Copenhagen Business School
HANDELSHØJSKOLEN

CFIVU1002U

Master's Thesis

May 15, 2020

Liquidity Risk Premium in Pairs Trading

Authors:	***Student Number:***
Simon Foldberg Jeppesen	S101278
Luciano Matoso	S124230

Advisor:

Martin Richter

Abstract

We employ the pairs trading algorithm popularized by Gatev et al. (1999, 2006) on U.S. stocks data from 1963 until 2019. As Do and Faff (2011) we find that the pairs trader does not earn significant returns after transaction costs. We hypothesize that the pairs trader is still making a profit as compensation for providing liquidity during shocks to market and funding liquidity proxied by the VIX and TED-spread, respectively. We find evidence that the pairs trader on average is net long illiquid stocks, measured as the Amihud (2002) illiquidity measure. During sub-periods with shocks to market and funding liquidity, the pairs trader increases his/her loading on even more illiquid securities, but equally so on the long and the short leg. The pairs trading strategy has become more unprofitable in recent years, but if one implements during sub-periods with scarce market and funding liquidity we do find an indication of slightly higher returns after transaction costs, but not a statistically significant compensation. A liquidity-adjusted CAPM model does not provide an indication that the pairs trader is compensated for taking liquidity risk. We therefore come to the conclusion that the pairs trader is not compensated, to a significant degree, for taking liquidity risk during sub-periods characterized by liquidity shocks. We speculate that the pairs trader in a Gatev et al. (1999, 2006) framework is sub-optimally selecting its pairs to ideally reflect a liquidity risk premium exploiting trade.

Keywords: Pairs trading, Statistical arbitrage, Liquidity risk premium.

Contents

1 Introduction

Pairs trading refers to a trading strategy that bets on the continuation of the historic co-movements in the prices of two assets, despite temporal divergences. If the pairs trader assumes that the future will be similar to the past, he takes a long position in the asset that is now "cheap" and shorts the "expensive" one. This indicates that the trade will be profitable if the historical relationship of the two assets is reestablished. The strategy was popularized by the seminal paper by Gatev et al. (1999), and has since then been studied for decades. According to Malkiel (2003), any abnormal returns to such a simplistic trading strategy must disappear with its high level of public attention. Yet somehow, studies still find that simple modifications of this strategy generates risk adjusted returns higher than the market portfolio. As the abnormal returns persist even after adjusting for common factors, this could suggest that pairs trading earns a risk premium yet to be accounted for.

We will in this study attempt to accept or reject the following initial hypothesis: The pairs trading profits are compensation for taking liquidity risk.

Empirical studies find persistence in liquidity, among those, Acharya and Pedersen (2005). Persistence in liquidity implies that returns are predictable. Stocks with a high illiquidity level today have a high expected illiquidity level in the following period. Therefore, investors required returns increase. A higher expected return leads to a lower price and this is where the pairs trader receives trading signals. Given a natural contrarian approach to investing, the pairs trader buys stocks that have gone down, while shorting stocks that have gone up. When shocks to market and funding liquidity occur, the pairs trader will therefore become a liquidity provider in the illiquid stocks, while taking liquidity in liquid stocks. Under such liquidity states, the pairs trader is subject to liquidity risk and must require a high expected return during times with shocks to market and funding liquidity which results in high barriers to arbitrage. We define liquidity risk as the risk of not being able to trade given a lack of liquidity in the market.

This paper seeks to answer following questions:

- Is the pairs trader on average a liquidity provider and therefore subject to liquidity risk?
- How does the level of liquidity provision change over time?

- Is the pairs trader earning a higher compensation during sub-periods with high barriers to arbitrage?
- Does a liquidity-adjusted CAPM model indicate that the pairs trader is compensated for taking liquidity risk?

These questions are relevant for any statistical arbitrage hedge fund, seeking to understand what the drivers of pairs trading returns are, and the potential implications of market-wide liquidity shocks. This could be relevant for funds seeking to hedge exposure to liquidity risk by pairing one of their existing investment strategy with a pairs trading strategy.

Over the years, academia has provided numerous ways of identifying the liquidity of a security. Amihud and Mendelson (1986) used the bid-ask spread as their measure of illiquidity to explain cross-sectional stock returns. Brennan and Subramanyan (1996) followed with price impact regressed on a unit trade size. Amihud (2002) provided a low-data-requirement measure considering an average price impact relative to daily dollar volume traded in a given security. Pastor and Stambaugh (2003) found that returns of securities varied with sensitivity to market liquidity. Brown, Crocker and Foerster (2007) found a relation between liquidity and momentum and information effects in the large capitalization segment. These findings imply that a single measure of liquidity is hardly able to stand alone.

Regardless of multiple accounts pointing to liquidity being a priced factor, liquidity is still not a common style to include when attempting to find the determinants of expected stock returns of an investment strategy (Subrahmanyam, 2010). One could perhaps argue that the lack of a universal measure of illiquidity has yet to be determined, but as Ibbotson (2013) states, he does not suggest that his measure of turnover is superior to other liquidity measures, but he argues that his measure is simple and works well. With that logic, Ibbotson simultaneously points out that the other common styles of investments can be measured in various ways. I.e. size, *SMB*, can be found under various capitalization constructions, quintiles, deciles or quartiles and even across multiple markets. Value versus growth, *HML*, is not limited to Fama and French (1992)'s definition as the book-to-market ratio could be measured as the price over earnings (Basu 1977), dividend over price, enterprise value over ebitda, or other fundamental ratios. Momentum, by Jegadeesh and Titman (1993), *UMD*, can be measured over a long list of different horizons and weightings. Therefore, Ibbotson places liquidity in the same category of accepted styles as size, value and momentum.

