitations. Aggregating firms into portfolios reduces the number of observations and may suppress important firm-level variation, potentially concealing key economic relationships (Lo and MacKinlay, 1990; Berk, 2000). Despite these drawbacks, portfolio grouping remains widely used in empirical asset pricing as an effective method for mitigating the influence of outliers and idiosyncratic risk (Fama and French, 2008; van Dijk, 2011).

The Portfolio Sorting Approach introduced by Fama and French (1993) captures the size premium by constructing mimicking portfolios, rather than including firm size directly as a variable in cross-sectional regressions. Central to this method is the Small-Minus-Big (SMB) factor, which represents the difference in average monthly returns between a portfolio of small-cap stocks (typically comprising the bottom 30% of firms by market capitalisation) and a portfolio of large-cap stocks (usually the top 30%). The SMB factor is designed to reflect the systematic risk premium associated with smaller firms, which have been shown to outperform larger firms on average (Banz, 1981; Fama and French, 1993).

To estimate a firm's exposure to this size premium, its monthly stock returns are regressed on the time series of the SMB premiums, typically using the Fama-French three-factor model, which includes the market factor (MKT), the size factor (SMB), and the value factor (HML) (Fama and French, 1993). This regression takes the form:

$$R_{it} - R_{ft} = \alpha_i + \beta_i\left(RM_t - R_{ft}\right) + s_iSMB_t + h_iHML_t + \varepsilon_{it}$$

where $R_{it}$is the return on stock i in month t , R_{ft} is the risk-free rate, RM_t is the market return, and SMB_t is the size premium. The resulting factor loadings (slope coefficients) for the SMB term, denoted as s_i , represent each firm's sensitivity to the size premium. These factor loadings can then be used as explanatory variables in Fama-MacBeth (1973) cross-sectional regressions to test whether these size exposures explain the cross-section of stock returns (Fama and MacBeth, 1973; Fama and French, 1993). This combined approach allows researchers to assess both the magnitude and significance of the size premium while accounting for time-varying risk exposure and other systematic risk factors (Cochrane, 2001; Shanken, 1992).

Despite its widespread use, the portfolio sorting approach has several notable limitations. First, the construction of SMB portfolios depends heavily on the sorting method, which can reduce data granularity by grouping firms with diverse characteristics into broad categories, potentially obscuring meaningful firm-level differences (Lo and MacKinlay, 1990; Berk, 2000). Second, the approach assumes that the SMB factor reliably captures the underlying size-related risk premium, which may not hold true if the sorting process introduces biases or if size is not the actual source of risk (Fama and French, 2008). Third, the method can amplify the influence of extreme observations, as small-cap stocks tend to exhibit higher volatility and are more susceptible to outliers (Shumway and Warther, 1999). Finally, portfolio sorting often requires a large number of firms to form well-diversified groups, which can be a challenge in studies involving emerging markets or specialised sectors with limited firm coverage (Barry et al., 2002). Nevertheless, due to its conceptual clarity and longstanding use in the asset pricing literature, the portfolio sorting approach remains a dominant method for estimating the size premium (Fama and French, 2012).

Taking the above criticism into consideration, this paper focuses on studies where the size effect is estimated by regressing stock returns on the natural logarithm of market value of equity or other proxies for firm size (e.g., Astakov et al.). This inclusion is not restricted to Fama-MacBeth (FM) regressions but extends to any research that estimates the relationship between stock returns and firm size using regression analysis. Once this criterion is met, we collect the size premium and associated standard errors from these studies to test for selective reporting bias. If a paper does not directly report the standard error, it will still be included if it provides other variables, such as t-statistics or standard deviations, that can be used to compute the standard error.

Testing for selective reporting bias is critical in financial research, as it can significantly distort effect sizes by over-representing statistically significant findings, leading to misleading conclusions about the strength of observed effects (Ioannidis, 2005; Stanley and Doucouliagos, 2012). This bias also affects the credibility of financial models by selectively emphasizing positive findings, creating a skewed picture of financial relationships that can mislead both investors and policymakers (Harvey et al., 2016; Campbell et al., 1997). Additionally, selective reporting increases the risk of false positives by inflating the apparent robustness of a result, while ignoring null or negative findings, leading to potentially unstable investment strategies and erroneous financial forecasts (Simmons et al., 2011; McLean and Pontiff, 2016). Finally, accurate effect size estimates are critical for evidence-based policy and investment decisions, as biased estimates can result in inefficient capital allocation and policy errors, exposing markets and portfolios to unanticipated financial risks (Lo, 2005; Campbell et al., 2001).

In the absence of selective reporting, estimated effect sizes should not be systematically related to their standard errors. However, under selective reporting or publication bias, less precise studies (characterized by larger standard errors) must report larger effect sizes in order to achieve statistical significance. This induces a systematic association between the magnitude of the reported estimate and its standard error, commonly referred to as the small-study effect. Egger et al. (2001) formalize this idea by testing for funnel plot asymmetry through a regression of standardized effect sizes on study precision.

Publication bias is a widespread problem that may seriously distort attempts to estimate the effect under investigation (Thornton & Lee, 1999). An early study found that dissertations and theses were three or four times more likely to be published if they were positive than if they were negative (Can Psychol, 1964). In the study, publication bias is tested to analyse the association between the size effect and their announced standard error using a structured four-step framework.

First, a funnel asymmetry plot is constructed by plotting the reported size premiums against a measure of study precision (the inverse of the standard error). Visual inspection of funnel symmetry provides an initial indication of potential small-study effects.

Second, publication bias is formally examined using a meta-regression framework estimated via both ordinary least squares (OLS) and precision-weighted weighted least squares (WLS). Within this specification, the intercept corresponds to the Funnel Asymmetry Test (FAT), which tests whether reported effect sizes are systematically related to their standard errors. A statistically significant intercept indicates funnel asymmetry consistent with selective reporting or small-study effects. The slope coefficient in the same regression represents the Precision Effect Test (PET), which assesses whether a genuine underlying size premium remains after accounting for potential publication bias.

If PET indicates the presence of a statistically significant genuine effect, the Precision Effect Estimate with Standard Error (PEESE) procedure is subsequently applied to obtain bias-adjusted estimates. The PEESE approach involves regressing reported effect sizes on their squared standard errors, providing an improved estimate of the underlying effect when selective reporting is present.

Together, the FAT–PET–PEESE framework provides a structured and robust approach to distinguishing between publication bias and genuine effects in the size premium literature. For completeness, Egger's regression specification is also reported as a complementary formulation of the asymmetry test. Fig 1 Publication bias detection framework shows a visual representation of the step used in this paper.

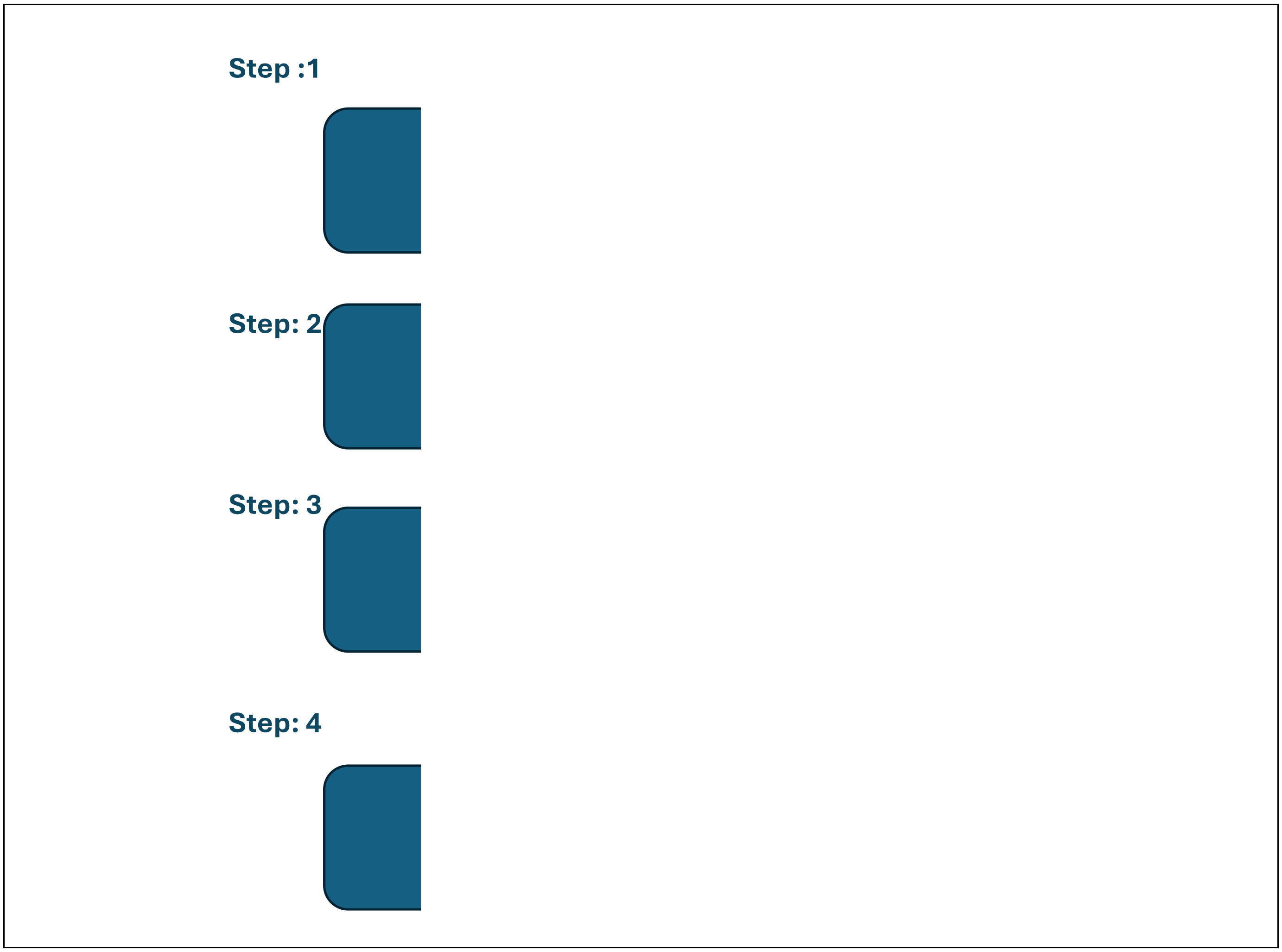


Fig: 1 Publication bias detection framework

Following Stanley (2005), publication bias is formally examined using the Funnel Asymmetry Test (FAT), implemented within a meta-regression framework. The FAT–PET regression is estimated as:

$$S_{it}=\alpha+\beta SE_{it}+\varepsilon_{it}$$

where S_{it} denotes the reported size premium from study i, SE_{it} represents the associated standard error, and ε_{it} is the disturbance term. The regression is estimated using both ordinary least squares (OLS) and precision-weighted weighted least squares (WLS), where WLS assigns greater weight to more precise estimates (i.e., those with smaller standard errors).

In this specification, the intercept α corresponds to the Funnel Asymmetry Test (FAT) and captures the extent of publication bias. Under the null hypothesis of no publication bias, the intercept should not differ significantly from zero, implying that reported size premiums are symmetrically distributed around the true underlying effect. A statistically significant intercept indicates funnel asymmetry consistent with selective reporting or small-study effects. The sign of α provides insight into the direction of bias: a positive intercept suggests an upward bias in reported size premiums, whereas a negative intercept indicates a downward bias. Estimating the regression using both OLS and precision-weighted WLS ensures robustness of inference, as WLS accounts for heteroskedasticity inherent in meta-analytic data by giving greater influence to more precisely estimated effects.

Although the FAT–PET meta-regression framework allows detection of funnel asymmetry and assessment of genuine effects, it does not by itself provide a bias-corrected estimate of the underlying effect size. When selective reporting is present, the estimated slope coefficient from PET may still be influenced by small-study effects. Therefore, following Stanley and Doucouliagos (2014), the PEESE procedure is applied conditionally to obtain a more reliable estimate of the true underlying size premium. Specifically, PEESE is preferred when PET indicates the presence of a statistically significant genuine effect, as it corrects for bias by modelling the relationship between reported effect sizes and their variances.

