



**HAL**  
open science

# Stock Market Liquidity and the Trading Costs of Asset Pricing Anomalies

Marie Brière, Tamara Nefedova, Amine Raboun

► **To cite this version:**

Marie Brière, Tamara Nefedova, Amine Raboun. Stock Market Liquidity and the Trading Costs of Asset Pricing Anomalies. 2023. hal-02283349

**HAL Id: hal-02283349**

**<https://hal.science/hal-02283349>**

Preprint submitted on 19 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stock Market Liquidity and the Trading Costs of Asset Pricing Anomalies

Marie Brière <sup>\*</sup>      Charles-Albert Lehalle <sup>†</sup>      Tamara Nefedova <sup>‡</sup>  
Amine Raboun <sup>§</sup>

January 13, 2020<sup>¶</sup>

## Abstract

Using a large database of the US institutional investors' trades, this paper sheds new light on the question of anomalies-based portfolio transaction costs. We find that the real costs paid by large investors to implement the well-identified Fama-French anomalies (size, value, investment and profitability) and Carhart momentum are significantly lower than documented in the previous studies. We show that the average investor pays an annual transaction cost of 16bps for size, 23bps for value, 31bps for investment and profitability and 222bps for momentum. The five strategies generate statistically significant net returns after accounting for transaction costs of respectively 4.29%, 1.98%, 4.45%, 2.69%, and 2.86%. When the market impact is taken into account, transaction costs reduce substantially the profitability of the well-known anomalies for large portfolios, however, these anomalies remain profitable for average size portfolios. The break-even capacities in terms of fund size are \$ 184 billion for size, \$ 38 billion for value, \$ 17 billion for profitability, \$ 14 billion for investment and \$ 410 million for momentum.

**Keywords:** Trading Costs, Market Impact, Liquidity, Anomalies-based Investments.

---

<sup>\*</sup>Amundi Asset Management, 75015 Paris, France. Université Paris-Dauphine, PSL Research University, LEDa, 75016 Paris, France. marie.brier@dauphine.psl.eu

<sup>†</sup>Capital Fund Management, 75007 Paris, France. Imperial College London, England. charles-albert.lehalle@cfm.com

<sup>‡</sup>Université Paris-Dauphine, PSL Research University, CNRS, UMR [7088], DRM, 75016 Paris, France. tamara.nefedova@dauphine.psl.eu

<sup>§</sup>Euronext Paris, 92054 Courbevoie, France. Université Paris-Dauphine, PSL Research University, LEDa, 75016 Paris, France. amine.raboun@dauphine.psl.eu

<sup>¶</sup>The authors thank Andrew Chen, Gaëlle Le Fol, Augustin Landier, Giuliano De Rossi, Paul Schneider and the participants of the 12th Financial Risks International Forum (Paris), the 2nd QFFE International Conference (Marseille), and 2018 Quant Vision Summit in Paris for their useful comments

# 1 Introduction

Fama and French (1993) Size and Value, Carhart (1997) Momentum and Fama and French (2015) Investment and Profitability factors (also known as “asset pricing anomalies”) are widely recognized as a source of rewarded risk. Nowadays, they are extensively used by investors to build new portfolios that aim to over-perform traditional market cap weighted portfolios. Broeders and Jansen (2019) show that pension funds manage factor exposures strategically on their equity portfolios. However, the success of anomalies-driven investments raises important questions regarding the liquidity and scalability of such strategies. Because they involve high turnover and potentially generate significant transaction costs, their attractiveness for portfolio construction may strongly depend on portfolio size and market impact. In this paper, we estimate the cost of trading the well-identified anomalies based on execution data of institutional investors. Using ANcerno database, composed of large investors’ trades including pension funds, mutual funds and asset managers, we find that the average investor from ANcerno pays an annual transaction cost of 16bps for size, 23bps for value, 31bps for both investment and profitability anomalies and 222bps for momentum. This is the cost of implementing the strategies for an average ticket size submitted by ANcerno institutional clients and corresponds to a total portfolio size of approximately \$ 1 million. This average estimation does not account for the potential additional cost induced by the market impact of large portfolios’ holders. Therefore, we explore how robust these asset pricing anomalies are to market impact by estimating the dependence of trading costs to the size of the trades following Kyle (1985) framework. We derive the break-even capacity for each strategy, i.e., the maximum attainable fund size before price impact eliminates profits. We estimate \$ 184 billion for size, \$ 38 billion for value, \$ 17 billion for profitability, \$ 14 billion for investment and \$ 410 million for momentum, the most frequently rebalanced strategy.

Several authors studied the limits to arbitrage of asset pricing anomalies. For instance, Novy-Marx and Velikov (2015) estimate trading costs of a large panel of anomalies including Fama-French size and value and Carhart momentum. They find that the implementation

of size, value and momentum would generate respectively 48bps, 60bps and 780bps annual transaction costs. These costs dramatically reduce market anomalies profitability. However, their trading cost measure, based on Hasbrouck (2009) Gibbs sampler estimation of the effective spread, is a proxy of the bid-ask spread derived from end-of-day data and can be quite different from real transaction costs incurred during the intraday session. In Chen and Velikov (2018), the authors use tick-by-tick databases such as NYSE Trades and Quotes (TAQ) to study the post publication trading costs of 120 stock market anomalies. They compute the average of 4 end-of-day estimates of the bid-ask spread and find an average cost per anomaly of 100bps, corresponding to an average negative net returns of -3bps. According to this study, only cost mitigation strategies are able to generate positive net returns. Their estimation of intraday trading costs is however limited to fixed cost and does not account for market impact. Korajczyk and Sadka (2004) apply various price impact models (Glosten and Harris (1988), Breen et al. (2002)) to TAQ data, to measure market impact of Jegadeesh and Titman (1993) momentum strategies. They conclude that the abnormal returns (Fama French 3-factor alpha) are reduced respectively by 204bps and 192bps for the 11/1/3<sup>1</sup> and 5/1/6 momentum strategies. Momentum could only be profitable for relatively small investors holding less than \$2 billion of AUM. Similarly, Lesmond et al. (2004) investigate the profitability of relative strength portfolios, including Jegadeesh and Titman (1993) 6/1/6 momentum. They confront the strategies to a battery of trading cost estimates such as quoted and effective spreads, and find that stocks that generate the largest momentum returns are precisely the ones with the highest transaction costs. Depending on the trading cost measure, the net alpha of the strategy is reduced by 544bps to 937bps. Using a dynamic trading model a la Garleanu and Pedersen (2013), Bonelli et al. (2019) develop a closed formula to estimate the capacity of a trading strategy with respect to its gross performance, the liquidity of the underlying securities and the dynamics of the signal

---

<sup>1</sup>Jegadeesh and Titman (1993) define momentum strategy using three parameters J/S/K, where J is the length of the period over which past returns are calculated in months, K is the holding period in months, and S is the waiting period. "The skip" is also in months

on which the strategy is based, that they apply to four well known anomalies including value, momentum and operating profitability. They find that even if the value signal is slow, the strategy has limited capacity, non-existent in the recent period between 2002-2017 because the pure Sharpe ratio is low. Momentum has low capacity as well, between \$64 and 73 million, because of the quick mean reverting signal. While operating profitability has a large capacity of \$43 billion for large-caps and \$14 billion for mid-caps, because the pure Sharpe ratio is high and the signal is slow. Contrary to the previous studies, Frazzini et al. (2012) base their estimation on the proprietary database of AQR Capital Management's executions, and find a lower transaction cost for the momentum anomaly of 354bps. They argue that TAQ database estimates are higher than what institutional investors pay in practice, for two main reasons. On one hand, the models employed are too conservative. On the other hand, TAQ database approximates the average trade, including informed traders, retail traders, liquidity demanders, and those facing high price impact costs. Patton and Weller (2019) rely on US based mutual funds returns to estimate anomalies implementation costs. Using Fama and MacBeth (1973) framework they assess the gap in factor-mimicking portfolio performance for each particular factor, and find that the difference in compensation per unit of market exposure between paper portfolios returns and mutual funds returns are respectively -0.97%, 2.09%, and 5.04% per year for size, value and momentum anomalies. Patton and Weller (2019) analysis differs from ours in a sense they are interested in estimating the cost of all potential sources of limits to arbitrage, including regulatory constraints, investability and borrowing costs, while our analysis assesses the transaction cost when the trades are possible. However, while their approach gives average estimates of implementation costs, it does not account for market impact related to fund size. Therefore, Patton and Weller (2019) result should be seen as a lower bound cost of anomalies implementation cost.