The thesis is organized as follows: In section 2, we review the key academic literature surrounding liquidity and pairs trading that will repeatedly be referred to throughout the analysis. Section 3 describes theoretical concepts applied in the methodology. Section 4 describes our methodology of replicating the pairs trading algorithm and it provides a robustness test of the portfolio construction process. Finally, it shows how we estimate a liquidity-adjusted CAPM model. Section 5 presents our empirical analysis, it is structured in four parts: The first part considers the validation of our methodology through presentation of replicated pairs trading returns, before and after transaction costs, as well as a disaggregation into a long and short leg portfolio performance. The second part seeks to answer what the pairs trader actually trades in terms of risk exposures and their relation to liquidity. The third part presents our results from liquidity shock sub-period returns. The last part presents the results from a liquidity-adjusted CAPM cross sectional model. Section 6 discuss our methodology and challenges our hypothesis. Section 7 concludes.

2 Literature Review

2.1 Amihud, Y. (2002)

In 1989, Amihud and Mendelson hypothesized that returns increases in illiquidity. While this relationship was examined widely across stocks, Amihud (2002) examined the relationship over time by proposing that: *"over time, the ex ante stock excess return is increasing in the expected illiquidity of the stock market"*, thereby suggesting that the expected return on stocks, in part, represents a premium on illiquidity.

Amihud (2002) defined an illiquidity measure, denoted *ILLIQ*. The idea follows that of Kyle (1985), who attempted to capture price impact by relating returns to order flow. The measure of illiquidity for security i was found as the daily ratio of absolute stock returns over the volume measured in dollar terms, averaged over some period of time:

$$ILLIQ_t^i = \frac{1}{Days_t^i} \sum_{d=1}^{Days_t^i} \frac{|R_{td}^i|}{V_{td}^i} \tag{1}$$

where $Days$, is the number of days averaged over. In Amihud (2002), he averages over a year. $|R|$ is the absolute return, and V is the dollar volume.

Amihud's measure of illiquidity provided a simple and low-data-requirement estimate of illiquidity. The Amihud measure stands in comparison to other illiquidity, such as the bid-ask spread or market impact, which both requires more microstructure data. This data was not readily available to the investor in less covered markets, and this is probably the reason for the Amihud measure's popularity.

The author considered NYSE stocks through 1964 to 1997, and found that expected stock returns are positively related to expected market illiquidity risk. In relation, Amihud found that sudden, unexpected, drops in market illiquidity led to drops in contemporaneous stock returns, assuming that firm cash flows remain unaffected by the liquidity of the market. These effects were found to be larger for small capitalization companies and smaller for large capitalization companies, indicating that the size-effect varies with market liquidity.

The intuition behind *ILLIQ* is that it captures the daily average relation between a unit change in volume and the price change. Another interpretation presented by the author relates to a study by Harris and Raviv (1993). They found that disagreements between investors about new information lead to increased trading, while agreements about the implications of news leads to less trading as prices change without much trading. Thereby, *ILLIQ* has the interpretation of consensus among investors about firm specific news.

2.2 Liquidity Factors

In order to justify the hypothesis that pairs trading profits are compensation for taking liquidity risk, one must first validate that a robust liquidity premium does in fact exist. Literature researching the liquidity premium was popularized by Amihud and Mendelson (1988), who was able to show that illiquid stocks outperformed liquid counterparts by using the bid-ask spreads as a liquidity measure. This section provides an overview of a selection of liquidity factors each using a different method of measuring liquidity.

2.2.1 Pastor and Stambaugh (2003)

Pastor and Stambaugh (2003) analyze whether market-wide liquidity is priced by testing whether changes in a stock's expected returns are cross-sectionally related to changes in the stock returns sensitivity to market liquidity. They measure liquidity from the following ordinary least squares (OLS) regression:

$$r^e_{i,d+1,t} = \theta_{i,t} + \phi_{i,t} r_{i,d,t} + \gamma_{i,t} \operatorname{sign}\left(r^e_{i,d,t}\right) \cdot v_{i,d,t} + \epsilon_{i,d+1,v} \quad d = 1, \ldots, D \tag{2}$$

Where $r_{i,d,t}$ is the return of the stock i on day d in month t and $r^e_{i,d,t} = r_{i,d,t} - r_{m,d,t}$, where $r_{m,d,t}$ measures the value-weighted return of the market from the CRSP database. $v_{i,d,t}$ is the dollar volume of the stock. The OLS estimate of $\gamma_{i,t}$ measures the liquidity for stock i in month t.

From equation 2, we see that the stock return in excess of the market is explained by a lagged stock return and the dollar volume signed by the lagged stock return in excess of the market. By using the return in excess of the market in both the dependent variable and to sign the volume, Pastor and Stambaugh attempt to isolate the stock-specific effects. Amongst low-price stocks the tick size may represent large relative price changes, this constitutes a problem with using a stock's total return of zero percent for signing the volume. However, using the excess return solves this issue as the excess of market returns are very unlikely to be zero. By adding the lagged stock return as an independent variable, this term captures non-volume related reversals, such as reversals imposed by a stocks tick size. Furthermore, the lagged stock return variable allows for isolation of volume-related return reversals in the second term.

The authors argue that one should in general expect a negative relation between the stock returns in excess of the market today relative to the excess returns of stock i yesterday. The reason for this expectation is not specified. One could speculate that the authors expect that stocks that outperform the market today are outperforming due to excess purchase pressure pushing stocks away from its 'true' market price and, subsequently, one could expect the stocks to revert on the following day. Moreover, the authors expect that with lower liquidity one obtains larger absolute magnitude of reversals.

Pastor and Stambaugh also provide a comparison between their liquidity measure and that of Lo and Wang (2000), Chordia, Roll and Subrahmanyam (2000, 2001, 2002), Jones (2002), and Amihud (2002). The authors point out a fallacy with using trading activity measures such as the volume and turnover used by the aforementioned studies. PS argue that while these trading activity measures seem to capture cross-sectional differences in liquidity well, they do not appear to capture time variation in liquidity well. Illiquid markets tend to have high trading volume, PS points to the stock market crash in 1987, liquidity was at a record low, but trading volume where set its record high. This presents a problem with the Amihud (2002) measure used in this study (Equation 1). During a crisis, one should expect illiquidity costs to be high. However, with record volumes, as the denominator in equation 1 increases, the illiquidity costs decrease. This relation is quantified by the Pastor and Stambaugh. They find a positive time-series correlation between their liquidity measure and dollar volume during the entire sample period, but when using sub-samples of low-liquidity months, that relation turns negative. For that reason Pastor and Stambaugh does not use trading activity measures to proxy for time variation in liquidity.