The PEESE regression is specified as:

Si=β0+β1SEi2+εi

where Si denotes the reported size premium, SEi2 is the squared standard error (i.e., the variance of the estimate), and εi is the disturbance term. In this specification, β0 represents the bias-adjusted estimate of the underlying true size premium, while β1 captures the relationship between the reported effect size and its variance.

Under selective reporting, imprecise estimates (with larger variances) tend to exhibit inflated magnitudes, generating a systematic association between effect size and variance. By explicitly modelling this relationship, the PEESE procedure provides an improved estimate of the true underlying effect size. Following the recommended decision rule, PEESE estimates are preferred when PET indicates the presence of a statistically significant genuine effect; otherwise, PET estimates are retained.

We performed an Egger’s test along with the FAT to address the small-study effects. Smaller studies usually lean towards to report greater and more significant findings, reasonably overstating the overall effect size in a meta-analysis (Sterne and Egger, 2005). It is often recommended as a complementary approach to FAT to provide a more robust assessment of publication bias (Sterne et al., 2011). The intercept β_0 in the equation below is considered evidence of funnel plot asymmetry and potential biases; significantly different β_0 from zero shows presence of publication bias.

$$\frac{S_i}{SE(S_i)} = \beta_0 + \beta_1 . \frac{1}{SE(S_i)} + \varepsilon_i$$

where $\beta_1$is the slope coefficient.

While the FAT–PET–PEESE framework allows us to detect and correct for potential publication bias, substantial variation in reported size premia may also arise from genuine differences in study design, sample characteristics, and methodological choices. Therefore, beyond assessing selective reporting, it is necessary to examine whether specific study-level characteristics systematically influence the magnitude of the reported size premium. To address this, we employ meta-regression techniques that incorporate moderator variables to explain heterogeneity across studies.

We begin with a single-factor (univariate) meta-regression, where each moderator variable is introduced separately into the regression model. Formally, this involves estimating the reported size premium as a function of one study-level characteristic at a time, allowing us to examine whether that moderator systematically explains variation in effect sizes across studies. This step serves as an exploratory analysis to identify potentially important drivers of heterogeneity prior to estimating a full multivariate meta-regression.

$$S_{it} = \beta_0 + \beta_1 . Z_{it} + \varepsilon_{it}$$

where S_{it} is the reported size premium from the studies and Z_{it} are the testing moderator variables from each study. The β_1 is the regression coefficient that measures the effect of the moderator on the reported size premium. The explanatory variables that are tested in this paper are:

(i) Fama French Portfolio Break Points

Fama and French (1992), for constructing the size premium (SMB) portfolios used a median split based on NYSE market capitalization. One of the research objectives is to test whether paper using the same portfolio breakpoints influences the size premium compared to paper using other portfolio breakpoints.

(ii) Market Index

In the Fama and French (1992) paper titled *"The Cross-Section of Expected Stock Returns"*, the market index they used as the market proxy was The CRSP Value-Weighted Index of all NYSE, AMEX, and NASDAQ stocks. This broad-based CRSP value-weighted index is considered to closely approximate the market portfolio in empirical asset pricing studies. For this paper a meta regression was run to understand whether using broader market index like Fama and French have elevated impact on size premium value compared to studies where they have used specific market index like Amex, Nasdaq or FTSE500, FTSE100.

(iii) Geographical Region

Both Banz (1981) and Fama and French (1992) established evidence of the size premium in the US market. However, Fama and French (1998) struggled to demonstrate a significant size premium in 12 of the 13 developed markets studied. In contrast, several authors, including Rouwenhorst (1999) and de Groot, Koedijk, and Verschoor (2012), reported a robust size premium in most emerging markets, particularly in Asia and Latin America. These contrasting findings raise concerns about the heterogeneity of the size premium across different regions. This paper tests the consequences for the size premium when the market is divided based on geographic region. The geographic regions are categorised into:

- USA: includes all the samples from USA market
- Asia Pacific: includes all the samples from any Asian market including Middle East, South Asia, and East Asia. This category also includes studies from Australia and New Zealand too.
- Europe: includes studies with samples from European countries including UK.
- Emerging: mainly includes studies where they specifically mentioned the samples are from emerging markets
- Others: includes studies from South America, Africa, Canada and any pooled groups like G7 and G8.

(iv) Survivorship Bias

Survivorship bias, excluding delisted firms, can inflate the observed performance of portfolios, especially those consisting of small-cap stocks (Brown, Goetzmann, Ibbotson, and Ross, 1992). van Dijk (2011) highlights how data selection practices, including survivorship bias, impact the replication and robustness of the size premium. Therefore, the impact on size premiums is tested in this paper.

(v) Publication Period

To investigate whether the reporting period influences the size premium, this study categorises publication years into four distinct intervals: 1981–1990, 1991–2000, 2001–2010, and 2011–present. This temporal breakdown allows for a structured analysis of how the size premium has evolved in the published literature over time. Prior research has documented that the size effect is not stable across decades. For instance, Horowitz, Loughran, and Savin (2000) reported the diminishing significance of the size premium after the early 1980s, while Dimson and Marsh (1999) observed that many anomalies, including size, tend to weaken post-publication. Similarly, van Dijk (2011) provided evidence that the magnitude and significance of the size effect vary considerably by publication period and dataset. To test this empirically, a single moderator meta-regression is conducted using publication year groupings to examine whether later studies report smaller or statistically weaker size premiums.

(vi) Sample End Period

This paper also tests whether the temporal scope of the data—as captured by the sample end year—affects the reported size premium. This analysis addresses the hypothesis that studies covering different economic regimes, market conditions, or historical phases may report varying outcomes. For instance, Fama and French (2012) show that the size premium varies globally and is sensitive to the chosen time period. Furthermore, van Dijk (2011) emphasises that both the sample period and its end date can influence the replication and robustness of asset pricing results. To examine this, additional single moderator meta-regressions are conducted on the grouped sample end years, allowing the analysis to assess how recent data coverage contributes to the heterogeneity in reported size premiums.

Univariate meta-regression is a useful exploratory tool that helps identify whether a single moderator variable systematically influences the variation in the reported size premium across studies. By focusing on one moderator at a time, it allows researchers to assess preliminary patterns in the data and detect potentially important sources of heterogeneity (Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009).

However, single-factor meta-regression also presents several methodological limitations. Including only one moderator at a time can result in biased and inconsistent estimates due to omitted variable bias, especially when other relevant variables that are correlated with the effect size are excluded (Stanley & Doucouliagos, 2012). In many meta-analytic datasets, moderators are often interrelated. A univariate model cannot disentangle these overlapping influences, which can lead to distorted estimates of moderator effects (Thompson & Higgins, 2002).

Moreover, running multiple single-factor regressions increases the risk of false positives, as each test is conducted independently without adjusting for multiple comparisons (Borenstein et al., 2009). These models also tend to overstate the effect of a given moderator by ignoring the shared variance it may have with other predictors. As a result, effects that appear strong in isolation may diminish or disappear entirely when tested within a multi-factor model (Viechtbauer, 2010). Additionally, since precision—typically measured as inverse variance weights—is a key determinant of effect size estimates, failing to control for it can lead to inaccurate moderator effects if precision is correlated with the moderator (Higgins & Thompson, 2004). Univariate models are also unable to detect interaction effects between moderators, limiting their explanatory power (Stanley & Jarrell, 1989).

For these reasons, most meta-analysis guides recommend using single-factor models primarily as an exploratory step, followed by multi-factor meta-regressions that control for confounding influences and allow for more robust and interpretable results (Stanley & Doucouliagos, 2012; Borenstein et al., 2009; Viechtbauer, 2010).

To address these limitations, the analysis progresses to multi-factor meta-regression models, in which all relevant moderators are included simultaneously.

By jointly estimating the effects of multiple moderators, multi-factor models control for confounding by allowing the unique contribution of each moderator to be estimated while holding others constant, thereby reducing biased attribution of effects (Thompson & Higgins, 2002). Multi-factor meta-regression mitigates omitted variable bias by incorporating the principal sources of heterogeneity identified in the literature and in the single-factor stage, thereby improving estimate validity (Viechtbauer, 2010). Since the influence of a moderator must persist in the presence of others to be considered robust, this approach lowers the likelihood of false positives arising from multiple comparisons (Borenstein et al., 2009). Moreover, multi-factor models allow for interaction testing, enabling exploration of whether the effect of one moderator depends on the level of another (Stanley & Jarrell, 1989).

The meta-regressions are estimated using ordinary least squares (OLS), precision-weighted weighted least squares (WLS), and restricted maximum likelihood (REML) estimators. While OLS and WLS provide conventional and precision-weighted inference, REML accounts for residual between-study heterogeneity by estimating the variance component directly, making it particularly suitable for random-effects meta-regression (Viechtbauer, 2010).

This two-stage approach — beginning with single-factor models to screen potential moderators and then advancing to multi-factor models — ensures that the final inferences about moderator effects are both statistically rigorous and robust to confounding influences.

$$S_{it} = \beta_0 + \beta_1 Z_{1,it} + \beta_2 Z_{2,it} + \cdots + \beta_k Z_{k,it} + \varepsilon_{it}$$

All the six moderators were tested in a multifactor meta regression to express their overall effect on the size premium.

$$S_{it} = \beta_0 + \beta_1 FF_{it} + \beta_2 Ind_{it} + \beta_3 Geo_{it} + \beta_4 SurvBias_{it} + \beta_5 PubPer_{it} + \beta_6 SSE_{it} + \varepsilon_{it}$$

Where:

S_{it} is the reported size premium from study *i* at time *t*,

FF_{it} is a dummy for whether the study uses Fama-French (1992) portfolio breakpoints,

Ind_{it} is the market index moderator (e.g., whole market vs. specified index),

Geo_{it} indicates the geographic region of the study (e.g., USA, UK, Emerging),

$SurvBias_{it}$ tests for the presence of survivorship bias,

$PubPer_{it}$ is the publication period group (e.g., 1981–1990, 1991–2000, etc.),

SSE_{it} represent the sample end years, respectively,

ε_{it} is the error term.

While multi-factor meta-regression is a powerful tool for controlling confounding and examining robustness, it also introduces several methodological challenges. One major issue is multicollinearity between moderators making it difficult to isolate their individual effects. This can inflate standard errors, reverse coefficient signs, and obscure meaningful relationships (Thompson & Higgins, 2002). Another concern is overfitting, particularly when the number of included moderators is large relative to the number of studies, resulting in unstable coefficient estimates and inflated values (Borenstein et al., 2009). Loss of statistical power is also common, as each additional moderator reduces degrees of freedom, making it harder to detect significant effects (Viechtbauer, 2010). Additionally, interpretation becomes more complex, as coefficients represent conditional effects which may be conceptually unrealistic if the moderators are interdependent (Higgins & Thompson, 2004). Multi-factor models are also sensitive to model specification: adding or removing a moderator can substantially alter the effects of others due to shared variance, necessitating robustness checks such as stepwise selection or leave-one-out analysis (Stanley & Doucouliagos, 2012). Lastly, there is the risk of over-controlling, particularly if moderators are intermediate variables on the causal pathway, potentially leading to underestimation of the true effect of interest (Greenland et al., 1999).

To address the limitations of both single-factor and multi-factor meta-regression, this study applies model-averaging techniques to account for model uncertainty in moderator selection. Two complementary model-averaging techniques are employed: Weighted Average Least Squares (WALS) and Bayesian Model Averaging (BMA).

BMA, a Bayesian approach, evaluates all possible combinations of moderators and computes posterior inclusion probabilities (PIPs) for each variable, representing the likelihood that a moderator belongs in the “true” model (Hoeting et al., 1999). In this study, moderators with PIP values exceeding 0.5 in the BMA analysis are designated as focus variables in the subsequent WALS estimation, ensuring that theoretically and empirically robust predictors are always included in the model.

WALS, a frequentist counterpart, combines information from all possible models, weighting the least squares estimates according to their model probabilities (Magnus, Powell, & Prüfer, 2010). In WALS, the focus variables are always included, while the remaining moderators are treated as auxiliary variables and included probabilistically. A regressor is considered robustly correlated with the outcome if its absolute t-ratio exceeds one, providing a practical selection criterion that mitigates omitted-variable bias and reduces overfitting. Comparing the WALS t-ratio criterion with BMA PIP rankings enables a cross-validation of results across frequentist and Bayesian paradigms, ensuring that the final set of moderators reflects both statistical robustness and model-selection stability.