Our results on ANcerno database of executions by institutional investors generalizes Frazzini et al. (2012) findings. We compute portfolios' trading costs in two ways. The first method is a non-parametric approach. It consists in averaging, for a given stock and

rebalancing day, the costs of ANcerno reported tickets executed during the same day, in the same direction that the simulated portfolio for each anomaly, regardless of the size of the tickets. This method accounts for the precise transaction costs borne on the rebalancing days of the strategy. The second method focuses specifically on capturing the transaction costs dependence on the volume treated due to market impact. We estimate Kyle (1985) model on ANcerno tickets using as explanatory variables the ratio of the ticket size with respect to an average daily turnover, bid-ask spread and the price returns volatility. Our estimations are lower than those documented from daily or intraday data for two main reasons. First, trade level databases usually referenced in the literature, such as TAQ, do not link single market trades to their originating parent orders. The resulting price deviation is shared by all investors participating in the same trading session and are not linked to the investor originating the trade. Therefore, these databases are more suited to study the price formation process resulting from investors' orders, than to estimate the trading costs paid by a single investor. Second, when we inflict to the strategy the cost of the spread plus the margin, we indirectly assume that large investors do not mitigate transaction costs and send only aggressive market orders that consume liquidity. Yet, large investors split their orders over several hours, or even days, depending on the size of the orders and the relevance of the trading signal. They can alternate between limit orders, executed only at the limit price or better, and market orders, which demand immediate execution at the best available price. These market practices can save institutional investors a significant amount of transaction costs.

The paper is organized as follows. In the first section, we describe present the databases we use for our study. In the second section, we define the trading cost measure, explain the parametric and non-parametric estimation methods and compare the results. Finally, we discuss the profitability of five of the most recognized anomalies, Fama-French size, value, investment, and profitability anomalies and Carhart momentum after accounting for transaction costs. We also assess the break-even capacity of each strategy.

## 2 Data

We obtain institutional trading data for the period from January 1<sup>st</sup> 1999 to June 30<sup>th</sup> 2015 from ANcerno Ltd. ANcerno, formerly Abel Noser Corporation, is one of the leading consulting companies in providing Transaction Cost Analysis (TCA) in the US. It provides equity trading costs analysis for more than 500 global institutional investors, including pension funds, insurance companies and asset managers. This database was largely used by academics to investigate institutional investors market practices (see for example Anand et al. (2011), Puckett and Yan (2011) and Eisele et al. (2019)). ANcerno clients contribute to the database by sending batches of their equity trades in order to monitor their execution quality. Therefore, costs estimated on ANcerno are representative of what institutional investors effectively pay for their executions. Previous research confirmed that ANcerno database is free from any survivorship or backfill bias (see Puckett and Yan (2011)), constitute approximately 8% of the total CRSP daily dollar volume (Anand et al. (2013)), and 10% of total institutional activity (Puckett and Yan (2011)).

For each execution, ANcerno reports information on the CUSIP and ticker of the stock, the execution time at minute precision, the execution date, execution price, side (i.e., buy or sell), number of shares traded, commissions paid, whether the trade is part of a larger order, and a number of trade-level benchmarks to evaluate the quality of the execution. For a limited period of time (until 2011), ANcerno database contained clients identifiers allowing to link parent orders to institutions executing them. An institution could be either a large mutual fund, a group of funds, or a single fund subscribing to Abel Noser analytical service. Each institution could have one or several accounts. Table 1 provides descriptive statistics of ANcerno trades. In our sample, we successfully track the activity of 1078 institutions with 149 thousands accounts, responsible of 51.3 trillion dollars of transactions, and using the service of 1488 different brokerage firms. Compared to market volume reported in CRSP, ANcerno accounts for an average of 5.44% over the whole period. However, this proportion varies in time. We observe an increase from 2.28% in 1999 to 8.28% at the end of 2004, then

a steady drop from 2005 to 2011, then a revamp of volume after 2011. The sharp decrease in ANcerno volume as percentage of CRSP after 2005 may be the direct result of the US market fragmentation happening after Reg NMS <sup>2</sup> regulation, while the high percentage volume after 2011 could be explained by the increase of passive investing [REF]. On the contrary, the traded dollar volume varies between \$2060 and \$4506 billion without a visible monotonicity. Part of the volume reported in ANcerno is executed outside the traditional market venues, thus is not reported in CRSP. The traded amount reported in ANcerno is over a trillion dollars every year and is, therefore, large enough to be relevant.

[Table 1 about here.]

ANcerno has the advantage of not being restricted to a single trading venue. It covers the three main US historical venues that compose CRSP and Computstat universe, namely, New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and NASDAQ. Table 2 Panel A shows the percentage coverage in terms of number of stocks and market capitalization. By construction, ANcerno is composed of only investible assets as all reported stocks were physically held by US institutional funds. Henceforth, it is slightly biased towards large capitalization stocks. ANcerno encompasses 92% of NYSE stocks, but slightly more than one third of AMEX companies. Yet, this difference in coverage should not bias the estimation of the studied asset pricing anomalies implementation cost, as 98% of CRSP overall market capitalization is present in ANcerno and Fama-French methodology is based on market cap weighted portfolios. ANcerno comprises a fair amount of small companies as well. In Panel B of Table 2, half of the companies are smaller than \$ 226 million worth. While the average firm size is \$ 2.4 billion, ANcerno also covers a broad range of value and momentum stocks, as shown by the spectrum of book-to-equity ratios and preceding 12 months returns covered.

---

<sup>2</sup>Reg NMS (Regulation National Market System) is a financial regulation in the US that came into force in 2005 to modernize and strengthen the National Market System for equity securities. One of its major rules concerns market fragmentation.



[Table 2 about here.]

ANcerno database has several advantages over any other trades-level database such as TAQ or TRTH (Thomson Reuters Tick History), which are both abundantly referenced in the academic literature (see, for example, Glosten and Harris (1988), Hasbrouck (1991)). First, ANcerno mainly reports the activity of institutional investors, which are the most likely to implement anomalies-based portfolios, which is the focus of this paper. Second, ANcerno is not restricted to a single stock exchange. In the US, it covers all trading venues present in CRSP. Third, ANcerno links child tickets to their corresponding parent tickets, whether the execution was split on several days or executed in one swoop. While other transaction databases tend to list the entire amount of their orders placed in the market or effectively matched, without putting any link between related trades. This may result in false estimations of portfolios trading costs, polluted by short term opportunistic investors and high frequency traders (Frazzini et al. (2018)). Finally, ANcerno provides more information on the fixed costs born by institutional investors such as broker commissions and trading fees. Table 3 gives an overview of ANcerno tickets' characteristics. By a parent-ticket we mean a buy or sell order sent by an individual fund or manager on a single stock, whether the trading firm chose to split the order across brokers or days. ANcerno provides an identifier per parent ticket, with the corresponding intended volume and execution period, which allow us to track the related child tickets. We observe that both, market conditions and institutional investors' trading behavior changed through time. We note that the number of parent-tickets has increased significantly during the studied period, starting from \$1.92 million in 1999 to attain \$11.66 million in 2008. Inversely, the average parent-ticket size and average participation rate have both constantly been dropping during the period. From \$ 1.12 million to \$ 0.39 million average size, and from 6.42 % to 0.58% participation rate. The execution period has also shrunk from almost 3 days on average to about the half 1.5 days at the end of the studied period. It suggests that institutional investors changed their trading scheme from sending relatively few very large orders (i.e., parent-tickets) to their preferred brokers, each

of which should be split on several days by the broker, to more controlled algorithms, where the asset manager takes care of the daily execution scheduling, sends relatively smaller orders and asks the broker to implement them in one or two days maximum. Note that this smooth change in market practices happened at the same time as the increase in automated trading, the decrease of over the counter trades and the multiplication of the trading venues resulting from post-RegNMS market fragmentation (Laruelle and Lehalle (2018)). Large investors adjusted their trading behaviour as a response. Commissions and fixed fees increased from 11.52bps in 1999 to 18.85bps in 2003, then dropped quickly afterward to reach 3.69bps in 2015.

[Table 3 about here.]

### 3 Methodology:

To estimate anomalies portfolios trading costs, we calculate the cost of rebalancing each stock belonging to the underlying anomaly portfolio and sum them up with their respective weights. The trading cost of each stock is measured as the sum of the implementation shortfall and fixed costs, including commissions, taxes and fees.

#### 3.1 Implementation Shortfall

Implementation shortfall, as defined in Perold (1988), measures the difference between a theoretical or benchmark price (in our case, the closing price at the time the strategy's desired holdings are generated, i.e., prior day) and an actual traded price, in percent of the benchmark price. Implementation shortfall measures the total amount of slippage a strategy might experience from its theoretical returns. In essence, our cost estimates measure how much of the theoretical returns to a strategy can actually be achieved in practice.

For a parent-ticket  $m$  of size  $Q_k(m)$  and side  $s_k(m)$  ( $= 1$  for buy tickets and  $-1$  for sell tickets) executed at date  $d$  with  $Ntrades$  child tickets, the implementation shortfall is

calculated as follows:

$$IS_k(m, d) = \frac{s_k(m)}{P_k^{ref}(d)} \left( \sum_{i=1}^{Ntrades} \frac{v_{k,m}(i)}{Q_k(m)} \times P_k(i) - P_k^{ref}(d) \right) \quad (1)$$

where  $Q_k(m) = \sum_{i=1}^{Ntrades} v_{k,m}(i)$ ,  $v_{k,m}(i)$  is the volume of each child ticket  $i$  related to the parent-ticket  $m$  and  $P_k^{ref}$  is the closing price of stock  $k$  at the review date  $d - 1$  of the back-tested strategy. All  $P_k(i)$  are happening after the open of day  $d$ .