One potential contributor to the negative autocorrelation in Pastor and Stambaugh's measure is nonsynchronous trading. Excess returns of stocks could simply be caused by a lack of liquidity in stocks relative to the market. Therefore, on the following day after an excess return for stocks, one should expect a reversal. However, the probability of a reversal on day $d + 1$ is higher when the volume on day d is low. This is contrasting with the interpretation of $\gamma_{i,t}$ from equation 2, where one should expect reversals if the volume of day d is high. Moreover, nonsynchronous trading is expected to have a greater impact in low-liquidity months, but average volume and turnover in those low-liquidity months are shown to be slightly larger than usual. Regardless of likelihood, the negative value of illiquidity could be affected by a contribution from nonsynchronous trading.

From cross sectional average of stock liquidity measures, the authors estimate their monthly market liquidity measure. Consistent with flight to quality and flight to liquidity, when market liquidity drops, stocks returns display negative correlations with fixed income returns. As found by Acharya and Pedersen (2005), Pastor and Stambaugh find commonality across stocks in the monthly liquidity measure. Both findings speak in favor of the hypothesis that market liquidity could be a priced variable by investors. To construct their liquidity factor, the authors use the monthly liquidity measure estimated above to construct a liquidity factor relying on the principle of greater return reversals induced by order flow when liquidity is low.

From 1966-1999, their factor returns 7.5% annually after adjusting for a multifactor model of market return, size, value and momentum. They further find that half the alpha in a momentum strategy is compensation for taking liquidity risk. The choice to include momentum in the analysis is motivated by the anomaly of buying winners and selling losers over the past year produces abnormal returns (Jegadeesh and Titman (1993)). Finally, Pastor and Stambaugh conclude that future studies should focus on attempting to explain pricing anomalies through liquidity.

2.2.2 Amihud (2014)

Amihud (2014) presents a liquidity factor, *IML*, defined as Illiquid-minus-Liquid stocks portfolios. Illiquidity is calculated following equation 1, from his own 2002 paper. Portfolios are constructed by sorting into 15 portfolios in two dimensions based on *ILLIQ* and standard deviation. As *ILLIQ* and standard deviation are positively correlated, Amihud seeks to avoid confounding between the two by sorting into tertiles based on the standard deviation and, within each tertile, he divides into quintiles based on the *ILLIQ* measure. To mitigate reversals and momentum in stock returns, Amihud calculates portfolio returns three months after the end of the formation period. Each portfolio return is equal to the value weighted average monthly stock return. Finally, the *IML* factor return is calculated by taking the average of the three highest and lowest *ILLIQ* quintiles, and buying the illiquid portfolios while short selling the liquid portfolios. The factor was constructed using a sample of NYSE and AMEX stocks.

Following a multi-factor CAPM model using the return factors of Fama and French (1992) and Carhart (1997), the factor generates an annual alpha of 4.06% from 1950 to 2012, and is robust to sub-periods. It has a positive and highly significant *SMB* beta, which seems rather intuitive as small stocks are naturally more illiquid relative to the large stocks. Estimating a multi-factor model without the SMB results in an annualized alpha of 5.6%, which Amihud suggests speaks to a limited limited size-effect on the *IML* factor.

Conclusively, Amihud points to the *IML* factor being priced, but when controlling for funding illiquidity, the risk factor is only statistically significant in periods of time where funding illiquidity is high, which is consistent with Brunnermeier and Pedersen (2009). Amihud measures funding illiquidity as the yield spread between corporate bonds rated BAA and AAA. Higher spread equals higher funding illiquidity.

2.2.3 Liu (2006)

As mentioned, liquidity can be measured in many ways, and does not have one universally accepted measure, as each measure captures something different. Amihud (2014) is therefore, naturally, not the only paper proposing of a liquidity factor. Liu (2006) proposes a return factor, *LIU*, using an illiquidity measure that evaluates the proportion of zero-volume days over the previous year and by turnover. By construction the two methods also differ. While *IML* are value-weighted, quintile portfolios with pre-rankings into terciles based on volatility of returns, *LIU* are equal weighted, decile portfolios with no pre-rankings. However, both regression capture about the same alpha indicating that the liquidity premium is not captured by the other factors. Amihud then includes the *LIU* factor to test whether the two factors capture the same liquidity premium. The results show that the two factors indeed capture different information about the liquidity premium, and that both systematic risk factors can together contribute to explain expected stock returns.

2.2.4 Ibbotson, Chen, Kim and Hu (2013)

Ibbotson et. al. (2013) present a measure of illiquidity by considering stock turnover. At the time, this was already an established measure of illiquidity with a negative correlation to equity returns (Haugen and Baker (1996) and Datar, Naik and Radcliffe (1998)). Turnover was found by Idzorek, Xiong, and Ibbotson (2012) to exhibit greater explanatory power than that of Amihud (2002)'s illiquidity measure. Ibbotson defines illiquidity as the annual share turnover defined as the sum of 12 monthly volumes divided by each month's shares outstanding. The authors make use of the top 3500 stocks in U.S. equity markets from 1971 to 2011.

They construct a liquidity factor from this measure, by splitting their stock universe into quintiles based on their ranked illiquidity measure. They construct a long-short portfolio by buying the least liquid quintile and shorting the most liquid quintile. The resulting liquidity factor premium, as with Pastor and Stambaugh, Liu and Amihud, could not be explained by the Carhart four-factor model constituting market risk premium, small-minus-big, high-minus-low and winners-minus-losers.

Besides a liquidity factor, Ibbotson et al. (2012) also analyze how a the returns of the first quartile of a liquidity sorted portfolio compared with the market portfolio as well as the first quartile of returns of size, value, and momentum. Their findings suggest that low liquidity outperforms the equally-weighted market portfolio, size and momentum, but still falls short to value.