Both methods are implemented on the meta-analysis dataset, with BMA conducted in R (BMS package) and WALS estimated in Stata using focus and auxiliary variable partitions derived from the BMA results.

Following the initial moderator screening and model uncertainty analysis using Bayesian Model Averaging (BMA) and Weighted Average Least Squares (WALS), we identify a subset of robust moderators that consistently exhibit explanatory power across specifications (Raftery et al., 1997; Magnus et al., 2010). These moderators are subsequently incorporated into a multivariate meta-regression framework to jointly assess their effects on the reported size premium. By estimating the selected moderators simultaneously, the model controls for residual heterogeneity and evaluates the unique contribution of each determinant while holding others constant.

To ensure the stability and reliability of the findings, the final multi-factor meta-regression is estimated using ordinary least squares (OLS), precision-weighted weighted least squares (WLS), and restricted maximum likelihood (REML). While OLS provides baseline inference, WLS accounts for heteroskedasticity inherent in meta-analytic data by weighting estimates according to their precision. REML further accommodates residual between-study heterogeneity by estimating the variance component directly, making it particularly suitable for random-effects meta-regression (Viechtbauer, 2010; Thompson & Higgins, 2002). Consistency of results across estimators strengthens the robustness of the inferred moderator effects.

Meta-Regression Flowchart

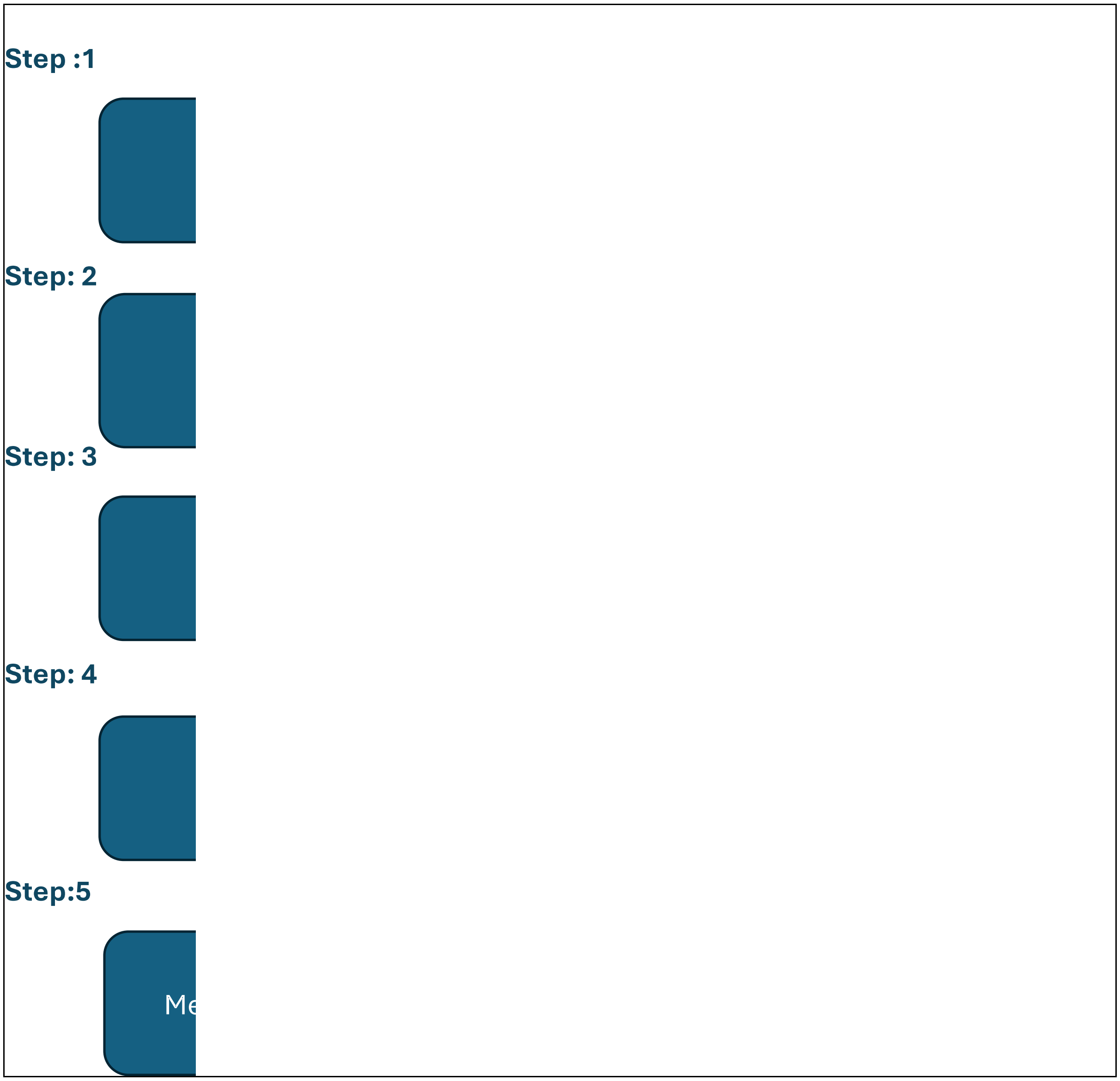


Fig:2 Meta- Regression Flowchart

Building upon the general meta-regression framework outlined above, we extend the analysis to examine specific combinations of moderators in order to test targeted hypotheses regarding the determinants of the size premium. While single- and multi-factor specifications allow for the identification of independent moderator effects, they do not capture potential interdependencies among study characteristics.

To address this limitation, we estimate additional meta-regression models incorporating interaction terms between selected moderators. This specification enables assessment of whether the effect of one study-level characteristic on the reported size premium varies systematically with the level of another. By modelling these interaction effects explicitly, we provide a more nuanced evaluation of the structural drivers of heterogeneity in the size premium literature. In particular two primary hypotheses are formulated and empirically tested:

Hypothesis 1 – Data construction choices interact to influence the magnitude of the size premium.

The portfolio breakpoint variable captures whether a study follows the Fama–French methodology in classifying firms into size-based portfolios or adopts an alternative sorting rule, directly affecting the definition of small and large firms. The index type variable reflects whether the market universe is constructed using Fama–French benchmarks (e.g., NYSE breakpoints) or alternative indices such as FTSE, MSCI, or country-specific benchmarks. While each of these design choices may independently influence the measured size premium, their effects may also be interdependent. In particular, the impact of portfolio breakpoints may differ depending on the underlying market universe employed. By introducing an interaction term between portfolio breakpoint and index type, we test whether methodological design choices jointly shape the magnitude of the reported size premium, thereby identifying whether observed heterogeneity reflects conditional data construction effects rather than isolated methodological differences.

$$S_{it} = \beta_0 + \beta_1 FF_i + \beta_2 Ind_i + \beta_3 (FF_i X Ind_i) + \varepsilon_i$$

- S_i: reported size premium (effect size) from estimate i
- FF_i: portfolio breakpoint indicator
- Ind_i: index type / market universe indicator
- $FF_i X Ind_i$: interaction term
- ε_i: error term

Hypothesis 2 – The impact of temporal trends on the size premium varies systematically across geographical regions.

To evaluate whether the size premium exhibits structural variation over time and across economic environments, we introduce temporal trends and geographical region as interacting moderators in the meta-regression framework. Temporal trends are proxied by the continuous sample end year of each study, capturing the market period to which the reported size premium corresponds. This specification allows us to assess whether the magnitude of the size effect evolves systematically over market time. Geographical region is categorised into five groups — USA (reference category), Asia-Pacific, Europe, Emerging markets, and Others — enabling comparison across distinct economic and institutional environments.

By interacting the continuous sample end year with regional indicators, we test whether the time trajectory of the size premium differs across markets. This approach allows us to determine whether any observed strengthening or attenuation of the size effect reflects a global temporal trend or whether it varies systematically by regional market structure. In doing so, the model disentangles uniform time effects from region-specific anomaly dynamics.

$$S_{it} = \beta_0 + \beta_1 SSE_i + \beta_2 Geo_1 + \beta_3 Geo_2 + \beta_4 Geo_3 + \beta_5 Geo_4 \\ + \beta_6 (SSE_i X Geo_1) + \beta_7 (SSE_i X Geo_2) + \beta_8 (SSE_i X Geo_3) + \beta_9 (SSE_i X Geo_4) + \varepsilon_i$$

Where:

- S_i: reported size premium (effect size) from estimate ii
- SSE_i: sample end year of estimate ii, capturing the temporal trend in market time (continuous variable)
- Geo_1 : dummy variable equal to 1 if estimate ii belongs to Region 1 (e.g., Asia-Pacific), 0 otherwise
- Geo_2 : dummy variable equal to 1 if estimate ii belongs to region 2 (e.g., Europe), 0 otherwise
- Geo_3: dummy variable equal to 1 if estimate ii belongs to Region 3 (e.g., Emerging markets), 0 otherwise
- Geo_4: dummy variable equal to 1 if estimate ii belongs to Region 4 (e.g., Others), 0 otherwise
- $SSE_i XGeo_{k,i}$: interaction term capturing whether the temporal trend in the size premium differs in region *k* relative to the USA
- ε_i: error term

For all meta-regression specifications described above, we estimate the models using alternative estimation techniques to ensure robustness of the findings. In particular, we implement Ordinary Least Squares (OLS), precision-weighted Weighted Least Squares (WLS), and random-effects meta-regression estimators.

OLS estimation treats each reported effect size equally, assigning identical weight to all observations regardless of their precision. This approach allows us to assess whether moderator effects remain consistent when precision-based weighting is not imposed.

In contrast, WLS accounts for heteroskedasticity inherent in meta-analytic data by weighting each estimate by the inverse of its sampling variance:

$$w_i = \frac{1}{SE_i^2}$$

so that more precise estimates exert greater influence on the regression results.
While OLS and WLS differ in their weighting schemes, both are estimated within the meta-regression framework. Comparing results across OLS, WLS, and random-effects (REML) specifications allows us to evaluate the sensitivity of the interaction effects to alternative assumptions regarding precision weighting and residual heterogeneity, thereby strengthening the robustness and credibility of the conclusions.

Data

Following the approach of Astakhov, the initial search was conducted in *Google Scholar* using the keywords:*“(size OR small) AROUND(4) (effect OR premium OR anomaly OR pattern OR puzzle) size small stock firm returns risk empirical regression portfolio sort effect premium "market value" OR "market capitalization”*.

From this search, the first 500 papers were collected for preliminary screening. The search period was restricted to publications between 1981—when Banz first documented the size effect—and 31 December 2025. To ensure methodological consistency and comparability across studies, a four-stage screening procedure was implemented.

First Screening – Regression-Based Estimation of the Size Effect
In line with the methodological considerations outlined in this thesis, only studies estimating the size effect through regression analysis were retained. Specifically, the analysis focuses on studies in which stock returns are regressed on the natural logarithm of market equity (or closely related proxies for firm size). This inclusion criterion is not limited to Fama–MacBeth (FM) regressions but extends to any empirical framework that estimates the relationship between stock returns and firm size using regression techniques. Consequently, studies relying solely on portfolio return differentials without reporting regression-based size estimates were excluded. After applying this criterion, 372 studies remained and proceeded to the second stage of screening.