Table 4 shows the distribution of ANcerno parent tickets implementation shortfall. Interestingly, we find that the implementation shortfall can be negative. The market movements could be either favorable or detrimental to the trade. In fact, during bullish periods, buy tickets are more expensive than sell tickets and vice-versa. For instance, in 2009, when the market daily average return was 22.79bps, the average implementation shortfall of buy order was 31.41bps, almost twice the bid-ask spread (16.37bps), whereas sell tickets benefited from this buying pressure with a low 5.84bps average cost. Similarly, in 2008 during the Sub-prime crisis, sell tickets had an excessive implementation shortfall of 50.72bps. Buy tickets, on the contrary, experienced a price improvement, i.e., negative implementation shortfall, of -16.95bps, the moment the bid-ask spread amounted for 17.27bps. The direction of the ticket compared to the direction of market movements (indicating potential market pressure) is thus an important factor in transaction costs analysis. In addition, Table 4 also highlights the dependence of transaction costs to firm size. As expected, large companies are cheaper to trade than smaller ones (8.69bps against 14.35bps). Transaction costs are increasing with the stocks' bid-ask spread and volatility. In 1999, when the bid-ask spreads were the highest on our sample (52.62bps), implementation shortfall was also the highest (42.03bps). Inversely in 2007, the bid-ask spread and implementation shortfall were at their lowest levels (9.19bps and 3.34bps respectively). More volatile periods are also associated with larger bid-ask spreads and higher implementation shortfall. During above average volatility periods, market-makers aware of adverse selection, revise their limits farther from the mid-price (Sandaas et, al. (2001)).

[Table 4 about here.]

## 3.2 Portfolio implementation cost

We propose two ways to assess the trading cost of anomalies portfolios. The first approach (non parametric estimation) consists of averaging, for each stock rebalanced on a given day, the transaction costs of all trades for that stock happening that day in the same direction in the ANcerno database. The second approach estimates a model capturing the dependency of trading costs with respect to the traded volume, and applies the estimation results to the backtested anomalies (parametric estimation).

### 3.2.1 Non Parametric Estimation

Fama-French and Carhart Momentum anomalies are based on the largest possible universe of all stocks listed in the 3 major US trading venues (NYSE, AMEX and NASDAQ). The paper strategies do not take into account any trading or liquidity constraints. Our non-parametric approach implies to restrict the universe to stocks traded by institutional investors and present in ANcerno database. Table 5 Panel A shows ANcerno average coverage of the Fama-French six sub-portfolios composing the size, value, investment and profitability anomalies, while panel B gives the coverage of the four sub-portfolios composing Carhart momentum anomaly (see details about the sub-portfolios construction in Appendix B) over the back-tested period. Large capitalization stocks are only marginally impacted by the universe restriction. More than 97% of the stocks (98% of the total market capitalization) in the original Fama-French and Carhart portfolios are traded in ANcerno database. The universe restriction impacts more heavily the small cap portfolios. For example, only 37.9% of the stocks traded in Fama-French Small-High sub-portfolio are traded in ANcerno, and 40.9% for the Small-Down sub-portfolio.

[Table 5 about here.]

When the new composition of the anomalies portfolio becomes effective after the close of day  $d - 1$ , we extract the weight of each stock  $k$  that needs to be rebalanced  $\delta w_k(d)$ . We consider all ANcerno tickets submitted in day  $d$ , on stock  $k$ , executed in the direction  $s_k(m)$  as the rebalancing trade ( $s_k(m) = \text{sign}(\delta w_k(d))$ ) and average the cost of ANcerno tickets that meet these criteria to obtain an estimation of trading cost  $\hat{\xi}_k(d)$  for each stock  $k$  (equation 2). Finally, we compute portfolio trading cost as the weighted sum of its  $Nsec$  components' trading costs multiplied by the size of the portfolio,  $AUM(d - 1)$ , at the review date.

$$\hat{\xi}_k(d) = \frac{1}{\sum_m \mathbb{I}_{\delta W_k(d)s_k(m) > 0}} \sum_m \mathbb{I}_{\delta W_k(d)s_k(m) > 0} (IS_k(m, d) + \text{fixedcost}_k(m, d)) \quad (2)$$

$$Prtf Cost(d) = AUM(d) \times \sum_{k=1}^{Nsec} |\delta W_k(d)| \times \hat{\xi}_k(d) \quad (3)$$

where  $\delta W_k(d)$  is the delta weight of stock  $k$  at the review date  $d - 1$ .  $s_k(m)$  is the side of ticket  $m$  on stock  $k$ .  $IS_k(m, d)$  and  $\text{fixedcost}_k(m, d)$  are respectively the implementation shortfall and the fixed cost of ticket  $m$  on stock  $k$ .  $AUM(d - 1)$  is the portfolio size at the end of day  $d - 1$ .  $Nsec$  is the number of stocks in the portfolio.

### 3.2.2 Parametric Estimation

One of the drawbacks of the non-parametric estimation is that it only considers stocks that are traded in ANcerno database. Also, this estimation of average transaction costs based on all trades executed in the database does not account for the well-known dependence of the trading costs to the size of the trades. Kyle (1985) model, also used by Bacry et al. (2015) and Frazzini et al. (2018), postulates a squared root relationship between the implementation shortfall and the size of the trade, measured as the fraction of daily volume traded in a stock (see equation 4 below).

$$IS_k(m, d) = \alpha \times \psi_k(d) + \beta \times \sigma_k^{GK}(d) \times \sqrt{\frac{Q_k(m)}{ADV_k(d)}} + \epsilon_k(m, d) \quad (4)$$

where  $IS_k(m, d)$  is the implementation shortfall of ticket  $m$  submitted on stock  $k$  at day  $d$ .  $\psi_k(d)$  is the quoted intraday bid-ask spread of stock  $k$  averaged on the month,  $Q_k(m)$  is the ticket size,  $ADV_k(d)$  is the daily traded volume averaged on a 12 months rolling window, and  $\frac{Q_k(m)}{ADV_k(d)}$  is the participation rate,  $\sigma_k^{GK}(d)$  is the Garman Klass intraday volatility of stock  $k$  estimated on a 12 month rolling window,  $\alpha$ ,  $\beta$  are model's parameters and  $\epsilon_k(m, d)$  is the error.

To calibrate the model, we consider all tickets reported in ANcerno database in the 12 months preceding the review date  $d-1$  of participation rate  $\frac{Q_k(m)}{ADV_k(d)}$  higher or equal to 0.01%. We find that below that threshold, the ticket size has a limited effect on trading cost because of intraday volatility noise. We then form 1000 bins based on ticket's participation rate and estimate the model on the average implementation shortfall of each bin. The left side of Table 6 describes the result of the calibration made at end of each month (152 regressions). We note that the average coefficient of the bid-ask spread is 0.4. When the market impact generated by the ticket is small ( $Q_k(m)$  close to 0), arbitrageurs on average do not pay the full bid-ask spread but only 40% of it. Therefore, for ANcerno institutional investors, the bid-ask spread is a conservative estimation of trading costs. The right side of table 6 shows the result of 10 thousand bootstraps. In each iteration, we withdraw randomly 1/12 of our ANcerno sample, constitute the bins and estimate the model. The bootstrap average coefficients 5% confidence interval confirms the rolling window approach. We use the model parameters 5% confidence interval to compute the portfolios' trading costs confidence interval.

[Table 6 about here.]

In the parametric estimation, we consider the full universe of all CRSP stocks. The estimated model parameters  $\hat{\alpha}$  and  $\hat{\beta}$  are used to estimate stocks trading costs as shown in equation 5. The intraday bid-ask spreads are retrieved from TRTH when available. For missing data, usually small capitalization, we use Abdi and Ranaldo (2017) CRSP daily spread estimates. It consists of averaging the spread of the daily last available bid and last

available ask over the month. For fixed costs, we take the average of commissions and taxes per ticket computed on the month preceding the review date.

$$\hat{\xi}_k(d) = \hat{\alpha} \times \psi_k(d) + \hat{\beta} \times \sigma_k^{GK}(d) \times \sqrt{\frac{Q_k(d)}{ADV_k(d)}} + \frac{1}{N} \sum_{k,m}^N fixedcost_k(m, d) \quad (5)$$

Where  $Q_k(d)$  is the size of the rebalancing trade. It derives explicitly from the size the portfolio AUM(d-1) such as  $Q_k(d) = \frac{AUM(d-1) \times \delta W_k(d)}{P_k^{close}(d-1)}$ . Finally, we sum stock trading cost  $\hat{\xi}_k(d)$  at portfolio level as in equation 3).