Conclusively, Ibbotson points to the clear-cut argument for a liquidity premium. Liquidity is desired by investors, who are clearly willing to pay more for liquid stocks. Illiquidity is costly as it takes longer to trade and usually costs more to trade an illiquid security relative to a liquid one. Therefore, in equilibrium, one finds a higher gross return as a compensation of illiquidity. Small stocks can be argued to be more risky, but should that necessarily be the case for value stocks or high momentum? With this argument, Ibbotson questions whether the same clear argument can be provided for, particularly, value and momentum, or whether these styles are rather compensation for other characteristics that markets put a lower price to.

In summary, these papers seem to suggest that liquidity is a priced factor, regardless of the choice of liquidity measure, while simultaneously not being explained simply by a *SMB* factor. Conclusively, we will maintain our initial hypothesis that pairs trading profits could be compensation for taking liquidity risk.

2.3 Acharya and Pedersen (2005)

This paper is described in more detail, as future calculations in this thesis will be based on the framework presented by Acharya and Pedersen (AP).

When liquidity is low, trading difficulty is usually higher, and small firms have a relatively lower liquidity than large firms, Pastor and Stambaugh (2003). When trading activity drops, implicit costs increase and prices turns more volatile. The explicit transaction costs similarly increase when liquidity is low, as market-makers, who provide liquidity, end up requiring higher collateral and margin requirements. With these mechanics in mind, one should expect illiquid securities being priced relatively lower than liquid securities as investors seek compensation for taking on the liquidity risk. Therefore, in 2005, AP introduced a liquidity-adjusted CAPM model (LCAPM), arguing that a shock to the market-wide liquidity is a source of systematic risk. Investors price this risk by requiring higher expected returns on securities with higher liquidity-related risk, resulting in lower prices for these high liquidity risk securities.

Introducing the Liquidity-Adjusted CAPM

The liquidity-adjusted CAPM model considers how asset gross returns,

$$r_t^i = \frac{D_t^i + P_t^i}{P_{t-1}^i} \tag{3}$$

depends on its relative illiquidity costs, market returns and relative market illiquidity.

$$c_t^i = \frac{C_t^i}{P_{t-1}^i} \tag{4}$$

$$r_t^M = \frac{\Sigma_i S^i (D_t^i + P_t^i)}{\Sigma_i S^i P_{t-1}^i} \tag{5}$$

$$c_t^M = \frac{\Sigma_i S^i C_t^i}{\Sigma_i S^i P_{t-1}^i} \tag{6}$$

Where P is the asset price, D is the dividend, C is the illiquidity costs measured as a per-share cost of selling security i, and S is the number of shares outstanding.

The authors define an imagined economy, in which security i has a dividend of $D_t^i - C_t^i$ and no illiquidity costs. In this imagined economy the standard CAPM holds (Markowitz (1952), Sharpe (1964), Lintner (1965) and Mossin (1966)). They then claim that equilibrium prices are the same in an economy with frictions as those of the imagined economy, arguing that net returns on a long position is identical between the two economies, and that all investors in the imagined economy hold the market portfolio while engaging in borrowing and lending at the risk free rate to satisfy their risk preferences. Based on these arguments, the CAPM in the imagined economy can be expressed as a CAPM in net returns with illiquidity costs.

The liquidity adjusted CAPM in net return of security i thereby becomes:

$$E_t\left(r_{t+1}^i - c_{t+1}^i\right) = r^f + \lambda_t \frac{\operatorname{cov}_t\left(r_{t+1}^i - c_{t+1}^i, r_{t+1}^M - c_{t+1}^M\right)}{\operatorname{var}_t\left(r_{t+1}^M - c_{t+1}^M\right)} \tag{7}$$

Where $\lambda = E_t(r^M_{t+1} - c^M_{t+1} - r^f)$ is the risk premium.

Equation 7 can be expanded into the conditional expected gross return:

$$\begin{aligned} E_t\left(r^i_{t+1}\right) = & r^f + E_t\left(c^i_{t+1}\right) + \lambda_t \frac{\operatorname{cov}_t\left(r^i_{t+1}, r^M_{t+1}\right)}{\operatorname{var}_t\left(r^M_{t+1} - c^M_{t+1}\right)} + \lambda_t \frac{\operatorname{cov}_t\left(c^i_{t+1}, c^M_{t+1}\right)}{\operatorname{var}_t\left(r^M_{t+1} - c^M_{t+1}\right)} \\ & - \lambda_t \frac{\operatorname{cov}_t\left(r^i_{t+1}, c^M_{t+1}\right)}{\operatorname{var}_t\left(r^M_{t+1} - c^M_{t+1}\right)} - \lambda_t \frac{\operatorname{cov}_t\left(c^i_{t+1}, r^M_{t+1}\right)}{\operatorname{var}_t\left(r^M_{t+1} - c^M_{t+1}\right)} \end{aligned} \tag{8}$$

Simplified to

$$E_t\left(r^i_{t+1}\right) = r^f + E_t\left(c^i_{t+1}\right) + \lambda_t\beta_1 + \lambda_t\beta_2 - \lambda_t\beta_3 - \lambda_t\beta_4 \tag{9}$$

Where $E_t(c^i_{t+1})$ is the expected relative illiquidity cost. The four betas will be presented below.

Sources of liquidity risk

From equation 9 the authors show that the LCAPM consists of the regular market beta and three betas each representing different sources of liquidity risk, for which investors require a return premium:

β_1 : The standard market beta, captures the relation between stock returns relative to market returns.

β_2 : The liquidity risk associated with the covariance between liquidity of a security, c^i, with market liquidity c^m, $cov(c^i, c^m)$. This beta provides an estimate of the extent to which the liquidity of the stock, depends on the liquidity of the markets. This beta is usually positive, as one should expect the liquidity of a security to drop when the market liquidity is dropping. This phenomenon is called *commonality in liquidity.* With this return premium, the authors argue that investors will require a premium for buying securities that are illiquid during an illiquid market period. The authors find a return premium to be 0.08%.

β_3 : The liquidity risk introduced by the covariance between security returns, r^i, and market illiquidity, $cov(r^i, c^m)$. This return premium is argued to exist as investors pay a premium for securities with high expected returns during drops in the market liquidity. Therefore, this beta tends to be negative. The return premium is estimated to be 0.16%. The existence of the return premium is supported by Pastor and Stambaugh (2003) findings focusing on the return reversals caused by order flow.