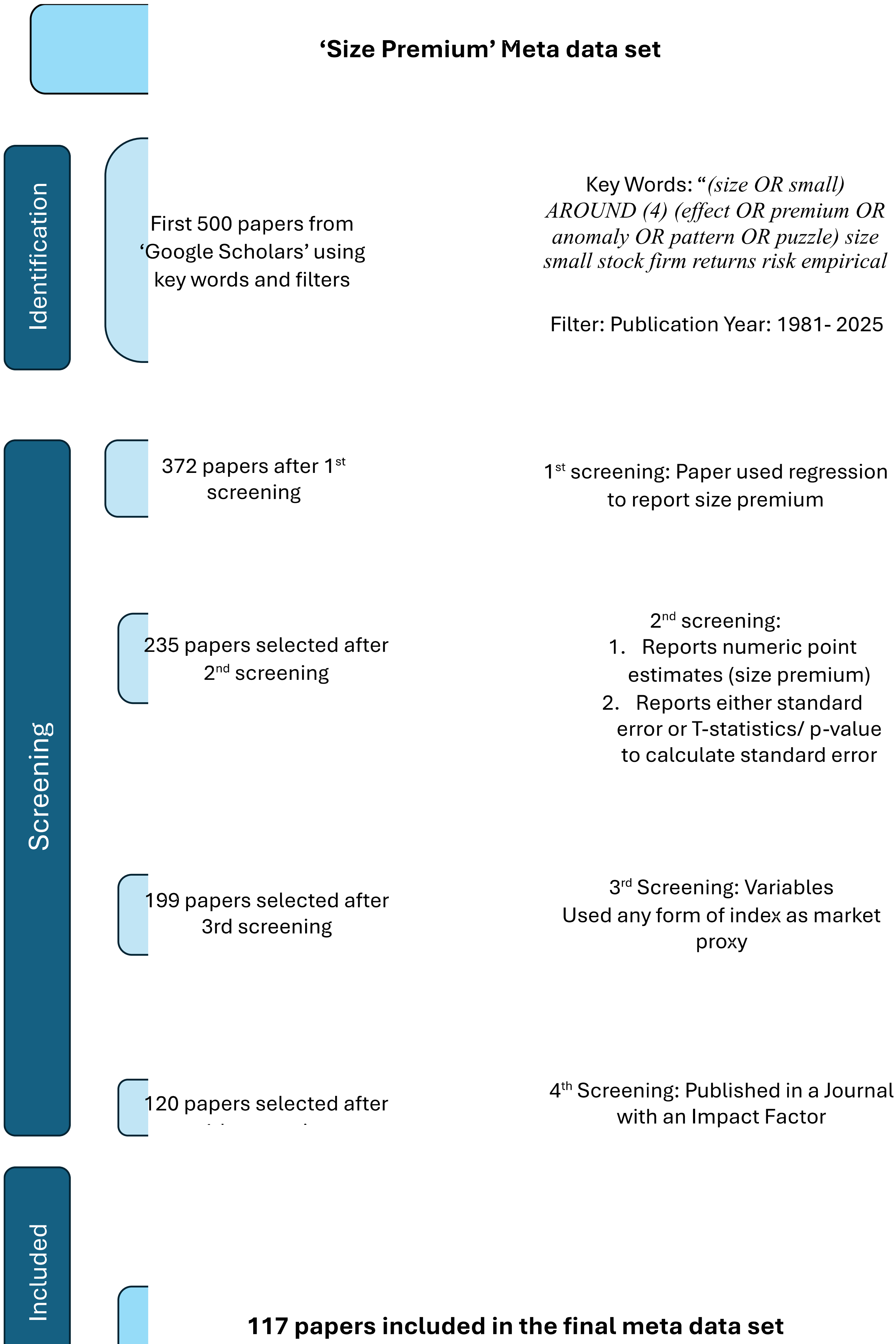
Second Screening – Availability of Statistical Precision Measures
The second criterion required that studies report the numerical size premium estimate together with a corresponding measure of statistical precision. Acceptable measures included reported

standard errors, t-statistics, or p-values from which standard errors could be computed. Studies that did not provide sufficient statistical information to recover a standard error were excluded, as they could not be incorporated into the meta-regression framework. Following this step, 137 studies were removed, leaving 235 studies eligible for the third screening stage.

Third Screening – Reporting of Market Index Proxy
Given that the market index used as a proxy for the market portfolio constitutes an important moderator in this meta-analysis, studies were required to explicitly report the index employed in estimating the size premium. Articles that did not clearly specify the market index were excluded to maintain transparency and to avoid introducing measurement ambiguity. This criterion led to the exclusion of 36 additional studies, resulting in a sample of 199 studies progressing to the final stage.

Fourth Screening – Journal Quality Requirement
Finally, to ensure academic rigor and consistency in publication standards, only studies published in peer-reviewed journals with a recognized impact factor were included in the final meta-dataset. Studies not meeting this requirement were excluded at this stage.

Applying these criteria yielded a final dataset of 117 papers, corresponding to 678 individual observations. The complete selection process is summarised in Figure 3, which presents the PRISMA flow chart.


'Size Premium' Meta data set
Identification
First 500 papers from 'Google Scholars' using key words and filters
Key Words: "(size OR small) AROUND (4) (effect OR premium OR anomaly OR pattern OR puzzle) size small stock firm returns risk empirical
Filter: Publication Year: 1981- 2025
Screening
372 papers after 1st screening
1st screening: Paper used regression to report size premium
235 papers selected after 2nd screening
2nd screening:
1. Reports numeric point estimates (size premium)
2. Reports either standard error or T-statistics/ p-value to calculate standard error
199 papers selected after 3rd screening
3rd Screening: Variables Used any form of index as market proxy
120 papers selected after
4th Screening: Published in a Journal with an Impact Factor
Included
117 papers included in the final meta data set

Results

Funnel plot

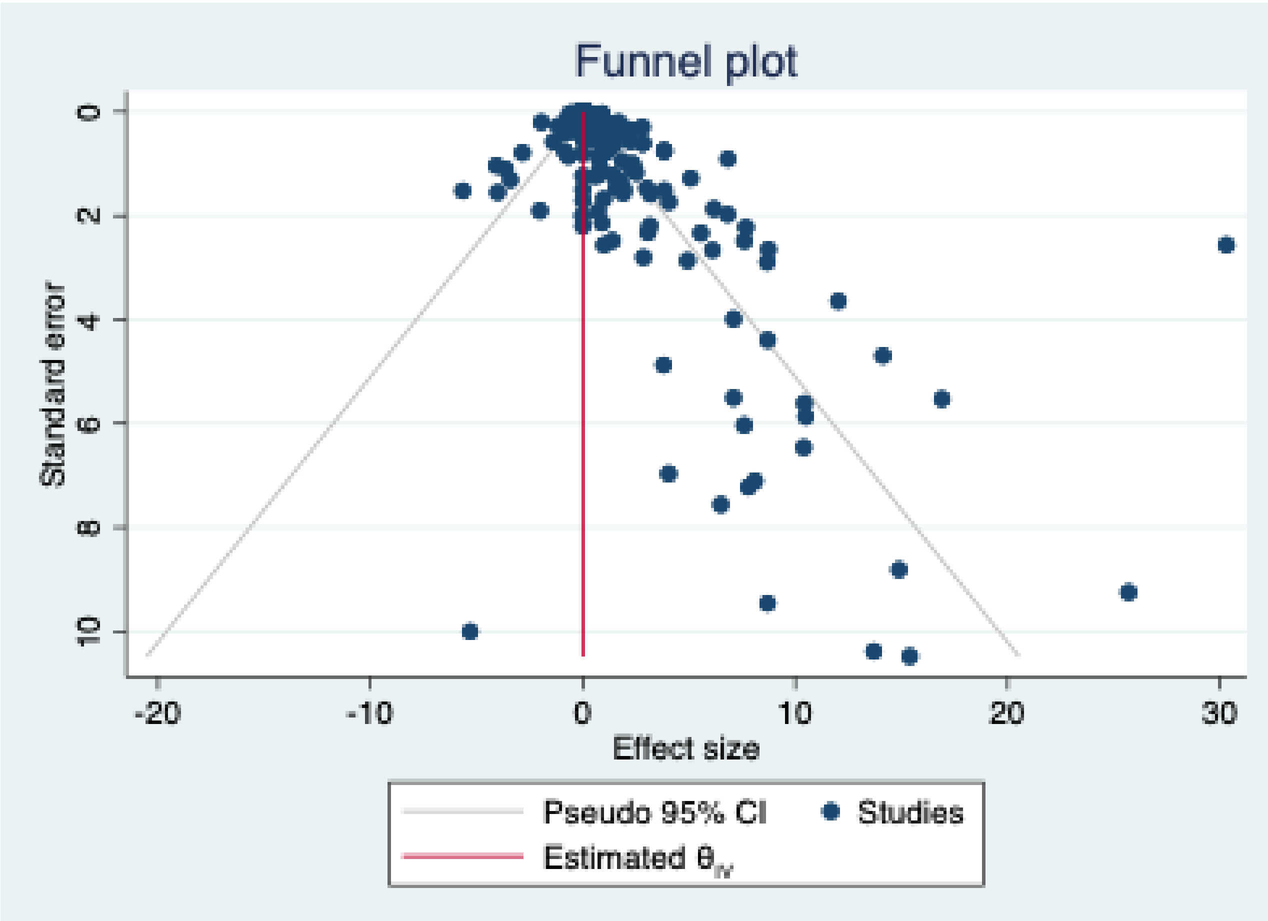


Fig 4: Funnel plot

The funnel plot in Fig 4 reveals pronounced asymmetry in the distribution of reported size premium estimates. Funnel plots are widely used in meta-analysis to detect publication bias and small-study effects (Egger et al., 1997; Sterne and Egger, 2001; Stanley, 2005). In the absence of publication bias, studies should be symmetrically distributed around the pooled inverse-variance weighted estimate (θ_{IV}), with dispersion increasing as standard errors rise, forming a balanced inverted funnel shape (Stanley (2005) and Stanley & Doucouliagos (2012). Instead, the plot shows a clear right-skewed pattern: while high-precision studies cluster around relatively modest effect sizes near the pooled mean, low-precision studies disproportionately report large positive size premiums. Conversely, imprecise studies with negative or near-zero estimates are notably underrepresented.

This systematic imbalance is consistent with strong small-study effects and provides visual evidence that the reported size premium is heavily influenced by publication bias. In relation to the central research question of this thesis—whether the documented size premium reflects a genuine risk-based anomaly or is driven by selective reporting—the funnel plot supports the latter interpretation. The overrepresentation of large positive estimates among less precise studies suggests that the magnitude and statistical significance of the size premium in the literature may be overstated due to selective publication practices.

FAT-PET-PEESE

	(1)	(2)	(3)
	FAT- OLS	FAT- WLS	PEESE
Standard Error (SE)	1.404354*** (7.19)	0.780635** (2.17)	-
Squared Standard Error (SE2)	-	-	0.1506667 *** (5.82)
Estimated True Size Premium (Intercept)	0.0233728 (0.36)	-0.0007896*** (-2.76)	0.3529972 *** (3.39)

Table 1. Funnel Asymmetry and Bias-Corrected Estimates of the Size Premium (FAT–PET–PEESE)

Standard errors are clustered at the study level (117 clusters; 672 observations). Columns (1) and (2) report FAT–PET estimates, where publication bias is tested using the standard error (SE) of each estimate. Column (3) implements the PEESE specification, which replaces SE with SE^2 to model bias as a function of sampling variance rather than sampling standard deviation. The intercept represents the estimated true size premium (PET in columns (1)– (2); PEESE in column (3)). t-statistics are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

	(1)
	EGGER's TEST
Precision (1/SE)	-0.0007896*** (-2.76)
Intercept (Funnel Asymmetry)	0.780635** (2.17)

Table 2. Egger's Regression Test for Funnel Asymmetry (Robustness Check)

Standard errors are clustered at the study level (117 clusters; 672 observations). The dependent variable is the standardized normal deviate (SND), defined as the reported size premium divided by its standard error. Precision is measured as the inverse of the standard error (1/SE). In the Egger specification, the intercept tests for funnel asymmetry and small-study effects (publication bias), while the precision coefficient captures the underlying effect in standardized form. t-statistics are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

The publication bias analysis reveals clear and economically meaningful evidence of small-study effects in the size premium literature. The FAT slope is positive and highly statistically significant across specifications. Under OLS, the coefficient on the standard error equals **1.404 ($t = 7.19$, $p < 0.01$)**, while the precision-weighted WLS specification yields a coefficient of **0.781 ($t = 2.17$, $p < 0.05$)**. These results indicate that less precise estimates (i.e., those with larger standard errors) are systematically associated with larger reported size premiums. The magnitude and significance of the FAT slope strongly suggest that sampling variance and selective reporting inflate published estimates of the size premium. The positive relationship between effect size and standard error is consistent with the pronounced funnel plot asymmetry documented earlier.