In Figure 1, we regroup ANcerno tickets in 1000 bins based on participation rate  $\frac{Q}{ADV}$  and plot the average implementation shortfall of the tickets in each of the buckets (blue dots) and the non parametric estimation (black dot). ANcerno tickets show a concave relation between implementation shortfall and ticket size relative to daily traded volume. We observe a sharp increase of the costs from -4bps to 20bps when ticket size increases from 0.01% to 2% of the ADV. The slope decays afterward. For instance a ticket with 40% participation rate, has 80bps trading cost. The parametric estimation captures well the dependence to volume and confirms the relevance of Kyle (1985) model of square root dependence of trading costs to traded volume. The non-parametric estimate does not capture volume dependence but represents the average cost paid by institutional investors.

[Figure 1 about here.]

## 4 Results

### 4.1 Non parametric estimation

In this section, we discuss the profitability of Fama-French anomalies size, value, profitability and investment and Carhart momentum anomaly from the perspective of the average trading costs experienced by ANcerno institutional investors. For each anomaly, we compare

(1) the gross performance of the anomaly, when constructed on CRSP universe of all US listed stocks following Fama-French methodology, (2) the gross performance after restricting to ANcerno universe of stocks traded by institutional investors, and (3) the net performance after accounting for transaction costs. Table 7 reports the result of the backtest starting on the 30<sup>th</sup> of June 1999 and ending the 30<sup>th</sup> of June 2015. Hou, Xue and Zhang (2017) report a significant reduction of most anomalies returns' when restricting the exposure to small and micro-capitalization stocks. Similarly, we find that the performance of size and value anomalies is reduced substantially by the restriction of the investment universe to ANcerno traded stocks, which are not well covered by ANcerno database (from 4.84% to 4.46% for size and from 2.94% to 2.22% for value). For instance, the long leg of size anomaly composed solely of small companies loses 0.42% in performance while the short leg gains 0.05%. For the same reason, both the long and short legs of the value anomaly, which both contain small capitalization stocks, experienced a substantial loss in performance (0.55% and 0.16% respectively). The momentum and investment anomalies are robust to the universe change (experience a small decay of 0.05%) while the profitability anomaly is impacted positively (4.20% on CRSP vs 4.77% on ANcerno). Israel and Moskowitz (2013) document a similar pattern for momentum, showing that it is equally strong among large and small capitalizations, while value is stronger among small caps.

Transaction costs significantly reduce the performance of the anomalies. Fama-French anomalies, rebalanced only once per year, have a low turnover and moderate trading costs, 16bps for size, 23bps for value, and 31bps for investment and profitability. Momentum strategy is much more costly to implement. It is rebalanced at the end of each month (60% monthly turnover) and has an average trading cost of 222bps. Trading costs accounts for 1/3 of the strategy gross performance. Note however that our estimates are around 0.4 times (two times and half lower than) Novy-Marx and Velikov (2015) transaction costs estimates based on daily effective bid-ask spreads ( 48, 60 and 780bps for size, value and momentum respectively). This 0.4 is also the bid-ask spread coefficient of the parametric model after



calibration. Therefore investors pay around 40% of bid-ask spread to trade small portfolios, and the full bid-ask spread is rather a conservative measure of trading costs. Moreover, Chen and Velikov (2018) argue that measures based on end-of-day data tend to over-estimate trading costs. Our estimates are more in line with those of Frazzini et al. (2012) for the momentum anomaly 351bps.

[Table 7 about here.]

## 4.2 Parametric estimation

The parametric method estimates the trading costs of all US listed stocks, based on Kyle (1985) model accounting for the dependence of trading costs to the volume traded. We back-test the anomalies on the wide universe of all CRSP stocks over the same sample period than the non-parametric estimation. Table 8, reports anomalies net performance for various levels of portfolio size: \$ 1 million, \$ 100 million, and \$ 1 billion of AUM. For comparison, BlackRock US Basic Value Fund, one of the biggest, has a total AUM of \$ 647 million. The table also reports the 5% confidence interval of portfolio trading costs derived from the 2.5% and 97.5% quantiles of model parameters bootstrapped distribution presented in Table 6. We note that the \$ 1 million portfolios generate small sized rebalancing orders that barely meets the threshold of 0.005% participation rate. Hence the orders have limited market impact. The trading costs estimates are of same magnitude than the non-parametric estimations: 19bps for size, 31bps for value, 43bps for profitability, 44bps for investment and 253bps for momentum anomaly. The net returns remain significant (4.46% for size, 2.11% for value, 4.24% for profitability, 3.78% for investment and 2.46% for momentum). The mid-sized portfolios of \$ 100 million generate rebalancing orders of the size of 1% to 2% of daily turnover. The trading costs reduce the performance of the anomalies (by 30, 54, 77, 81, and 417 bps respectively) but the net returns still remain attractive for the Fama-French anomalies (4.35%, 1.88%, 3.89% and 3.40% respectively). For large portfolios of \$ 1 billion, transaction costs are twice as big (56, 108, 156, 167, and 741 bps respectively)

and thus significantly reduce the net returns of the anomalies. While size remains largely profitable over the sample period (4.09% net return), value, profitability and investment are more heavily impacted (1.33%, 3.10%, 2.55% net return), and momentum loses all appeal (net return are significantly negative [-2.35%, -1.69%]) as the trading costs exceed the gross return of 5.15%.

[Table 8 about here.]

### 4.3 Break-even capacity

In this section, we study how robust the asset pricing anomalies are to market impact by assessing the break-even capacity of each strategy. Figure 2 document the break-even capacities of each strategy on the period ranging from the 30<sup>th</sup> June 1999 to the 30<sup>th</sup> June 2015. The continuous lines plots anomalies trading costs with respect to portfolio size, while dashed lines express anomalies returns. The intersection between dashed lines and continuous lines point the break-even capacity of each anomaly, which is the maximal fund size attainable before price impacts eliminate profits. We find that size is the most capacitive, with \$ 184 billion break-even capacity corresponding to 4.84% average cost. Followed by value with \$ 38 billion capacity and 2.94% trading costs, then investment and profitability with respectively \$14 and \$17 billion. Finally, the momentum is the most challenged by trading cost \$ 410 million and 5.15% costs. The limited capacity is partially due to bad performance of the momentum during our sample period, as it suffered from the 2001 Internet burst and 2007 financial crisis drawdown.

[Figure 2 about here.]

## 5 Conclusion

Using a broad database of investors' trades, accounting for 10% of institutional investors activity and 8% of total market volume in the US stock market, we estimate the cost of trading of five of the most recognized anomalies in the literature: Size, Value, Investment, Profitability and Momentum. We find that the average cost paid by investors to trade small orders represent less than 40% of the effective quoted bid-ask spread. This means that institutional investors are able to mitigate their transaction costs in an efficient way. Quoted bid-ask spreads, used in previous studies as a proxy for transaction costs estimates (Novy-Marx and Velikov (2015), Chen and Velikov (2018)) are rather a conservative measure of trading costs as it assumes that all institutional orders are liquidity consuming. For average sized portfolios (approximately \$ 1 million), corresponding to average volumes traded in ANcerno database, we find low trading costs for Fama-French (16bps for size, 23bps for value, and 31bps for investment and profitability), because of their low annual turnover. Momentum is more heavily impacted by trading costs (222bps, one third of the gross returns). Our estimates for the momentum anomaly trading costs are more in line with Frazzini et al. (2012) estimates. After accounting for market impact, estimating the dependence of trading costs to the traded volume using Kyle (1985) model, we compare the trading costs at three different levels of portfolio size and seek the break-even capacity of each strategy. We conclude that the momentum anomaly is the most expensive to trade. It is monthly rebalanced and involves a high turnover (around 60%). It reaches its capacity at \$410 million. The Fama-French are rebalanced once per year and thus, are more robust to trading costs. Size has an estimated capacity of \$ 184 billion, while the estimated capacity of value, investment and profitability are respectively \$38, \$14 and \$17 billion. The trading costs estimated on ANcerno database present the great advantage that they are representative of what institutional investors effectively pay. However, the exact cost of trading the long-short Fama-French portfolios should take into account the costs of short-selling (stocks borrowing costs, and leverage funding costs), an extension that we leave for future work.

## References

- Anand, A., Irvine, P., Puckett, A., and Venkataraman, K. (2011). Performance of institutional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies*, 25(2):557–598.
- Anand, A., Irvine, P., Puckett, A., and Venkataraman, K. (2013). Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis. *Journal of Financial Economics*, 108(3):773–797.
- Bacry, E., Iuga, A., Lasnier, M., and Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02):1550009.
- Bonelli, M., Landier, A., Simon, G., and Thesmar, D. (2019). The Capacity of Trading Strategies. *HEC Paris Research Paper No. FIN-2015-1089*.
- Breen, W. J., Hodrick, L. S., and Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, 48(4):470–483.
- Broeders, D. and Jansen, K. A. (2019). Do institutional investors manage factor exposures strategically? *SSRN Working Paper 3446239*.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.
- Chen, A. Y. and Velikov, M. (2018). Accounting for the anomaly zoo: A trading cost perspective. *SSRN Working Paper 3073681*.
- DeMiguel, V., Martin-Utrera, A., Nogales, F. J., and Uppal, R. (2019). A transaction-cost perspective on the multitude of firm characteristics. *SSRN Working Paper 2912819*.
- Eisele, A., Nefedova, T., Parise, G., and Peijnenburg, K. (2019). Trading out of sight: An analysis of cross-trading in mutual fund families. *Journal of Financial Economics*.

- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Frazzini, A., Israel, R., and Moskowitz, T. (2012). Trading costs of asset pricing anomalies.
- Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs.
- Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, pages 67–78.
- Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123–142.
- Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3):1445–1477.
- Hou, K., Xue, C., and Zhang, L. (2017). Replicating anomalies. Technical report, National Bureau of Economic Research.
- Israel, R. and Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108(2):275–301.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.
- Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59(3):1039–1082.

- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.
- Laruelle, S. and Lehalle, C.-a. (2018). *Market microstructure in practice*. World Scientific.
- Lesmond, D. A., Schill, M. J., and Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, 71(2):349–380.
- Novy-Marx, R. and Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.
- Patton, A. J. and Weller, B. (2019). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics (JFE)*, *Forthcoming*.
- Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3):4.
- Puckett, A. and Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, 66(2):601–633.
- Sandaas, P. (2001). Adverse selection and competitive market making: Empirical evidence from a limit order market. *The Review of Financial Studies*, 14(3):705–734.
- Zuo, Y. and Kita, E. (2012). Stock price forecast using bayesian network. *Expert Systems with Applications*, 39(8):6729–6737.

## A Data preparation

This section of the appendix provides a detailed description of databases content, and how we link them together to get the final data source. The main ANcerno dataset is a sample of institutional transaction-level data directly submitted by ANcerno’s clients. The data are submitted in batches that include all trades executed by a client during the interval of time covered by the batch. The exact length covered by each batch, however, is not predefined and can range from a few trading days to several months of trades. A large variety of clients rely on ANcerno’s monitoring services. The dataset includes transactions reported by several of the main mutual fund families domiciled in the United States, a small number of hedge funds, several pension plan sponsors and multitude of asset managers. A client from ANcerno’s perspective is any entity that submits trades, which generally consists of an individual mutual fund, a group of funds, or a fund manager. ANcerno assigns unique codes to the clients (variable `clientcode`) and the corresponding institution as reported by the client (variable `clientmgrcode`). However, the exact identity of the client is always anonymized.

For a limited period of time, ANcerno also provided a file (“MasterManagerXref”) including the list of the overarching institutions to which the individual clients were affiliated (i.e., the fund families in the case of mutual funds). This additional file includes the name of the institution (variable `manager`), e.g., PIMCO, and a number identifying the institution (variable `managercode`), e.g., 10. We match this file to another file (“ManagerXref”) (that includes both the numbers identifying the institutions (variable `managercode`) and the client codes (variable `clientcode`). In this way, we are able to match the main institution name with the original ANcerno trade data via client codes (as the variables `clientcode` and `clientmgrcode` are included both in the “ManagerXref” file and in the main ANcerno file)—see Figure3. Similarly, we link broker identifiers to trades using a fourth file obtained by ANcerno, which is called “BrokerXref”. All these information are necessary to better define a ticket and thus related trades, whether the originating house chose to split their trading on time or across different brokers. The main variables that we use from the ANcerno database

are reported in Table 9.

As far as stocks market data are concerned, we use CRSP daily and monthly stock files provided by WRDS (Wharton Research Data Services), from which we retrieve stock prices (open high low close), daily traded volume, outstanding shares, exchange code, and share code. Stocks balance sheets and fundamental information are obtained from Compustat Annual fundamental files, also provided by WRDS. We first match these two databases using CRSP-COMPUSTAT historical link table that maps Compustat GVKEY stock identifier to CRSP (PERMNO, PERMCO) couple, then we use the resulting table to cross with ANcerno main dataset using the CUSIP. Finally, we check that ANcerno reported prices fall within the range of CRSP daily low-high prices. We drop the few tickets that do not fulfill this condition.

[Figure 3 about here.]

[Table 9 about here.]

## B Fama-French Portfolio Construction

We reproduce portfolio construction described in Fama and French (2012) paper and Ken French’s webpage.

**Size, Value:** The Fama/French Size and Value anomalies are constructed using the 6 value-weighted portfolios formed on size and book-to-market. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed with respect to the 30% and 40% NYSE quantiles of book to market BE/ME (Value, Neutral and Growth) .

$$SMB = \frac{1}{2} (Small\ Value + Small\ Neutral + Small\ Growth) - \frac{1}{2} (Big\ Value + Big\ Neutral + Big\ Growth) \quad (6)$$



$$HML = \frac{1}{2} (Small\ Value + Big\ Value) - \frac{1}{2} (Small\ Growth + Big\ Growth) \quad (7)$$

**Investment:** The Fama/French investment anomaly is constructed using the 6 value-weighted portfolios formed on size and investment. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed with respect to the 30% and 40% NYSE quantiles of investemnt Inv (Conservative, Neutral and Aggressive) .

$$CMA = \frac{1}{2} (Small\ Conservative + Big\ Conservative) - \frac{1}{2} (Small\ Aggressive + Big\ Aggressive) \quad (8)$$

**Profitability:** The Fama/French profitability anomaly is constructed using the 6 value-weighted portfolios formed on size and operating profitability. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed with respect to the 30% and 40% NYSE quantiles of operating profitability OP (Robust, Neutral and Weak) .

$$RMW = \frac{1}{2} (Small\ Robust + Big\ Robust) - \frac{1}{2} (Small\ Weak + Big\ Weak) \quad (9)$$

**Momentum:** Fama-French implementation of momentum anomaly is based on the 6 value-weighted portfolios formed on size and prior ( $t-2$ - $t-13$ ) returns. The portfolios, which are formed monthly, are the intersections of 2 portfolios formed on size (Small vs Big with respect to NYSE median market equity) and 3 portfolios formed on prior ( $t-2$ - $t-13$ ) returns (High, Neutral, Low compared to NYSE stocks quantiles).

$$Mom = \frac{1}{2} (Small\ High + Big\ High) - \frac{1}{2} (Small\ Low + Big\ Low) \quad (10)$$

## C Definitions / Equations

### C.1 Portfolio's turnover

Portfolio turnover at the review date  $t$  is computed as follows :

$$Turnover_t = \sum_{s=1}^{Nsec} \left| \frac{AUM_t w_t^s - AUM_{t-1} w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)}{AUM_t} \right| \quad (11)$$

$$AUM_t = AUM_{t-1} \sum_{s=1}^{Nsec} w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s) \quad (12)$$

$$Turnover_t = \sum_{i=1}^{Nsec} \left| w_t^s - \frac{w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)}{\sum_{s=1}^N w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)} \right| \quad (13)$$

Where  $AUM_t$  and  $AUM_{t-1}$  are the portfolio size respectively at review date  $t$  and  $t - 1$ .  $Nsec$  is the number of securities composing the portfolio.  $w_t^s$  is the weight of the stock  $s$  after the review date  $t$ .  $r_{t-1,t}^s$  is the return of stock  $s$  on the period between review date  $t - 1$  and  $t$ .

### C.2 Garman Klass Volatility definition

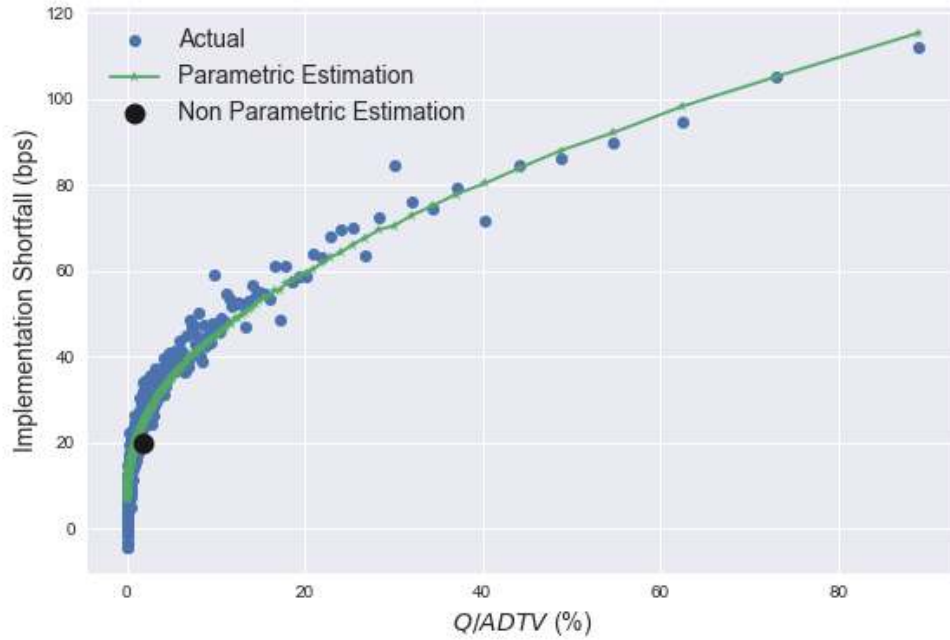
Garman-Klass estimate of the volatility uses the open, high, low and close prices of the day. This estimate is robust and very close in practice to more sophisticated ones. The formula is given by:

$$\sigma_k^{GK}(d) = \sqrt{\frac{1}{N} \sum_{t=1}^N \frac{1}{2} \log \left( \frac{H_{d-t}^k}{L_{d-t}^k} \right)^2 - (2 \log(2) - 1) \log \left( \frac{C_{d-t}^k}{O_{d-t}^k} \right)^2} \quad (14)$$

where the indexation  $k$  refers to the stock.  $d$  to the calculation day.  $N$  is the length of the rolling window in day. In our case 252 trading days.  $O_t^k$ ,  $H_t^k$ ,  $L_t^k$ ,  $C_t^k$  are respectively the open, high, low, close prices of stock  $k$  at day  $t$ .