β_4 : The liquidity risk associated with the covariance between security liquidity and market returns, $cov(c^i, r^m)$. The intuition behind this return premium is that investors are willing to pay a premium for liquid securities when the market returns drop. As such, risk averse investors prefer assets with stable liquidity costs during market drop. This premium was not studied before the publication of Acharya and Pedersen paper in 2005. The return premium due to the covariance between the illiquidity of securities with market returns is found to be 0.82%, and seem to be the most important of the three illiquidity betas.

These relations are illustrated in the beta matrix below.

	r^m	c^m
r^i	β_1	β_3
c^i	β_4	β_2

Unconditional Liquidity Adjusted CAPM

As liquidity, empirically, has been shown to be persistent and time-varying (Amihud (2002), Pastor and Stambaugh (2003)), the LCAPM model shows that returns are predictable. The persistence in liquidity conflicts with an assumption of independence between dividends and illiquidity costs. Therefore, to derive an unconditional function to estimate the liquidity-adjusted CAPM, the authors instead assume a constant conditional covariance of innovations in illiquidity and returns. This assumption results in the following unconditional functions.

$$E\left(r_t^i - r_t^f\right) = E\left(c_t^i\right) + \lambda\beta^{1i} + \lambda\beta^{2i} - \lambda\beta^{3i} - \lambda\beta^{4i} \tag{10}$$

where

$$\beta^{1i} = \frac{\operatorname{cov}\left(r_t^i, r_t^M - E_{t-1}\left(r_t^M\right)\right)}{\operatorname{var}\left(r_t^M - E_{t-1}\left(r_t^M\right) - \left[c_t^M - E_{t-1}\left(c_t^M\right)\right]\right)} \tag{11}$$

$$\beta^{2i} = \frac{\text{cov}\left(c_t^i - E_{t-1}\left(c_t^i\right), c_t^M - E_{t-1}\left(c_t^M\right)\right)}{\text{var}\left(r_t^M - E_{t-1}\left(r_t^M\right) - \left[c_t^M - E_{t-1}\left(c_t^M\right)\right]\right)} \tag{12}$$

$$\beta^{3i} = \frac{\text{cov}\left(r_t^i, c_t^M - E_{t-1}\left(c_t^M\right)\right)}{\text{var}\left(r_t^M - E_{t-1}\left(r_t^M\right) - \left[c_t^M - E_{t-1}\left(c_t^M\right)\right]\right)} \tag{13}$$

$$\beta^{4i} = \frac{\text{cov}\left(c_t^i - E_{t-1}\left(c_t^i\right), r_t^M - E_{t-1}\left(r_t^M\right)\right)}{\text{var}\left(r_t^M - E_{t-1}\left(r_t^M\right) - \left[c_t^M - E_{t-1}\left(c_t^M\right)\right]\right)} \tag{14}$$

They find the average illiquidity, $E(c_t^i)$, for a portfolio by taking the average using the entire time-series of illiquidity observations. The return and illiquidity innovations are elaborated on in the following section.

Empirical results

AP considers all common shares from CRSP listed on NYSE and AMEX from July 1962 to December 1999. As a measure of illiquidity, the authors utilize the Amihud (2002) measure defined in equation 1.

The authors form sets of 25 test portfolios sorted on illiquidity, variation in illiquidity, size and book-to-market by size. Following equations 10 to 14, the authors run cross-sectional regressions under a number of different model specifications. First, for comparison to the CAPM, they present a single lambda applied to a net beta. To distinguish between the empirical goodness of fit of the standard CAPM with a market beta and the liquidity adjusted CAPM while both exploit the same degrees of freedom, the authors define the net beta as:

$$\beta_p^{net} = \beta_p^1 + \beta_p^2 - \beta_p^3 - \beta_p^4 \tag{15}$$

By defining this single variable, the CAPM and LCAPM are more comparable, as the LCAPM is not simply improved by adding factors, rather it is improved by the liquidity adjustment. The standard CAPM results in an R^2 of 0.653, while the liquidity adjusted CAPM results in an R^2 of 0.732, with a statistically significant parameter for β_p^{net} at a 1% alpha level.

Second, four separate lambdas associated with each of the liquidity betas are introduced to determine where the liquidity effects are stronger. By considering these results the authors mention the multicollinearity problem inherently present in the model. The authors find that illiquid securities, securities with a high average c^i, display high liquidity risk, measured as $cov(c^i, c^m)$, $cov(r^i, c^m)$ and $cov(c^i, r^m)$. These findings provided evidence of the theory of 'flight-to-liquidity'. AP proposed that flight-to-liquidity is a result of investors selling what they perceive as the less liquid investments, and turning towards more liquid investments instead. Consider a financial crisis where market liquidity dries up and stocks diverge due to excess sales pressure, and subsequently, investment opportunities arise. From the perspective of the pairs-trader, AP's findings imply that the pairs-trader will end up buying illiquid stocks and selling the liquid ones. This is a result of the liquid security increasing in value relative to its counterpart as a result of excess demand, and the illiquid security dropping in value due to excess supply. This proposition suggests that the pairs trader provides liquidity. The pairs trader buys the stock that performs relatively poorly during a crisis, and sells the stock that has done relatively well during a crisis. As an effect, according to AP, the pairs trader earns a premium for liquidity provision. In the following sections we will consider which of the three liquidity effects introduced by AP determine the profits for the pairs trader.

2.4 Gatev, Goetzmann and Rouwenhorst (1999, 2006)

Gatev, Goetzmann and Rouwenhorst (GGR) were the first in documenting pairs trading and making it available to the public. Even though the authors mention that the strategy had been implemented since the 80's, it was not until their publication in 1999 that pairs trading became popular among arbitrageurs willing to take advantage of temporary relative mispricings of stocks. In 1999, they reported the results of the strategy implemented with U.S. markets daily data ranging from 1962 until 1997, and in 2006 they extended their research with data covering the period 1962-2002.