Turning to the intercept, which represents the Estimated True Size Premium under the PET specification, the results change materially once sampling error is accounted for. Under OLS, the intercept is **0.023 ($t = 0.36$)** and statistically insignificant, implying that the size premium becomes economically small and indistinguishable from zero once linear bias correction is applied. Under WLS, the intercept turns slightly negative at **−0.00079 ($t = -2.76$, $p < 0.01$)**, suggesting that after precision-weighting, the size effect effectively disappears and may even reverse in sign.

However, applying the quadratic PEESE correction alters the inference. When bias is modelled as a function of sampling variance (SE^2), the coefficient on SE^2 is positive and highly significant at **0.151 ($t = 5.82$, $p < 0.01$)**, and the bias-adjusted intercept increases to **0.353 ($t = 3.39$, $p < 0.01$)**. This estimate implies a statistically significant and economically meaningful corrected size premium under the quadratic specification. The divergence between PET and PEESE indicates that the bias-corrected magnitude of the size premium is sensitive to the assumed functional form of the variance–effect relationship.

Egger's regression test further corroborates these findings. The intercept measuring funnel asymmetry equals **0.781 ($t = 2.17$, $p < 0.05$)**, confirming statistically significant small-study effects, while the precision coefficient mirrors the PET estimate in standardized form at **−0.00079 ($t = -2.76$, $p < 0.01$)**. The consistency between FAT and Egger strengthens the conclusion that reported size premiums are systematically related to study precision.

Taken together, these findings provide strong evidence that the reported size premium is substantially influenced by small-study effects and is likely overstated in the published literature. While the anomaly does not appear to be purely an artefact of publication bias, its economic magnitude is fragile and highly dependent on the correction method employed. Under linear PET correction, the size premium largely dissipates; under quadratic PEESE correction, it re-emerges as statistically significant. This sensitivity underscores the importance of bias adjustment in asset pricing meta-analysis and calls for caution when interpreting the strength and persistence of the size effect in the broader empirical literature.

Meta Regression

i. Single factor meta regression

A single-factor meta-regression examines the relationship between the reported effect size and one moderator variable at a time, allowing researchers to identify whether a specific study characteristic systematically explains variation in reported estimates across the literature (Stanley and Jarrell, 1989; Stanley and Doucouliagos, 2012). By isolating individual moderators, this approach provides an initial assessment of heterogeneity in effect sizes and helps identify potential sources of bias or structural differences across studies before estimating more comprehensive multi-factor meta-regression models (Stanley, 2005). Table 3 lists all the moderators tested along with their respective reference groups, while Table 4 summarises the estimated effects of the single-factor moderators on the reported size premium.

Moderators	Categories	Reference group
Fama French portfolio breakpoints	Yes- the paper followed the FF breakpoints No- the papers followed something else	Yes
Index	Whole Market- the paper derived the data set from whole market Index- the paper derived the data from specific index	Whole Market
Geo	Only USA Europe – Any countries in the EU including UK Asia Pacific- Include countries from Middle East, Asia, Australia and New Zealand Emerging – Includes countries categorised as emerging in the papers Others – Include countries from Africa, South America, any pooled data set like G7 and all and Cananda	USA
Survivorship Bias	Yes – if there is any survivorship present in the paper No – if there are no survivorship bias	No
Publication Period	1981 -1991 1992-2001 2002- 2011 2012+	1981- 1991
Sample End Period	Continuous	

Table 3: List of moderators and the reference group

		(1)	(2)
Moderators	Reference Group	OLS	WLS
Fama French Portfolio Breakpoint (Yes or No)	Yes	0.0007247* (1.81)	-0.5483789*** (-2.64)
Index (Whole Market or Specific Index)	Whole Market	-0.0002918 (-0.76)	0.9983535*** (4.79)
Geo	USA		
Asia Specific		0.0006677* (1.67)	-0.208687 (-0.76)
Emerging Countries		0.0017477*** (3.32)	0.8856988*** (2.91)
Europe		0.0018401** (2.45)	0.7303482*** (2.72)
Other		-0.0012563** (-2.29)	-0.1945755 (-0.44)
Publication Period	1981-1991		
1992-2001		-0.0007188* (-1.83)	-0.8726082** (-2.09)
2001-2011		0.0012482*** (3.02)	0.1338704 (0.33)
2012+		0.0045477*** (4.03)	-0.2930048 (-0.71)
Survivorship Bias (Yes or No)	No	0.0010065** (2.38)	0.2846663 (1.40)
Sample End Year (Continuous)	Continuous	0.0001077*** (4.77)	-0.0135831 (-1.49)

Table 4: single factor meta regression table

This table reports single-factor meta-regression results for the size premium using Ordinary Least Squares (OLS) and precision-weighted Weighted Least Squares (WLS) estimators. The dependent variable is the reported size premium estimate. For categorical moderators, coefficients represent differences relative to the indicated reference group. t-statistics are reported in parentheses. Statistical significance is denoted as ***(p < 0.01), **(p < 0.05), and *(p < 0.10). OLS assigns equal weight to each observation, while WLS applies inverse-variance weights ($1/SE^2$), giving greater influence to more precisely estimated size premiums.

For the Fama–French portfolio breakpoint moderator, studies using the Fama–French methodology serve as the reference group. Under the OLS specification, studies that do not use Fama–French breakpoints report size premiums that are **0.0007247 higher** on average relative to the reference group, with the effect being statistically significant at the **10% level**. However, the WLS specification produces a markedly different result: studies not following the Fama–French breakpoint methodology report size premiums that are **0.548 lower**, and this difference is statistically significant at the **1% level**. The divergence between OLS and WLS estimates suggests that the relationship between portfolio construction methodology and reported size premiums is sensitive to precision weighting. Since WLS weights studies by the inverse of their variance

(1/SE2), giving greater influence to more precise estimates, the result indicates that the observed OLS relationship may be driven by less precise studies.

The index moderator compares studies using the whole market (reference group) with those relying on specific indices. Under the OLS specification, the coefficient is negative but statistically insignificant (−0.0002918), indicating no systematic difference in size premiums between these two approaches. In contrast, the WLS estimate is positive and highly significant (0.9983535, $p < 0.01$), suggesting that studies relying on specific indices report substantially larger size premiums once study precision is taken into account. This result indicates that index selection becomes an important determinant of reported size premiums when more precise estimates receive greater weight.

The geographical moderator evaluates differences in reported size premiums across regions relative to the USA. The OLS estimates show that studies focusing on Asia report slightly higher size premiums (0.0006677, $p < 0.10$), while studies examining emerging markets (0.0017477, $p < 0.01$) and Europe (0.0018401, $p < 0.05$) report significantly larger size premiums compared with the USA. Conversely, studies classified under "Other" regions report significantly lower estimates (−0.0012563, $p < 0.05$). The WLS specification confirms the importance of emerging markets (0.8856988, $p < 0.01$) and Europe (0.7303482, $p < 0.01$), although the Asia-specific and "Other" regional categories become statistically insignificant after precision weighting. These results suggest that geographical differences, particularly between developed and emerging markets, play an important role in explaining variation in reported size premiums.

The publication period moderator examines how reported size premiums vary across time relative to the baseline period of 1981–1991. The OLS estimates indicate that studies published between 1992–2001 report slightly smaller size premiums (−0.0007188, $p < 0.10$). In contrast, studies published between 2001–2011 (0.0012482, $p < 0.01$) and after 2012 (0.0045477, $p < 0.01$) report significantly larger size premiums relative to the baseline period. However, under the WLS specification only the 1992–2001 period remains statistically significant (−0.8726082, $p < 0.05$), while the later periods become insignificant. This suggests that the apparent increase in reported size premiums in more recent studies may partly reflect the influence of less precise estimates.

The survivorship bias moderator compares studies affected by survivorship bias with those that explicitly avoid it (reference group). The OLS estimate indicates that studies subject to survivorship bias report significantly higher size premiums (0.0010065, $p < 0.05$). However, the WLS coefficient becomes statistically insignificant (0.2846663), suggesting that the relationship weakens once estimates are weighted by their precision. This implies that the apparent effect of survivorship bias may be influenced by less precise estimates and is not robust under precision weighting.

Finally, the continuous moderator capturing the sample end year shows a positive and statistically significant relationship with the reported size premium under OLS (0.0001077, $p < 0.01$), indicating that more recent samples tend to report larger size premiums. However, this effect becomes statistically insignificant under the WLS specification (−0.0135831), suggesting that the positive trend observed in the OLS model may be driven by less precise studies. Once study precision is taken into account, the temporal increase in the size premium appears less robust.

The single-factor meta-regression results provide important insights into the central research question of this thesis: whether the magnitude of the reported size premium is systematically influenced by study characteristics rather than reflecting a stable underlying anomaly. Overall, the results suggest that several methodological and contextual moderators significantly influence the reported size premium across the literature. In particular, geographical focus, index selection, portfolio construction methodology, and survivorship bias all exhibit statistically significant effects in at least one specification, indicating that differences in research design and market context contribute to the heterogeneity observed in reported size premium estimates. The comparison between OLS and precision-weighted WLS specifications further reveals that some moderator effects weaken or reverse once study precision is taken into account, highlighting the role of small-study effects in shaping the reported magnitude of the size premium.

From a temporal perspective, the results also provide evidence of evolving patterns in the literature. While OLS estimates suggest that more recent studies tend to report larger size premiums, this relationship becomes statistically insignificant under precision weighting, implying that the apparent increase in the size premium over time may partly reflect the influence of less precise estimates rather than a genuine strengthening of the anomaly. Taken together, these findings indicate that the reported size premium is not uniform across studies but is sensitive to both methodological choices and market characteristics. This reinforces the importance of accounting for study design and precision when interpreting empirical evidence on the size effect and suggests that part of the variation in the documented anomaly may be driven by differences in research methodology rather than underlying asset pricing fundamentals.

ii. Multi factor meta regression

The multi-factor meta-regression results indicate that only a limited number of study characteristics exert robust influences on the reported size premium once multiple moderators are considered simultaneously. Under the REML specification, publication timing remains significant, with studies published after 2012 reporting larger size premiums relative to the baseline period of 1981–1991 (0.218, $p < 0.05$), suggesting a moderate upward shift in reported magnitudes in the more recent literature. The OLS estimates identify several additional significant moderators: more recent sample periods are associated with smaller reported size premiums (−0.109, $p < 0.01$), implying a negative time trend in unweighted estimates. Furthermore, studies published during 2002–2011 and after 2012 report substantially larger size premiums (1.716 and 2.581, respectively; both $p < 0.01$), indicating sizeable publication-period effects. Geographical focus also matters, as European studies report higher size premiums (0.702, $p < 0.05$), while studies subject to survivorship bias report smaller premiums (−0.528, $p < 0.05$). Methodological choices further influence reported magnitudes: studies using specified indices rather than whole-market indices report larger size premiums (0.651, $p < 0.01$), whereas studies that do not follow the Fama–French portfolio breakpoint methodology report smaller premiums (−0.424, $p < 0.10$). However, once study precision is accounted for in the inverse-variance weighted WLS specification, most of these effects lose statistical significance. Only two moderators remain robust: studies focusing on emerging markets report moderately larger size premiums (0.00182, $p < 0.05$), and studies employing non–Fama–French breakpoint methodologies report slightly higher premiums (0.00192, $p < 0.05$). The reduction in significance across estimation methods suggests that several apparent relationships are driven by less precise studies, while the precision-weighted results highlight market context and portfolio construction as the principal systematic determinants of reported size premiums.

	(1)	(2)	(3)
Variables	REML	OLS	WLS
Fama-French Portfolio Breakpoint: No	-0.016605 (-0.62)	-0.4244049* (-1.94)	0.0019155** (2.58)
Specified Index	-0.003389 (-0.12)	0.6505509*** (2.76)	-0.0004071 (-0.68)
Geo: Asia Pacific	0.0424157 (1.33)	0.1564271 (0.56)	0.0003694 (0.71)
Geo: Emerging	0.0029857 (0.07)	0.4980679 (1.54)	0.0018159** (1.98)
Geo: Europe	-0.0511905 (-1.49)	0.7022151** (2.40)	0.0013037 (1.37)
Geo: Other	-0.0386495 (-0.81)	-0.3795976 (-0.86)	0.0013167 (1.08)
Publication Period: 1992–2001	-0.0825569* (-1.65)	-0.2010946 (-0.44)	-0.0008456 (-1.11)
Publication Period: 2002–2011	0.0280334 (0.39)	1.7164*** (2.97)	-0.0003512 (-0.23)
Publication Period: 2012+	0.2176594** (2.10)	2.581208*** (3.28)	0.0016033 (0.65)
Survivorship Bias: Yes	0.0011838 (0.05)	-0.5275735** (-2.33)	-0.0000422 (-0.06)
Sample End Year	0.0008819 (0.28)	-0.1089341*** (-4.93)	0.0001178 (1.51)

Table 5 **Multi-Factor Meta-Regression Results**

Notes: Dependent variable is the reported size premium (N = 672). Model (1) reports Random-Effects Meta-Regression (REML). Model (2) reports OLS estimates. Model (3) reports inverse-variance weighted WLS estimates. t/z statistics are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Reference categories: Publication period = 1981–1991; Geography = North America; Index type = Whole Market; Fama-French = Yes; Survivorship bias = No.