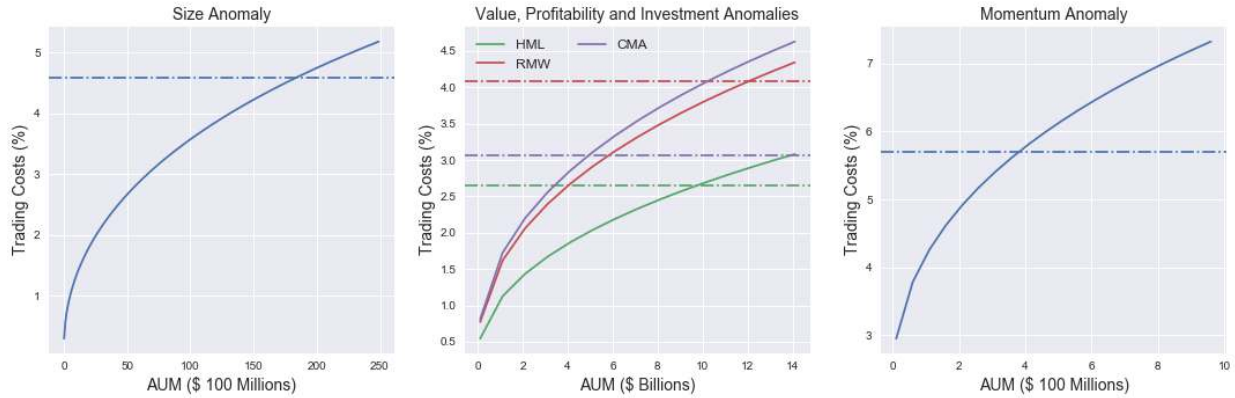
### Figure 1. Parametric vs Non-Parametric Estimation of Implementation Shortfall

This figure plots the average implementation shortfall for ANcerno tickets. We sort all trades into 1000 bins based on their participation rate  $\frac{Q}{ADV}$ . The blue dots represent the average implementation shortfall of each of the bins. The green line, is the corresponding parametric estimate based on Kyle (1985) model averaged by bin. Finally the black marker indicates the average tickets participation rate and the average non-parametric estimation

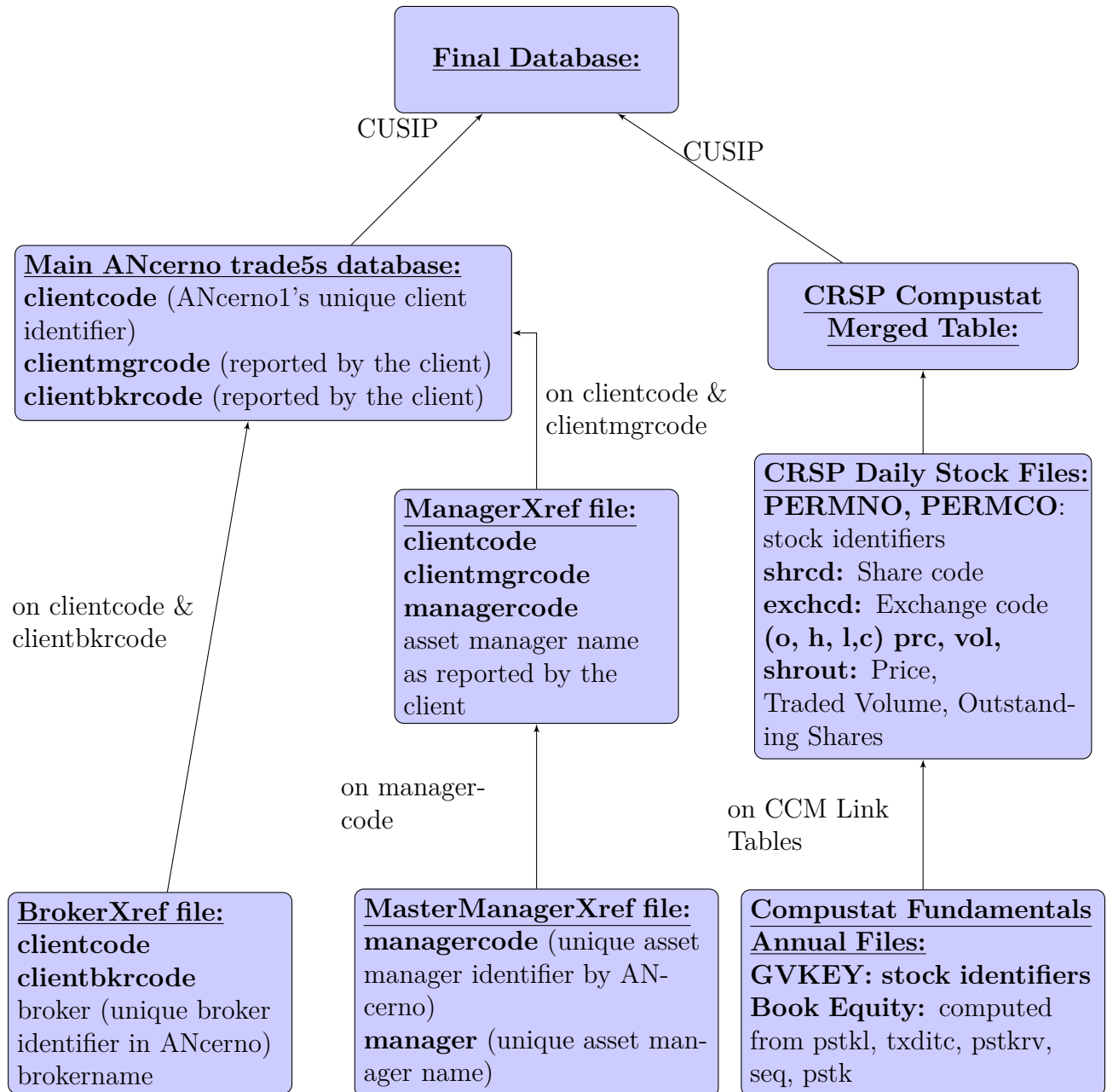


## Figure 2. Break even capacity

The left hand figure plots the trading costs of the low turnover anomalies with respect to portfolio size in \$ billion. The green line for size and the blue lines for value). The right hand figure shows the result for the monthly rebalanced momentum anomaly. The Dashed lines represent the annualized average return of the anomalies on their respective back-testing period



**Figure 3.** Merging referential, market data, fundamental data and trade level data together



**Table 1. Descriptive Statistics of ANcerno Institutional Execution Database**

The table gives descriptive statistics on ANcerno trading data for each year of our sample period (From January 1999 until June 2015). The number of institutions refers to the number of unique clientcodes. The number of Funds, Managers or Accounts is the number of unique clientmgrcodes. The number of brokers corresponds to the number of unique broker identifiers from BrokerXref file where the couple clientcode-clientbkrcode is present in ANcerno. The amount traded in \$ is the sum of the dollar volume executed by ANcerno institutions in the sample. The amount traded in % of market volume is the ratio of ANcerno reported volumes w.r.t CRSP daily turnover

	# Institu- tions	# Funds, Managers or Accounts	# Brokers	# Stocks	Amount traded (\$ billions)	Amount traded (% of market volume)
Full Sample	1078	148621	1488	10044	51310	5.44
1999	381	6153	657	6291	2060	2.28
2000	374	6390	669	6239	3181	2.46
2001	401	13654	716	5396	3026	3.37
2002	428	16847	765	4935	3096	5.00
2003	405	26861	751	4930	2667	6.03
2004	408	23112	716	4126	4122	8.28
2005	379	18928	761	4912	3930	6.17
2006	403	22081	753	4773	4232	5.98
2007	381	28999	738	4941	4506	5.35
2008	338	26600	701	4507	4187	3.95
2009	303	41848	650	4207	2875	3.70
2010	258	43227	632	3951	2508	3.69
2011	-	-	675	3884	1828	5.96
2012	-	-	723	3715	2596	8.30
2013	-	-	647	3755	2714	9.39
2014	-	-	531	3809	2742	7.12
June 2015	-	-	392	2753	1033	3.64