The methodology of the strategy proposed by GGR consists of pairing stocks which experienced similar relative pricing during the formation period, and subsequently trade them, opening long and short positions when the normalized prices diverge by two standard deviations from their historical spread. The authors define a formation period of twelve months in which they calculate the sum of squared differences in the normalized price series for each pair of stocks (including reinvested dividends). To be included in the formation period, the stock has to satisfy the requirements of having zero days without trading volume during the previous twelve months. This filter helps to alleviate some concerns about the possibility to trade highly illiquid stocks.

After identifying the pairs with the minimum distance in their normalized price series, the five pairs with the lowest SSD are assigned to the portfolio *top 5*, and with the same criteria they form the portfolio *top 20* and *top 110-120* (top 20 pairs after the top 100). Then, they begin to trade the pairs during a period of six months, opening positions when the spread exceeds two standard deviations from the historical measure, and closing them when the prices converge, when one of the stocks in the pair is delisted, or at the end of the trading period. The whole process is repeated every month, so at each point in time there six overlapping portfolios, mimicking the structure of an institutional investor with different portfolio managers.

The excess returns of the strategy are reported as what they call *committed capital* and *employed capital* (or *fully-invested capital*). The first considers the number of pairs that are selected for trading, while the second calculates the returns based on how many pairs are actually traded during the period. The difference in these methods arises when one (or more than one) pair never diverges enough to trigger a trading signal. Therefore, the committed capital approach still considers the cost of committing capital even though a trade is not executed, while the fully-invested approach calculates returns based on capital actually invested in trading positions.

They further distinguish between returns based on positions that open (closed) on the same day of the trading signal (*no waiting*), and returns calculated from opening positions on the day following the signal (*1 day waiting*). The returns from the *no waiting* approach may be biased upwards due to the bid-ask bounce, so they also report the returns from trading one day after a trading signal.

In their publication from 1999, they find statistically and economically significant returns both for the *committed capital* and the *fully-invested* portfolios, with positions initiated on the day of the signal and on the following day. Similar conclusions arise in the paper published in 2006, in which they extend their previous study. In 1999 they reported the returns over 6-months periods, while in 2006 they reported returns over 1-month periods. To make their results comparable, we report the returns from both papers compounded to 1-year. Despite the circulation of the paper published in 1999, the returns from the pairs trading strategy did not diminish significantly for the committed capital portfolios, and even increased for the fully-invested portfolios.

Portfolio	Top 5		Top 20	
Publication	’99 paper	’06 paper	’99 paper	’06 paper
No-wait				
Fully-invested	0.1238	0.1688	0.1249	0.1866
Committed Capital	0.1232	0.0982	0.1238	0.1010
Wait-one-day				
Fully-invested	0.0754	0.0932	0.0816	0.1128
Committed Capital	0.0750	0.0570	0.0808	0.0642

The authors argue that the profits are not caused by mean reversion, but to the presence of a latent risk factor in the returns, different from conventional risk measures. They arrive to this conclusion after analyzing the returns up until and after 1988, which consider the year that marks the rise in the hedge fund industry. The average monthly return for the top 20 strategy drops from 118 bps per month to around 38 bps. However, after adjusting for the Fama-French-Momentum-Reversal factors, the risk-adjusted returns for each of the periods remain significantly positive. GGR argues that pairs trading returns are a compensation for the risk arbitrageurs bear when enforcing the *Law of one price*.

They employ some robustness checks to analyze pairs trading results under different circumstances. First, the returns presented before come from the strategy implemented with unrestricted pairs, so any stock can be paired with any other that minimizes the sum of squared distances in their normalized prices. This approach is entirely mechanical, and may lead to pairing stocks which do not have much in common nor share common factor exposures. Therefore, they restrict the pairs to stock that belong to the same industry group within the grouping proposed by Standard and Poor’s: Utilities, Transportation, Financials and Industrials. After this change, the returns are somewhat lower but still both statistically and economically significant.

Secondly, they argue that the returns from pairs trading do not simply come from mean reversion. To support this statement, they perform a bootstrap analysis to compare the performance of pairs trading implemented following their methodology, and the results to forming pairs with a random selection of stocks. They start by defining pairs and obtaining the dates in which each pair would start a position, but instead of then trading the stocks from the pair, they replace them with two other stocks with similar returns in the previous month. If pairs trading is simply a reversal strategy, by performing the bootstrap several times, one could expect to get on average, similar returns to those from pairs trading. However, they find that the returns from the random pairs are negative and more volatile than those from the true pairs, concluding that the strategy is not solely benefiting from mean reversion.

Thirdly, given that the strategy may not be possible to implement due to short-selling constraints, especially with small capitalization stocks, the authors restrict the analysis by only considering stocks that belong to the top 30% market capitalization and they find that the returns remain mostly unchanged and highly significant for all portfolios. Additionally, they test if the strategy may be affected by short recalls so they simulate recalls on the short positions on days with high volume and they find that the returns are slightly affected but remain positive and highly significant.

Finally, they also analyze if pairs trading works just because of the exposure to bankruptcy risk. As stocks with an increasing probability of bankruptcy would normally go down in price, if pairs trading is simply profiting from the non-realization of a bankruptcy from the long-end of the pairs, the returns would be expected to come from the long positions. However, GGR show that much of the returns come from the short positions.

2.5 Do and Faff (2010, 2012)

Do and Faff (2010) analyze pairs trading in U.S. markets for the period 1962-2009 and provide further evidence of the decreasing trend in the profitability of the strategy. Their argument is that the main the driver for the lower returns is a higher arbitrage risk caused by fundamental, noise-trader and synchronization risks. Therefore, they disagree with GGR's hypothesis that this phenomenon is produced by an increased competition to arbitrage away these opportunities.

Fundamental risks refers to the increased disruption in the relationship between stock prices, which leads to non-convergent pairs. This disruption could be because the condition of the business of one (or both) of the companies changes substantially, and this is reflected in the price during the trading period. The noise-trader risks alludes to irrational trading, such as market bubbles or panic selling, which works against the traders betting on a closing of the spread. Finally, synchronization risk refers to the uncertainty about when the market will react to a potential mispricing of assets, which could not happen during the time the arbitrageur holds a position.