The multi-factor results provide important clarification regarding both the recent evolution of the size premium and the systematic influence of study characteristics on its reported magnitude. First, the evidence does not support a robust conclusion that the size premium is either strengthening or disappearing over time. Although unweighted OLS estimates suggest that recent studies report larger size premiums, this pattern is not preserved once study precision is incorporated, indicating that apparent temporal trends are driven by less precise estimates rather than genuine changes in the anomaly. The absence of a consistent time effect in the precision-weighted specification suggests that the size premium has not experienced a uniform structural shift in recent decades. Second, the magnitude and direction of the size premium are shown to depend on methodological and contextual moderators. In particular, studies focusing on emerging markets consistently report larger size premiums, indicating that the anomaly is more pronounced in certain market environments. Portfolio construction methodology also exerts a systematic influence: deviations from the Fama–French breakpoint approach lead to statistically significant differences in reported magnitudes, implying that the size premium is sensitive to how portfolios are formed. In contrast, several moderators that appear influential in simpler models—such as publication period, index selection, and survivorship bias—do not retain significance once study precision is considered, suggesting that their effects are not robust. Overall, these findings indicate that the size premium is not a fixed empirical regularity but a conditional phenomenon whose reported magnitude varies according to study design and market context rather than exhibiting a clear long-term directional trend.

iii. BMA and WALS

Variables	PIP	Posterior Mean	Posterior SD	Prob (Coef > 0)
Sample End Year	**0.997**	**−0.098**	**0.028**	**0.000**
Publication Period 2011+	**0.900**	**2.496**	**0.999**	**0.999**
Publication Period (2001-2010)	**0.896**	**1.664**	**0.662**	**0.999**
Index Specification	**0.828**	**0.589**	**0.341**	**1.000**
Geo- Europe	0.264	0.156	0.300	1.000
Fama-French Portfolio Breakpoint	0.224	0.097	0.208	1.000
Survivorship Bias	0.181	−0.077	0.190	0.000
Geo- Emerging Markets	0.161	0.094	0.254	1.000
Publication Period (1991-2000)	0.143	−0.142	0.416	0.044
Geo- Others	0.077	−0.041	0.186	0.000
Geo- Asia Pacific	0.043	−0.005	0.062	0.141

Table 6: Bayesian Model Averaging

Notes: Results are obtained from Bayesian Model Averaging (BMA) estimation. PIP represents the posterior probability that a variable belongs to the true model. Posterior means and standard deviations are computed across all possible model specifications weighted by their posterior model probabilities. The probability that the coefficient is positive indicates the stability of the estimated effect direction across models. Variables with PIP values greater than or equal to 0.75 are highlighted in bold.

Table 6 reports the Bayesian Model Averaging (BMA) estimates for the determinants of the reported size premium. The results reveal that four moderators demonstrate strong and robust evidence of inclusion, as indicated by posterior inclusion probabilities exceeding the conventional threshold of 0.75 (bolded entries). Sample End Year exhibits an exceptionally high PIP of 0.997, implying near-certainty of inclusion in the true model. Its posterior mean of −0.098 indicates that reported size premiums decline systematically over time, suggesting a weakening of the anomaly in more recent sample periods. The probability that the coefficient is positive is effectively zero, confirming a highly stable negative relationship across model specifications. Similarly, the publication period indicators display strong evidence of importance. The dummy for **Publication Period 2011+** shows a PIP of 0.900 and a posterior mean of 2.496, indicating that studies published in the most recent decade report substantially larger size premiums relative to the base period. The probability that the coefficient is positive equals 0.999, demonstrating near-perfect stability in the positive direction of this effect. The **Publication Period (2001–2010)** indicator exhibits comparable strength, with a PIP of 0.896 and a posterior mean of 1.664, again suggesting systematically higher reported premiums in later publication periods. The high sign probability (0.999) reinforces the consistency of this positive association.

The **Index Specification** moderator also demonstrates strong inclusion evidence, with a PIP of 0.828. Its posterior mean of 0.589 indicates that differences in index construction significantly influence reported size premiums, reflecting methodological heterogeneity across studies. The probability of a positive coefficient equals one, implying complete stability in the direction of this effect across alternative model combinations.

In contrast, other moderators exhibit substantially lower posterior inclusion probabilities, indicating weaker evidence of systematic influence once model uncertainty is accounted for. For example, geographical indicators such as **Geo–Europe**(PIP = 0.264) and **Geo–Emerging Markets** (PIP = 0.161) show limited explanatory power, while methodological controls such as the **Fama-French portfolio breakpoint** (PIP = 0.224) and **Survivorship Bias** (PIP = 0.181) display only modest inclusion probabilities. These findings suggest that cross-study variation in reported size premiums is driven primarily by temporal dynamics and publication-related factors rather than by regional characteristics or certain methodological design choices.

Overall, the BMA analysis confirms that the reported magnitude of the size premium is highly sensitive to the inclusion of specific methodological moderators and highlights that temporal evolution, publication era, and index construction are the principal systematic determinants of reported size premiums in the empirical asset-pricing literature.

To ensure that the results are not driven by the choice of Bayesian methodology, the analysis is extended using Weighted-Average Least Squares (WALS) as an alternative model-averaging framework. Whereas BMA evaluates variable importance using posterior model probabilities, WALS derives model-averaged estimates within a frequentist setting. Applying both approaches enables a cross-method robustness assessment of the determinants of the size premium. Agreement between the two sets of estimates provides stronger empirical support for the stability and reliability of the identified drivers.

Variable	Coefficient	t-statistic
Sample End Year	**−0.087**	**(−4.11)**
Publication Period (1991–2000)	−0.220	(−0.51)
Publication Period (2001–2010)	**1.299**	**(2.32)**
Publication Period (2011+)	**1.937**	**(2.54)**
Geo – Asia Pacific	0.052	(0.21)
Geo – Emerging Markets	0.395	(1.46)
Geo – Europe	**0.531**	**(2.18)**
Geo – Other Regions	−0.354	(−0.93)
Index Specification	**−0.543**	**(−2.58)**
Fama–French Portfolio Breakpoint	**0.407**	**(2.16)**
Survivorship Bias	**−0.433**	**(−1.97)**

Table 7: Weighted-Average Least Squares (WALS) Estimates

Notes: This table reports Weighted-Average Least Squares (WALS) estimates using a Laplace prior. Coefficients represent model-averaged effects across alternative specifications. t-statistics are reported in parentheses. Bold values indicate coefficients with $|t| \geq 1.96$, representing statistically meaningful effects. WALS is employed as a robustness check to the Bayesian Model Averaging results.

The Weighted-Average Least Squares (WALS) estimates presented in Table 7 provide a frequentist model-averaging robustness check for the Bayesian Model Averaging results. The findings indicate that several moderators exert statistically meaningful and stable effects on the reported size premium. In particular, **Sample End Year** exhibits a strong and consistently negative association with the size premium (coefficient = −0.087, t = −4.11), suggesting that the magnitude of the size effect declines systematically over time. The publication-period indicators for **2001–2010**(coefficient = 1.299, t = 2.32) and **2011+** (coefficient = 1.937, t = 2.54) are both positive and statistically meaningful, indicating that more recent studies tend to report larger size premiums relative to earlier periods.

Methodological design choices also play an important role. **Index Specification** displays a significant negative effect (coefficient = −0.543, t = −2.58), implying that differences in index construction materially influence reported estimates. Similarly, the **Fama–French portfolio breakpoint** variable shows a positive and statistically meaningful effect (coefficient = 0.407, t = 2.16), highlighting the impact of portfolio-sorting methodology on anomaly measurement. The coefficient for **Survivorship Bias** is negative and marginally meaningful (coefficient = −0.433, t = −1.97), suggesting that datasets affected by survivorship considerations tend to report smaller size premiums.

Among regional controls, only **European markets** exhibit a statistically meaningful positive effect (coefficient = 0.531, t = 2.18), whereas other geographical indicators remain statistically weak. Overall, the WALS results reinforce the conclusion that temporal dynamics, publication characteristics, and methodological design choices are the primary systematic determinants of reported size premiums, thereby corroborating the evidence obtained from Bayesian model averaging.

iv. Multi meta using robust moderators

A comparison of the Bayesian Model Averaging (BMA) and Weighted-Average Least Squares (WALS) results reveals a consistent set of moderators that robustly explain cross-study variation in reported size premiums. Both model-averaging frameworks identify **Sample End Year** as the most influential determinant, with a consistently negative effect indicating a gradual attenuation of the size premium over time. The publication-period indicators for **2001–2010** and **2011 onwards**also exhibit strong and stable positive effects across both approaches, suggesting that more recent studies tend to report larger size premiums relative to earlier periods. Similarly, **index specification** emerges as a robust methodological determinant, confirming that differences in index construction materially influence anomaly estimates. The convergence of Bayesian inclusion probabilities and frequentist model-averaged estimates for these variables indicates that their importance is not sensitive to the choice of model-averaging methodology.

In contrast, several moderators display weaker and less consistent evidence. Variables capturing geographical variation generally exhibit low posterior inclusion probabilities in the Bayesian framework and remain statistically weak under WALS estimation, with the partial exception of European markets. Methodological controls such as the Fama–French portfolio breakpoint and survivorship bias show moderate support, appearing more influential under WALS than under BMA, but lacking the strong and consistent evidence observed for the core determinants. Taken together, these findings suggest that temporal dynamics, publication characteristics, and index construction represent the most reliable drivers of reported size premiums. Guided by this cross-method consistency, the subsequent meta-regression analysis focuses on these robust moderators to provide more precise inference on their effects.

	1	2	3
Variables	**OLS**	**WLS**	**REML**
Sample End Year	-0.100***	0.000	-0.002
	(-5.28)	(0.86)	(-0.99)
Publication Period (2001–2010)	1.846***	0.001*	0.117***
	(6.01)	(1.73)	(3.10)
Publication Period (2011+)	2.727***	0.004**	0.362***
	(5.45)	(2.43)	(5.70)
Index Specification	-0.689***	0.001	0.022
	(-3.28)	(1.59)	(0.94)
Observations	672	672	672

Table 8: Multi meta regression using robust moderators from BMA and WALS

Notes
OLS estimates treat all observations equally. WLS estimates weight observations by the inverse variance of the reported size premiums. Column (3) reports results from a random-effects meta-regression estimated using Restricted Maximum Likelihood (REML). t-statistics are reported in parentheses for the OLS and WLS models, while z-statistics are reported for the REML meta-regression. The constant represents the baseline predicted size premium when all moderators are at their reference levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8 reports the results of the multi-moderator meta-regression examining the determinants of the reported size premium across studies. The specification includes the moderators identified as robust through the Bayesian Model Averaging (BMA) posterior inclusion probabilities and the Weighted Average Least Squares (WALS) procedure. Three estimation approaches are presented to assess robustness: Ordinary Least Squares (OLS), precision-weighted Weighted Least Squares (WLS), and a random-effects meta-regression estimated using Restricted Maximum Likelihood (REML).

The results indicate that **temporal dynamics play an important role in explaining variation in the reported size premium**. Under the OLS specification, the coefficient on Sample End Year is negative and statistically significant at the 1% level (−0.100), suggesting that the magnitude of the size premium declines over time. This finding implies that studies using more recent sample periods tend to report smaller size effects. However, the WLS and REML estimates do not provide strong statistical support for this relationship, indicating that the temporal decline observed in the OLS model may be sensitive to the weighting scheme and to between-study heterogeneity. Overall, while the direction of the effect remains negative, the evidence for a systematic temporal decline in the size premium appears less robust once study precision and random effects are considered.