**Table 2. Characteristics of ANcerno Traded Stocks**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1<sup>st</sup>, 1999 to June 30<sup>th</sup>, 2011. The left part of panel A gives the ratio of common stocks (those with a sharecode of 10 or 11) present in ANcerno relative to CRSP database. The right part of panel A displays the ratio on market-capitalization coverage. It is computed as the sum of market capitalization of stocks present in ANcerno divided by CRSP universe total market capitalization. Panel B reports descriptive statistics for stocks traded by ANcerno institutions. We obtain price data (prices, traded volume, outstanding shares) from CRSP database, Book-to-Market from Compustat. GK volatility is the Garman Klass (1980) estimation of the volatility. Turnover is the average percentage of outstanding shares traded on a single day

Panel A: ANcerno Coverage								
	% Number of Stock				% Market Capitalization			
	ALL	NYSE	AMX	NASDAQ	ALL	NYSE	AMX	NASDAQ
Full Sample	0.72	0.92	0.36	0.66	0.98	0.99	0.73	0.96
1999	0.63	0.92	0.35	0.55	0.99	0.99	0.83	0.97
2000	0.67	0.92	0.32	0.62	0.99	0.99	0.83	0.99
2001	0.66	0.93	0.28	0.61	0.99	0.99	0.85	0.98
2002	0.71	0.94	0.33	0.66	0.99	0.99	0.88	0.98
2003	0.77	0.96	0.40	0.73	0.99	1.00	0.88	0.98
2004	0.84	0.97	0.52	0.83	0.99	1.00	0.90	0.98
2005	0.84	0.98	0.51	0.83	0.99	1.00	0.86	0.98
2006	0.85	0.98	0.52	0.84	0.99	1.00	0.81	0.98
2007	0.86	0.98	0.56	0.85	0.99	1.00	0.81	0.98
2008	0.83	0.98	0.43	0.82	0.99	1.00	0.73	0.98
2009	0.83	0.97	0.43	0.80	0.99	1.00	0.68	0.98
2010	0.83	0.97	0.48	0.79	0.99	1.00	0.84	0.98
2011	0.83	0.97	0.50	0.79	0.99	1.00	0.87	0.98
2012	0.43	0.75	0.02	0.29	0.96	0.98	0.21	0.91
2013	0.44	0.74	0.03	0.30	0.94	0.97	0.31	0.89
2014	0.43	0.69	0.03	0.30	0.92	0.95	0.34	0.85
June 2015	0.37	0.59	0.02	0.27	0.86	0.89	0.33	0.80

Panel B: Stock Characteristics				
	mean	25%	50%	75%
Market Capitalization (\$ 100 billion)	24.0	0.57	2.26	9.33
Average Traded Volume (\$ million)	23.7	0.16	1.08	7.72
Book-to-Market Ratio	1.05	0.32	0.57	0.93
Lagged 12-month Return (%)	0.16	-0.36	-0.04	0.42
Turnover (%)	1.25	0.21	0.48	1.03
GK Volatility (%)	29.4	13.0	21.8	37.4

**Table 3. ANcerno Tickets Characteristics**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1<sup>st</sup>, 1999 to June 30<sup>th</sup>, 2015. Parent Ticket is an order sent by an institutional investor. It could be split into multiple child tickets. Participation rate is the size of the parent ticket compared to the 12 months average daily volume. Duration is the execution period calculated in days. Commissions are computed as the ratio of Commissions per share divided by the open price of the day the ticket is issued.

	# Parent Tickets ( million)	Av Parent Ticket size (\$ million)	% of buy	# Child Tickets	Particip. rate (%)	Dura- tion (days)	Commiss- ions (bps)
Full Sample	6.99	0.62	50.81	2.80	2.27	1.73	9.27
1999	1.92	1.12	53.73	2.66	6.42	2.85	11.52
2000	3.02	1.09	54.64	2.54	4.43	2.24	12.47
2001	3.71	0.87	55.83	2.45	3.89	2.03	14.62
2002	4.30	0.80	54.87	2.66	3.82	2.15	18.85
2003	4.99	0.62	53.62	2.90	3.53	1.96	17.08
2004	6.18	0.92	52.86	3.25	3.38	1.96	13.00
2005	6.73	0.80	51.35	3.26	2.74	1.79	9.99
2006	9.47	0.56	49.75	3.66	1.74	1.56	9.73
2007	10.23	0.55	49.81	4.13	1.48	1.50	7.07
2008	11.66	0.48	48.94	2.71	0.90	1.40	8.60
2009	10.31	0.39	47.64	2.29	0.94	1.47	8.05
2010	10.62	0.37	47.22	2.11	0.84	1.35	4.04
2011	8.20	0.24	48.09	2.25	0.42	1.22	2.89
2012	10.93	0.26	47.82	2.40	0.44	1.26	3.13
2013	7.69	0.37	48.79	2.66	0.65	1.38	3.62
2014	5.93	0.49	48.02	2.93	0.73	1.54	3.69
June 2015	2.82	0.39	49.50	3.59	0.58	1.56	3.69



**Table 4. ANcerno Parent Tickets Average Implementation Shortfall**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1<sup>st</sup>, 1999 to June 30<sup>th</sup>, 2015. Our sample include only common stocks (those with a share code of 10 or 11 in CRSP). The split by trading venue is based on CRSP. The split between large and small caps is based on the NYSE median capitalization in December of each year. Buy (respectively Sell) correspond to the average implementation shortfall for buy (respectively sell) tickets.  $R^M$  is the average daily return of the equally weighted basket composed of CRSP stocks.  $\sigma^{GK}$  is the average Garman Klass Volatility of CRSP stocks computed over 1 year rolling window. Spread is the monthly average quoted Bid-Ask spread relative to the mid price obtained from TRTH trade database

	mean (bps)	25% (bps)	50% (bps)	75% (bps)	Buy (bps)	Sell (bps)	$R^M$ (bps)	Spread (bps)	$\sigma^{GK}$ (%)
Full Sample	15.05	-103.83	7.31	127.78	9.86	20.74	5.13	17.58	22.78
Large Cap	8.69	-64.25	4.21	79.67	8.54	8.98	7.72	14.35	9.69
Small Cap	14.35	-92.60	8.32	117.81	13.08	15.83	3.79	21.33	19.63
NYSE	6.96	-64.69	3.32	77.42	6.30	7.72	4.53	14.96	10.49
AMEX	11.38	-118.30	7.11	136.37	1.45	20.41	5.33	27.18	18.41
NASDAQ	16.19	-77.22	8.55	105.35	16.56	15.93	6.16	17.73	13.54
1999	43.81	-126.23	18.60	184.33	48.29	38.56	12.57	52.62	26.16
2000	32.48	-183.93	15.36	232.42	22.01	45.01	-3.95	51.73	33.06
2001	25.32	-148.23	15.38	191.28	8.25	46.67	12.17	33.03	32.42
2002	23.36	-138.25	11.32	173.83	-15.00	69.70	-3.93	25.30	26.11
2003	20.00	-92.29	11.41	125.51	30.71	7.52	23.76	15.62	23.74
2004	11.23	-79.01	7.21	100.64	12.42	9.88	8.46	11.66	17.08
2005	8.66	-72.67	6.49	91.10	9.98	7.23	2.01	10.73	15.51
2006	6.30	-76.93	4.60	91.18	7.29	5.22	6.56	9.49	15.67
2007	3.34	-88.32	3.23	98.88	1.74	5.07	-2.27	9.19	16.12
2008	16.23	-161.71	5.86	182.76	-16.95	50.72	-21.02	17.27	25.45
2009	18.88	-134.41	8.67	162.38	31.41	5.84	22.79	16.37	36.70
2010	10.98	-90.16	6.67	110.98	18.99	3.22	11.10	10.52	21.84
2011	12.28	-99.19	6.50	120.03	5.84	18.94	-3.02	10.90	18.03
2012	1.53	-69.74	0.00	72.71	4.27	-1.21	8.54	17.91	9.35
2013	7.39	-51.66	4.47	64.97	9.49	5.28	15.92	13.04	7.40
2014	3.84	-59.32	2.23	66.47	0.52	7.15	2.82	12.53	7.60
June 2015	0.75	-59.71	0.00	61.16	1.41	0.09	-1.30	12.46	7.35

**Table 5. ANcerno % Coverage of Anomalies Sub-Portfolios**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1<sup>st</sup>, 1999 to June 30<sup>th</sup>, 2015. Prices data are obtained from CRSP database while fundamental data are derived from Compustat. We implement the size, value, profitability and investment anomalies based on Fama-French 6 sub-portfolios methodology (described in the Appendix B) and momentum anomaly following Carhart. The table below shows the percentage number of stocks in each sub-portfolio present in ANcerno and how much it represent in terms of market value.