They decompose the returns from pairs trading by periods, sectors and the frequency of convergence of the pairs, and they find that most of the profitable pairs are those with the highest number of zero-crossings (ZC), defined as the number of times the normalized spread series changes sign (convergence). They argue that restricting pairing to stocks within the same industry and choosing those with the lowest SSD during the formation period (like in GGR), but then sorting again by the number of ZC, produces portfolios with superior returns to those proposed by GGR.

As a continuation of their paper published in 2010, in 2012 Do and Faff investigated if pairs trading is still profitable after accounting for trading costs. They argue that the existence of pairs trading returns is not by itself an argument against market inefficiency because after transaction costs, the returns from all portfolios considered vanish or even become negative. As pairs trading requires the execution of several trades because the pairs may diverge and converge many times, one must take into consideration the cost of applying this strategy.

They form 29 different portfolios based on sectors, industries, sum of squared deviations and zero crossings, and they focus on estimating transaction costs for the period 1963-2009 and decomposing the costs into three groups: commissions, market impact and short selling fees. To estimate the commissions, they focus on Jones' (2002) work in which he presents a time series of commissions paid by all investors from 1925 until 2000. These costs are most likely above the commissions paid by institutional traders, so Do and Faff apply a 20% discount to the reported costs while citing other papers that support their estimations. For the period 2001-

2009 they use the quarterly reports of institutional trading costs from Investment Technology Group (ITG). The values reported range from 70 basis points (bps) in 1963 to as low as 9 bps in 2009, for an average of 34 bps for the entire period. Therefore, commissions on trades have dramatically decreased in the period studied.

To estimate the market impact of pairs trading, the authors measure the price movements which occur after the divergence signals, defined as two standard deviations from the historical spread in the normalized price series. They analyze the spreads from one day prior to divergence until two days after to draw conclusions about how arbitrageurs affect prices when they seize to exploit these opportunities. Do and Faff argue that we should expect a gradual narrowing in the spreads after the divergence when market participants see the current spread as an opportunity. Then, to arbitrage away the divergence, the cheap stock is expected to appreciate with the inflow of new buy orders, while the expensive one is expected to fall in price with the new selling orders. They report that on average the spread declines from 7.56% to 7.02% and 6.78% one and two days after divergence, respectively. These figures imply that an average market impact of 26 bps in the following two days after divergence signals for the entire period 1963-2009.

Additionally, they mention that short positions in pairs trading are subject loan fees and short selling constraints in the form of recalls or the inability to short certain stocks. For loans fees, they use a constant 1% yearly fee payable during the period in which a trader holds a short position. Regarding short selling constraints, they mention that around 84% of the stocks considered (or 99% of the total market value) can be shorted, and that historically only around 2% of the short positions are recalled.

Do and Faff show that all portfolios considered have positive monthly returns of 93 bps on average, and in all cases the returns are statistically significant. However, after adjusting for the estimated costs of trading, the average monthly returns amongst all portfolios goes down to 12 bps, with some portfolios even showing negative returns. This leads to the conclusion that, pairs trading is only profitable before accounting for all trading costs, but after trading costs most of the returns disappear.

2.6 Engelberg, Gao and Jagannathan (2009)

After a modified replication of the methodology of GGR, Engelberg et al. (2009) attempt to answer the questions of why some selected pairs perform better than others, what causes the spread of normalized prices between pairs to diverge from one another and if the reason for divergence affects subsequent returns and convergence horizon.

The authors find that pairs trading returns are driven by events with impact on both the long and short positions. Their analysis is focusing on news published around the date of divergence. The returns of pairs trading are found to be partly driven by a difference in liquidity level, and by differences in the responses to i.e. industry specific shocks. This suggests that by focusing on pairs where one stock is slower at reacting to news common to both stocks in the pair, one can improve pairs trading returns. They further find that returns tend to be smaller once sell-side analysts take an interest in the stocks, and when institutional owners hold a large proportion of stocks. In summary this analysis shows that the pairs trading return have two primary drivers: liquidity provision and lagged reactions to information affecting both companies in the pairs.

Furthermore, they analyze how the reason for initial divergence affects profits. They find that returns are higher for pairs that diverge due to news that temporarily affect the liquidity of one of the stocks in the pair. On the other hand, returns tend to be lower for news that have a more fundamental and persistent relevance for a given stock. The intuition here is clear, if a stock in a pair suddenly changes fundamentally, there is a lower probability that the two stocks will converge again. As liquidity was found as an important factor, the authors combined the Pastor and Stambaugh (2003)'s liquidity risk factor with the Carhart four factor model, and found that pairs trading returns has a significant exposure to the liquidity risk factor, regardless of the holding period tested. They can, however, not explain all pairs trading returns, which are both economically significant, ranging from 0.8% to 1.9% per month depending on holding period, and statistically significant at a 1% alpha level. It should be noted that the authors do not account for transaction costs. Conclusively, they find that liquidity provision and price discovery contribute about equally to explaining the abnormal returns in pairs trading.

Finally, the authors analyze the convergence horizon. They found that size and liquidity seem to affect which pairs converge faster than others. Small, illiquid and volatile stocks tend to converge faster than the stable large and liquid securities. Furthermore, they found that pairs trading returns can be improved by focusing on the holding period. Specifically, they found that pairs that do not converge during the first ten days of the initial signal have lower risk adjusted returns than pairs held for six months.

3 Theory

3.1 Market neutrality

This section seeks to inform the reader of the implications of market neutrality and how it relates to a pairs trading portfolio. We consider the pairs trading strategy to fall into the category of statistical arbitrage strategies. Statistical arbitrage strategies are commonly characterized by a short-term holding period and broadly diversified portfolios, and it falls into the category of market neutral strategies because it takes long positions and short positions simultaneously. This method seeks to take advantage of mispricings in securities with high correlations, which is the main focus of a pairs trading strategy.