In contrast, **publication period effects exhibit strong and consistent explanatory power across specifications**. Relative to the reference publication period, studies published between 2001–2010 report significantly larger size premiums across all models. The effect remains positive and statistically significant under OLS, WLS, and REML, indicating that studies published during this period systematically document stronger size effects. An even larger increase is observed for studies published after 2011, where the coefficients are positive and highly significant across all three estimators. This pattern suggests that publication timing is a key determinant of the reported magnitude of the size premium.

The strong influence of publication period may reflect **structural shifts in empirical asset-pricing research or publication incentives over time**. As the asset-pricing literature expanded and replication studies became more common, later studies may have increasingly focused on documenting significant anomaly effects, leading to larger reported size premiums. Alternatively, improvements in data availability and econometric methodologies may have enabled more precise identification of cross-sectional return patterns. Regardless of the underlying mechanism, the consistent significance of publication period across estimation methods indicates that publication timing represents one of the most robust moderators of the reported size premium.

The index specification moderator captures whether studies rely on a whole-market universe or a more specific index when constructing the sample of firms. The OLS estimates indicate a negative and statistically significant coefficient, suggesting that studies using alternative index constructions report smaller size premiums relative to the reference group. However, this relationship is not statistically significant under the WLS or REML specifications. This lack of robustness implies that the influence of index construction may depend on the weighting scheme or may be driven by less precise studies, reducing confidence in its role as a systematic determinant of the size premium.

Taken together, the results suggest that **publication-period effects represent the most robust determinants of the reported size premium**, while the influence of temporal sample characteristics and index construction appears less stable across estimation approaches. These findings provide evidence that heterogeneity in the size premium across studies may partly reflect differences in research design and publication timing rather than purely underlying market dynamics.

Hypothesis 1 – Data construction choices interact to influence the magnitude of the size premium.

	1	2	3
Variables	**REML**	**WLS**	**OLS**
Non–Fama-French Breakpoints	-0.159***	0.0021***	-0.582*
	(-3.55)	(3.06)	(-1.67)
Specific Index (vs Whole Market)	-0.020	0.0015**	-1.086***
	(-0.44)	(2.07)	(-3.18)
Non-FF × Specific Index	0.075	-0.00235***	0.227
	(1.35)	(-2.62)	(0.52)

Table:9 Interaction Between Portfolio Breakpoints and Market Index Construction

Notes:
Dependent variable is the reported **size premium**. Model (1) reports random-effects meta-regression estimated using REML. Model (2) reports precision-weighted weighted least squares (WLS) estimates using inverse-variance weights. Model (3) reports unweighted OLS estimates. t-statistics (or z-statistics for REML) are reported in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Reference categories: **Fama–French breakpoints** and **Whole market sample**.

The interaction specification examines whether the effect of portfolio breakpoint methodology on the reported size premium depends on the underlying market universe used in the study. The reference category corresponds to studies employing **Fama–French portfolio breakpoints within whole-market samples**.

Across the three specifications, the coefficient on the **non–Fama-French breakpoint variable** indicates how studies using alternative portfolio sorting rules differ from those adopting the Fama–French methodology when the market universe covers the whole market. In the REML meta-regression, the coefficient is negative and statistically significant ($\beta = -0.159$, $p < 0.01$), suggesting that studies not following the Fama–French breakpoint methodology report smaller size premiums relative to those adopting the standard Fama–French portfolio construction approach. The OLS specification produces a similar directional effect, although it is only marginally significant at the 10% level. In contrast, the precision-weighted WLS model yields a small positive and statistically significant coefficient. This divergence suggests that the relationship between portfolio construction methodology and the reported size premium is sensitive to precision weighting, indicating that less precise studies may influence the unweighted estimates.

The **index construction moderator** compares studies relying on a specific index with those using the entire market universe. In the REML specification, the coefficient is small and statistically insignificant, suggesting no systematic difference in the size premium between these two sampling approaches. However, the WLS specification indicates a positive and statistically significant relationship, implying that studies based on specific indices report slightly larger size premiums once study precision is accounted for. In contrast, the OLS model produces a large negative and highly significant coefficient, suggesting that when all estimates are weighted equally, studies relying on specific indices report substantially smaller size premiums relative to whole-market samples.

The key test of the hypothesis lies in the **interaction term between portfolio breakpoint methodology and index construction**. The interaction coefficient captures whether the effect of using alternative portfolio breakpoints changes when studies rely on specific indices instead of the whole market. Across the REML and OLS specifications, the interaction term is statistically insignificant, indicating no robust evidence that the joint choice of portfolio sorting rule and index construction systematically alters the magnitude of the size premium. Although the WLS specification produces a statistically significant interaction term, the effect is small in magnitude and not consistent across specifications.

Marginal predictions were computed for each combination of portfolio breakpoint methodology and index type using the margins FamaFrench#Index_num command. The results indicate that studies using **non–Fama-French breakpoints with specific indices** report the smallest predicted size premium (0.00397), which is statistically insignificant. In contrast, **non–Fama-French breakpoints combined with whole market samples** produce a significantly larger predicted size premium (0.0592, p < 0.01). When studies follow the **Fama–French portfolio breakpoint methodology**, the predicted size premium is higher across both index constructions, with estimates of **0.1631 for specific indices** and **0.1430 for whole market samples**, both statistically significant at the 1% level. These results suggest that studies adopting the Fama–French portfolio sorting methodology tend to report larger size premiums regardless of the index construction used.

Portfolio Breakpoint	Index Type	Predicted Size Premium
Non–Fama-French	Specific Index	0.00397
Non–Fama-French	Whole Market	0.05924***
Fama–French	Specific Index	0.16308***
Fama–French	Whole Market	0.14303***

Table: 10 margins FamaFrench#Index_num

Notes: Predicted values are obtained using the margins FamaFrench#Index_num command following the interaction meta-regression model. The table reports adjusted predictions of the size premium for each combination of portfolio breakpoint methodology and index construction.

*** p < 0.01, ** p < 0.05, * p < 0.10.

Figure 5 provides a graphical representation of these interaction effects by plotting the adjusted predictions of the size premium across the four methodological combinations with 95% confidence intervals. As illustrated in the figure, the predicted size premium increases substantially when studies adopt the Fama–French portfolio breakpoint methodology, particularly when combined with specific market indices. In contrast, the combination of non–Fama-French breakpoints and specific indices produces the lowest predicted size premium.

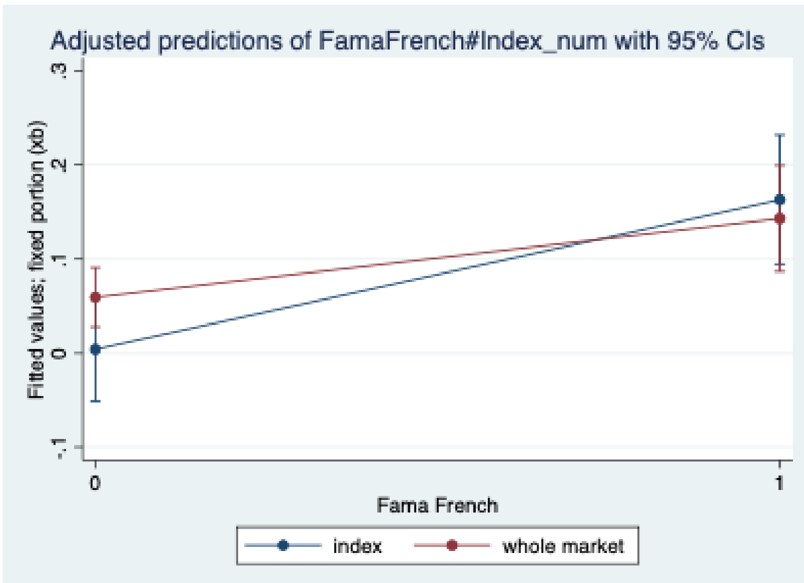


Fig 5: Graphical representation of interaction between fama French portfolio breakpoints and index

Overall, the results provide **limited support for the hypothesis that methodological design choices interact to influence the reported size premium**. While both portfolio breakpoint methodology and index construction appear to affect reported size premiums individually in some specifications, the interaction between the two variables is not robust across estimation approaches. This suggests that heterogeneity in the size premium literature is more likely driven by the independent influence of data construction choices rather than by a strong conditional relationship between them.

Hypothesis 2 – The impact of temporal trends on the size premium varies systematically across geographical regions.

	1	2	3
Variables	**REML**	**WLS**	**OLS**
Sample_End_Year	0.0078247***	0.000081***	0.011923
	(5.38)	(2.65)	(0.94)
Asia-Pacific	-23.9661***	-0.4367*	54.8105
	(-4.18)	(-1.89)	(1.15)
Emerging Markets	6.1567	0.7620	239.1252***
	(0.65)	(1.65)	(3.20)
Europe	-11.2564*	-0.0363	204.4177***
	(-1.68)	(-0.18)	(3.64)
Other Regions	2.8735	-0.3107	4.3338
	(0.35)	(-0.99)	(0.05)
Sample_End_Year × Asia-Pacific	0.0119845***	0.00022*	-0.0275
	(4.19)	(1.90)	(-1.15)

Sample_End_Year × Emerging	-0.003071	-0.00038	-0.1191***
	(-0.65)	(-1.65)	(-3.19)
Sample_End_Year × Europe	0.0055791*	0.000018	-0.1016***
	(1.67)	(0.19)	(-3.62)
Sample_End_Year × Other	-0.0014514	0.000155	-0.00228
	(-0.35)	(0.99)	(-0.06)

Table 11: Temporal Trends and Regional Variation in the Size Premium

NotesThe dependent variable is the reported **size premium**. Model (1) reports random-effects meta-regression estimated using REML. Model (2) reports precision-weighted weighted least squares (WLS) estimates using inverse-variance weights. Model (3) reports unweighted ordinary least squares (OLS) estimates. North America (USA) serves as the reference category for geographical regions. t-statistics (or z-statistics for REML) are reported in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Hypothesis 2 examines whether the temporal evolution of the size premium differs systematically across geographical regions. To test this hypothesis, the sample end year of each estimate is used as a continuous proxy for market time, while geographical location is represented by categorical regional indicators. North America (USA) serves as the reference category. Interaction terms between the sample end year and each regional dummy allow the time trend in the size premium to vary across markets.

Three estimation approaches are reported for robustness: random-effects meta-regression estimated using restricted maximum likelihood (REML), precision-weighted weighted least squares (WLS), and unweighted ordinary least squares (OLS).

The REML meta-regression results indicate a statistically significant positive temporal trend in the size premium for the reference region (North America). The coefficient on **Sample_End_Year** is positive and highly significant ($\beta = 0.0078$, $z = 5.38$, $p < 0.01$), suggesting that the magnitude of the size premium increases over time in U.S. markets.

The regional dummy variables capture differences in the average size premium relative to North America. The coefficient for **Asia-Pacific** is negative and highly significant ($\beta = -23.97$, $p < 0.01$), indicating that studies focusing on Asia-Pacific markets report lower size premiums relative to the U.S. baseline. The coefficients for **Emerging markets**, **Europe**, and **Other regions** are statistically insignificant, suggesting no consistent difference in the average size premium relative to the reference region once temporal effects are controlled for.

The interaction terms assess whether the temporal trajectory of the size premium differs across regions. The interaction between **Sample_End_Year and Asia-Pacific** is positive and statistically significant ($\beta = 0.0119$, $p < 0.01$), indicating that the size premium increases more rapidly over time in Asia-Pacific markets relative to the United States. The interaction term for **Europe** is positive and marginally significant ($\beta = 0.0056$, $p < 0.10$), suggesting a slightly stronger time trend relative to the U.S. market. In contrast, the interaction effects for **Emerging markets** and **Other regions** are statistically insignificant, implying that their temporal dynamics do not significantly differ from those observed in North America.

The precision-weighted WLS estimates yield somewhat weaker evidence of regional differences in the time evolution of the size premium. In this specification, the coefficient on **Sample_End_Year** remains positive and statistically significant ($\beta = 0.000081$, $p < 0.01$), indicating that the size premium increases gradually over time when greater weight is assigned to more precise estimates.