Panel A: Fama-French Anomalies							
Sub-portfolios	Big High	Big Low	Big Medium	Small High	Small Low	Small Medium	Average
				Size, Value			
Number Stocks (%)	97.7	98.4	97.6	37.9	57.3	54.0	73.82
Market Cap (%)	98.88	99.76	98.95	75.45	82.77	80.23	89.3
				Investment			
Number Stocks (%)	96.5	99.1	97.5	40.0	56.8	56.2	74.3
Market Cap (%)	99.5	99.1	98.8	77.3	81.9	79.9	89.4
				Profitability			
Number Stocks (%)	98.2	98.8	98.0	45.3	51.5	55.9	74.6
Market Cap (%)	99.1	99.3	97.9	80.6	82.5	77.2	89.4
Panel B: Carhart Momentum Anomaly							
Sub-portfolios	Big Down	Big Up	Small Down	Small Up	Average		
Number Stocks (%)	97.5	97.7	40.9	51.1	71.80		
Market Cap (%)	99.12	99.26	76.25	78.96	88.4		

**Table 6. Model calibration on ANcerno tickets**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1<sup>st</sup>, 1999 to June 30<sup>th</sup>, 2015. Quoted intraday bid-ask spreads are obtained from TRTH database averaged over the month.  $\sigma_k^{GK}$  and  $ADV$  are respectively the Garman Klass volatility and the average trading volume computed from CRSP database over a 12 months rolling window. The coefficients of the rolling regressions are estimated at the end of each month on 1000 bins based on tickets participation rate  $\frac{Q_k}{ADV_k}$  where only tickets submitted by institutional investors on the latest 12 months are considered. The bootstrap draws randomly 1/12 of the database tickets. 10 000 regressions are estimated on the 1000 corresponding bins

	1Y Rolling Window		10 <sup>4</sup> Bootstrap	
	$\psi_k$	$\sigma_k^{GK} \sqrt{\frac{Q_k}{ADV_k}}$	$\psi_k$	$\sigma_k^{GK} \sqrt{\frac{Q_k}{ADV_k}}$
Av coef	0.40	0.77	0.42	0.75
2.5% quantile	0.35	0.68	0.37	0.70
97.5% quantile	0.45	0.76	0.46	0.79
std err	0.02	0.001	0.02	0.001
$t_{stat}$	17.68	37.34	20.11	39.67
$P >  t $	0.01	0.00	0.00	0.00
Average $R_{adj}^2$		0.88		0.91
Nbr of regressions		152		10000
Nbr of observations per regression		1000		1000

**Table 7. Non-parametric estimation results**

The back history is run from 30<sup>th</sup> June 1999 to 30<sup>th</sup> June 2015. Ann Gross return is the strategy annualized average gross return in %. Volatility is the annualized monthly standard deviation in %. Ann Net return is the annualized average return after trading costs are deduced. Ann turnover is the annual turnover computed as in Appendix B.2.

	Ann Gross Returns (%)	Volatility (%)	Ann Net Returns (%)	Ann Turnover	Trading Costs (bps)
Panel A: Size Anomaly (SMB)					
CRSP Univ Portfolio	4.84	12.41	-	0.45	-
CRSP Univ Long Leg	11.99	20.61	-	0.25	-
CRSP Univ Short Leg	-7.16	15.41	-	0.20	-
Ancerno Univ Portfolio	4.46	11.11	4.29	0.54	15.69
Ancerno Univ Long Leg	11.57	20.35	11.42	0.35	13.65
Ancerno Univ Short Leg	-7.11	15.37	-7.086	0.19	2.04
Panel B: Value Anomaly (HML)					
CRSP Univ Portfolio	2.94	12.05	-	0.76	-
CRSP Univ Long Leg	10.60	18.35	-	0.44	-
CRSP Univ Short Leg	-7.67	18.93	-	0.32	-
Ancerno Univ Portfolio	2.22	12.16	1.98	0.76	22.50
Ancerno Univ Long Leg	10.05	18.84	10.02	0.44	2.56
Ancerno Univ Short Leg	-7.83	18.36	-7.62	0.32	19.95
Panel D: Profitability Anomaly (RMW)					
CRSP Univ Portfolio	4.20	10.80	-	0.71	-
CRSP Univ Long Leg	9.80	15.47	-	0.31	-
CRSP Univ Short Leg	-5.60	21.82	-	0.40	-
Ancerno Univ Portfolio	4.77	10.38	4.45	0.76	30.61
Ancerno Univ Long Leg	10.57	15.40	10.42	0.33	13.90
Ancerno Univ Short Leg	-5.80	21.66	-5.62	0.42	16.71
Panel E: Investment Anomaly (CMA)					
CRSP Univ Portfolio	3.08	6.35	-	1.25	-
CRSP Univ Long Leg	9.96	17.77	-	0.68	-
CRSP Univ Short Leg	-6.89	18.69	-	0.57	-
Ancerno Univ Portfolio	3.02	7.00	2.69	1.19	31.39
Ancerno Univ Long Leg	10.47	17.61	10.34	0.64	12.44
Ancerno Univ Short Leg	-7.46	18.63	-7.25	0.55	18.95
Panel C: Momentum Anomaly (UMD)					
CRSP Univ Portfolio	5.15	18.22	-	7.32	-
CRSP Univ Long Leg	11.71	18.46	-	3.28	-
CRSP Univ Short Leg	-6.56	24.52	-	4.04	-
Ancerno Univ Portfolio	5.09	18.52	2.86	8.34	222.49
Ancerno Univ Long Leg	11.71	18.65	10.78	3.84	93.30
Ancerno Univ Short Leg	-6.62	24.81	-5.33	4.50	129.19

**Table 8. Parametric estimation results**

The back history is run from 30<sup>th</sup> June 1999 to 30<sup>th</sup> June 2015. Ann. Net return is the annualized average return after trading costs are deduced. Volatility is the annualized monthly standard deviation in (%). Annual turnover is computed as in Appendix B.2. Av Participation Rate is the average ticket size w.r.t daily turnover rebalanced by the anomalies. The intervals between brackets correspond to the 5% confidence interval derived from the 2.5% and 97.5% quantiles of bootstrapped model parameters

AUM	Ann. Net Return (%)	Volatility (%)	Trading Costs (bps)	Turnover (Monthly)	Av Particip Rate (%)
Panel A: Size Anomaly (SMB)					
\$1 million	4.46 [4.45 4.48]	12.32	19.17 [17.86 20.48]	0.04	0.00
\$100 million	4.35 [4.33 4.37]	12.32	30.11 [28.14 32.07]	0.04	0.47
\$1 billion	4.09 [4.06 4.13]	12.32	56.04 [52.53 59.54]	0.04	4.57
Panel B: Value Anomaly (HML)					
\$1 million	2.11 [2.09 2.14]	11.61	30.84 [28.71 32.98]	0.06	0.01
\$100 million	1.88 [1.85 1.91]	11.61	54.22 [50.75 57.68]	0.06	1.28
\$1 billion	1.33 [1.27 1.40]	11.61	108.77 [102.27 115.24]	0.06	12.21
Panel D: Profitability Anomaly (RMW)					
\$1 million	4.24 [4.21 4.27]	11.21	42.56 [39.77 45.3]	0.06	0.02
\$100 million	3.89 [3.85 3.94]	11.21	77.05 [72.35 81.74]	0.06	1.82
\$1 billion	3.10 [3.01 3.19]	11.21	156.44 [147.54 165.31]	0.06	17.00
Panel E: Investment Anomaly (CMA)					
\$1 million	3.78 [3.75 3.80]	7.54	43.76 [40.88 46.63]	0.10	0.02
\$100 million	3.40 [3.35 3.45]	7.54	81.07 [76.11 86.02]	0.10	1.97
\$1 billion	2.55 [2.45 2.64]	7.54	166.64 [157.12 176.12]	0.10	18.31
Panel C: Momentum Anomaly (UMD)					
\$1 million	2.46 [2.30 2.62]	18.20	253.45 [236.95 269.93]	0.610	0.01
\$100 million	0.93 [0.70 1.16]	18.19	417.43 [393.59 441.11]	0.610	0.49
\$1 billion	-2.03 [-2.35 -1.69]	18.17	741.13 [705.26 776.40]	0.610	3.61

**Table 9. ANcerno Variables**

The table describes the main ANcerno variables used to compute the results of this paper

Variables	Description	Source File
client/ Family/Broker Identification Variables		
clientcode	unique client identifier by ANcerno	Main dataset
clientmgrcode	trading investment manager code as reported by the client	Main dataset
clientbkrcode	code of the broker executing the trade as reported by the client	Main dataset
managercode	unique trading investment manager code attributed by ANcerno	ManagerXref
manager	unique investment manager name by ANcerno	Master ManagerXref
broker	unique numeric broker identifier by ANcerno	BrokerXref
Order Identification Variables		
cusip	stock cusip	Main dataset
stockey	ANcerno stock identifier	Main dataset
onumber	Ticket identifier for a single stock, side and date	Main dataset
lognumber	ANcerno identifier for data source	Main dataset
odtOrderDate	Date where the broker receives the ticket	Main dataset
odtLastDate	Last date allowed to liquidate the order	Main dataset
ov	Ticket size	Main dataset
Side	buy or sell (1 = Buy; -1 = Sell)	Main dataset
Trade Identification Variables		
tradedate	date of the trade	Main dataset
xdtX	Execution time	Main dataset
Price	price per share	Main dataset
Volume	number of traded shares	Main dataset
Commission USD	per trade commission in USD	Main dataset