Among the regional indicators, the coefficient for **Asia-Pacific** is negative and marginally significant ($\beta = -0.437$, $p < 0.10$), suggesting that the average size premium reported in Asia-Pacific markets may be lower than that observed in the United States. The coefficients for **Emerging markets**, **Europe**, and **Other regions** are statistically insignificant.

The interaction effects are generally weak in the WLS specification. The interaction between **Sample_End_Year and Asia-Pacific** is positive and marginally significant ($\beta = 0.00022$, $p < 0.10$), suggesting a slightly stronger upward trend in Asia-Pacific markets. The interaction terms for **Emerging markets**, **Europe**, and **Other regions** are statistically insignificant.

The unweighted OLS estimates produce somewhat different results. In this specification, the coefficient on **Sample_End_Year** is statistically insignificant, indicating that the overall temporal trend in the size premium is not robust when all observations are treated equally.

However, the regional dummy variables indicate substantial differences in the average size premium across markets. The coefficients for **Emerging markets** and **Europe** are positive and highly significant, suggesting that studies focusing on these regions report larger average size premiums relative to the United States.

The interaction terms provide evidence that the temporal evolution of the size premium differs across regions. The interaction between **Sample_End_Year and Emerging markets** is negative and statistically significant ($\beta = -0.119$, $p < 0.01$), indicating that the size premium declines more rapidly over time in emerging markets relative to the United States. A similar negative interaction effect is observed for **Europe** ($\beta = -0.102$, $p < 0.01$), suggesting a weaker time trend compared with the U.S. market. The interaction terms for **Asia-Pacific** and **Other regions** are statistically insignificant.

Marginal Effects of Temporal Trends Across Regions

To facilitate interpretation of the interaction effects, marginal effects of **Sample_End_Year** were computed for each geographical region. These estimates represent the slope of the temporal trend in the size premium within each region.

The marginal effects indicate that the size premium increases significantly over time in several regions. The strongest temporal increase is observed in **Asia-Pacific**, where the marginal effect is positive and highly significant (dy/dx = 0.0198, $p < 0.01$). A similarly positive trend is observed in **Europe** (dy/dx = 0.0134, $p < 0.01$) and **North America** (dy/dx = 0.0078, $p < 0.01$). In contrast, the marginal effect for **Emerging markets** is statistically insignificant, indicating no clear temporal pattern in these markets.

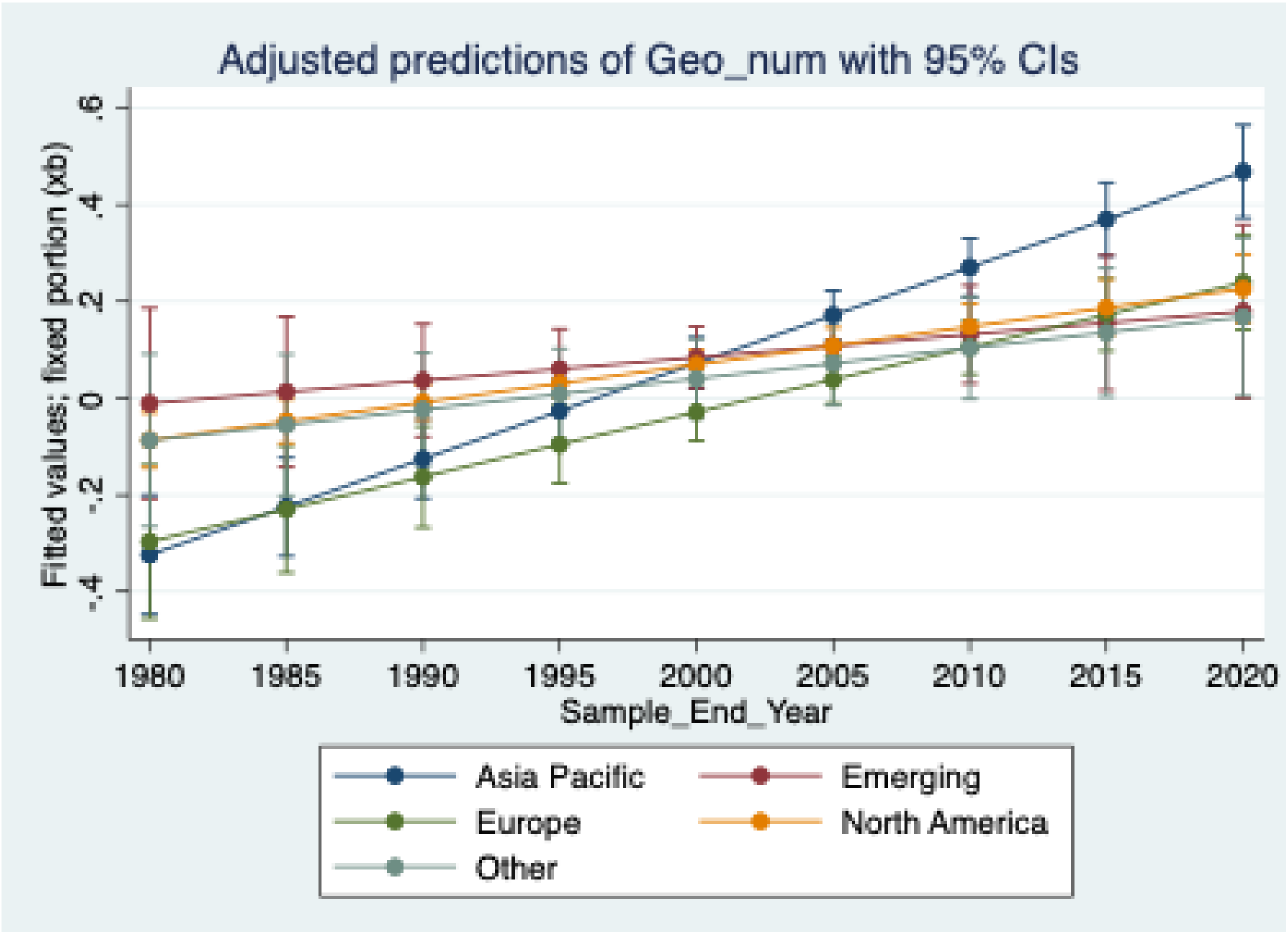
Region	Marginal Effect
Asia-Pacific	0.0198***
Emerging Markets	0.0048
Europe	0.0134***
North America (USA)	0.0078***
Other Regions	0.0064*

Notes: Marginal effects are computed from the interaction meta-regression model using margins Geo_num, dydx(Sample_End_Year). The values represent the estimated temporal trend in the size premium within each geographical region. North America serves as the reference region in the regression specification.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Graphical Illustration of Regional Time Trends

Figure 6 presents the adjusted predictions of the size premium across regions over time based on the interaction model. The figure illustrates the fitted values of the size premium as the sample end year increases, with 95% confidence intervals for each region. As shown in the figure, the size premium exhibits a generally upward trajectory across most regions, particularly in Asia-Pacific and North America, while the trend is more moderate in Europe and relatively flat in emerging markets.



Adjusted predictions of Geo_num with 95% CIs
Fitted values; fixed portion (xb)
-.4
-.2
0
.2
.4
.6
1980
1985
1990
1995
2000
2005
2010
2015
2020
Sample_End_Year
Asia Pacific
Emerging
Europe
North America
Other

Conclusion

This study provides a comprehensive meta-analysis of the size premium, examining whether its documented presence reflects a genuine risk-based phenomenon or is influenced by methodological choices and publication bias. By combining publication bias diagnostics, single and multi-moderator meta-regressions, and model-averaging techniques, the findings offer a nuanced understanding of the drivers of the size effect across the empirical literature.

The evidence from the publication bias tests provides strong and consistent support for the presence of publication bias in the size premium literature. The FAT–PET–PEESE framework indicates that the estimated true size premium is sensitive to model specification, with the intercept becoming statistically insignificant or even negative once standard error adjustments are introduced. Similarly, Egger's test confirms funnel asymmetry, suggesting that smaller or less precise studies tend to report larger size premiums. Taken together, these findings imply that the magnitude of the reported size premium is systematically inflated by selective reporting and small-study effects, rather than reflecting a stable underlying economic relationship. These results are consistent with the broader meta-analysis literature, which documents that empirical anomalies are often affected by publication selection bias (Stanley, 2005; Doucouliagos & Stanley, 2009).

The meta-regression results further demonstrate that methodological and study design choices play a critical role in shaping the reported size premium. Among the moderators examined, publication period emerges as the most robust determinant across estimation methods and model selection approaches (BMA and WALS). Later publication periods, particularly post-2011, are consistently associated with larger reported size premiums, reinforcing the evidence of publication bias and suggesting that the evolution of the academic literature itself influences reported anomaly magnitudes. This finding aligns with Schwert (2002), who argues that anomaly returns often weaken following their initial discovery but may persist in published findings due to research incentives.

In contrast, the effect of the sample period (Sample End Year) is less stable across specifications. While OLS results indicate a negative and significant time trend—suggesting that the size premium declines over time—this effect weakens or disappears under WLS and REML estimations. This inconsistency suggests that the apparent temporal decline in the size premium may be sensitive to estimation methodology and influenced by study precision and heterogeneity. Consequently, the results indicate that true market-time dynamics are less robust than publication-driven effects, further supporting the role of publication bias in explaining the variation in reported size premiums.

The findings also highlight the importance of data construction choices, particularly portfolio breakpoint methodology and index selection. The results show that studies employing Fama–French portfolio breakpoints tend to report larger size premiums compared to alternative sorting methods, indicating that the definition of "small" and "large" firms materially affects the measured anomaly. Similarly, index construction influences the size premium, although the direction and significance of this effect vary across estimators, suggesting that index choice interacts with other methodological features rather than exerting a uniformly independent effect. Geographical heterogeneity also contributes to variation in the size premium. The results indicate that emerging markets and European samples tend to exhibit stronger size premiums, while other regions show weaker or inconsistent effects. This supports the view that the size effect is not universal but depends on market structure, liquidity conditions, and institutional characteristics, consistent with prior literature (Fama & French, 1998; van Dijk, 2011).

The role of survivorship bias is also evident, although less consistently across models. In some specifications, survivorship bias reduces the reported size premium, indicating that excluding delisted firms may distort anomaly estimates. However, its relatively lower importance in the model-averaging framework suggests that survivorship bias, while relevant, is not among the most robust determinants of the size premium.

The interaction analyses provide further insight into the mechanisms underlying the size premium. For Hypothesis 1, the interaction between portfolio breakpoint methodology and index construction reveals that methodological choices do not operate independently but jointly influence the reported size premium. The marginal predictions show that the highest size premiums are observed when Fama–French breakpoints are used, regardless of index type, while non-Fama–French breakpoints combined with specific indices produce the smallest and often insignificant estimates. This indicates that the magnitude of the size premium is highly sensitive to how firms are classified and the market universe is defined, reinforcing the argument that methodological design choices are a key source of heterogeneity.

For Hypothesis 2, the interaction between temporal trends and geographical regions demonstrates that the evolution of the size premium is not uniform across markets. The marginal effects indicate that the size premium increases over time in regions such as Asia-Pacific and Europe, while remaining weaker or insignificant in emerging markets and other regions. This suggests that time trends in the size premium are region-specific rather than global, reflecting differences in market development, liquidity, and investor behaviour. Importantly, these results imply that any apparent global trend in the size premium may mask substantial regional heterogeneity.

Overall, the findings of this study suggest that the size premium is not a stable or universal asset pricing anomaly, but rather a phenomenon that is highly sensitive to publication dynamics, methodological design choices, and market context. The strong and consistent evidence of publication bias indicates that a substantial portion of the reported size premium in the literature may be driven by selective reporting and research practices rather than underlying economic fundamentals.

These results have important implications for both academic research and practical asset pricing applications. For researchers, the findings highlight the need for greater transparency, replication, and methodological consistency in anomaly studies. For practitioners, the results suggest caution in relying on the size premium as a persistent source of excess returns, particularly given its sensitivity to sample selection and model specification.

In conclusion, this study contributes to the asset pricing literature by demonstrating that the size premium is better understood as an artefact of empirical methodology and publication processes rather than a robust and persistent risk factor, thereby challenging its role in explaining cross-sectional stock returns.