A market neutral portfolio is defined as a portfolio that is independent of market fluctuations, which is commonly measured as the correlation between the market return and the market neutral portfolio. Different methods of obtaining a market neutral portfolio can be utilized, and we will in this section focus on three forms of neutrality: dollar neutrality, beta neutrality and sector neutrality. Needless to say, 'market neutral' is a very liberal use of the word as the trader is not immune to all kinds of fluctuations in the markets simply by having zero dollar, beta or sector exposure, and in that sense being risk free. It simply means that the portfolio is not subject to market risk, which the capital asset pricing model (CAPM) assumes is the only risk factor that the speculator is compensated for. While this assumption is widely challenged in academia and subject to endless testing, it is now empirically fair to assume that investors can earn risk premiums from exposure to other risk factors. We will touch upon these risk factors in section 3.2.

Dollar neutral

The straight forward approach to market neutrality, which is utilized by the pairs trader in a GGR (2006) and Do and Faff (2010, 2012) setup, is monetary neutrality. If you go long \$1 in stock A, you simultaneously go short \$1 in stock B.

$$\$_{Long} = \$_{Short}$$

As such you obtain a net investment of 0. This method does not account for any historic patterns in the relationship between the fluctuations in the stock prices of A and B. The dollar neutral portfolio is in theory self-financing, as the proceeds from borrowing the stock and short selling it can be used to buy another stock. This makes a dollar neutral strategy suitable for empirical research. In practice, this notion of self-financing is however overly optimistic due to, amongst others, short-selling restrictions - in particular on the small capitalization segment of stocks, as short selling was a scarce opportunity in the early sample used for this study. Moreover, short selling most often comes with margin requirements. Margin requirements impose the necessity of collateral posted by the short seller to add a buffer for the stock lender, in the case of the short seller going bust. Margin restrictions are ignored in this study, but it does affects the extent of potential implementation for the prospective pairs trader.

One should note that being dollar neutral does not imply that the investor is necessarily independent of market fluctuations as the risk profile on the long and short position may differ when the betas of stock A and stock B are not the same.

Beta neutral

Beta neutral is in a CAPM framework a risk free strategy, and can therefore not earn anything north of the risk free rate. The idea is to have a weighting scheme that ensures the same amount of market risk is taken on both the long and the short leg. As a concrete case, if we buy stock A, that has a beta of 2, and short stock B, that has a beta of 1, our beta neutral portfolio consists of \$-2 short in stock B and \$1 long in stock A. This example illustrates that we take a smaller position in the more risky stock A, while taking a larger position in the less risky stock B to compensate for the lower market risk. As such, our weighting allows for betas on the long side and short side to be equal.

$$\beta_{Long} = \beta_{Short}$$

It is evident from the example that the strategy is not dollar neutral, and thereby in theory not self-financing. It is however theoretically market neutral. This method is commonly applied by practitioners, yet it is liable to the choice of beta estimate. Beta can be estimated in an unlimited amount of ways. Should one use daily, weekly, monthly, quarterly or annual return data? Should one use a half year estimation period, one year, two years or five years? As the underlying company changes over time, so does the stock's beta. Therefore, to keep the portfolio beta neutral, one must frequently adjust the portfolio weighting.

Sector neutral

Arguably, a stock should be more affected by shocks to the sector it belongs than by changes in the overall market. For example, technology stocks suffered greater losses than utilities stocks during the dot-com bubble. Sector neutrality is an attempt to immunize the portfolio to sector specific shocks. Again, for the sector neutral trader, he can select between being dollar neutral and beta neutral to a given sector.

Benefits of market neutrality

The question is then, why would an investor spend resources attempting to be neutral to an endless amount of risk factors?

The answer is simple. By being long and short at the same time, you avoid having any directional bias. In other words, you do not bet on the direction of the market, you instead bet on the relative performance of stock A in relation to stock B. If you have reason to believe stock A is overvalued, and B is undervalued, one can - in theory - obtain risk free profits by buying the undervalued stock and selling the overvalued counterpart. As such market neutrality is beneficial for investment strategies exploiting small market inefficiency and anomalies.

Market efficiency

The question then becomes, are markets actually efficient? Fama (1970) first introduced the concept of an 'efficient' market defining it as a market reflecting all available information. This definition implies that one cannot exploit any mispricings, as markets will always reflect the true value of a security. Three forms of market efficiency were defined. First, we have the weak-form efficiency. It implies that security prices reflect all information about past prices. One cannot use past prices to predict future prices. Second, we have semi-strong efficiency. It implies that security prices reflect all publicly available information, and only inside information can generate abnormal returns. Third, we have strong-form efficiency. This third level implies that all publicly available information as well as inside information is priced. Note that the pairs trader is betting on markets not being weak-form efficient as the pairs trader is using past price information to form pairs and generate signals. Should the pairs trader generate abnormal returns, it would be evidence of an inefficient market, breaching all three levels of efficiency.

3.2 Factor models

One cannot discuss market efficiency without taking the choice of equilibrium model into account. In order to claim that one generates abnormal returns, one must prove that returns are not only compensation for taking exposure to certain risk factors. We will in this section present most commonly used risk factors, as these will be used repeatedly throughout this study.

Fama and French (1993) introduced two additional factors to assist the CAPM market risk premium in explaining stock returns. The first factor constructs portfolios based on size, as it buys the stocks with small market capitalization and shorts the large market capitalization, therefore the name Small-minus-Big, or *SMB*. This factor captures the, at the time, empirical anomaly of small capitalization stocks outperforming large capitalization stocks. The second factor they introduced is the High-minus-Low factor, or *HML*. It is another statistical anomaly, where one buys stocks with a high book value relative to market value (B/M ratio), also denoted value stocks, and short sell stocks with a low B/M ratio, also denoted growth stocks. The rationale is that first with a high B/M ratio are "cheap", in the sense that you pay less for

$1 of the company. Whereas growth stocks, as the name indicates, are priced high, based on expectations of future growth which in turn is reflected in the market value of the company.

In the same category of commonly used return factors, we find the Momentum factor, introduced in 1993 by Jegadeesh and Titman. The authors find that by buying recent winners and selling recent losers, the investor generates abnormal returns. These findings have lead to much empirical research and they hold across asset classes and across subperiods. Yet some evidence suggests that the momentum outperformance relative to the market is not caused by systematic exposure to a risk factor, but rather by delayed reactions to firm-specific news. While one could come up with a convincing story as to why investors should require a premium for holding the small illiquid stock, relative to holding large stocks, it is hard to tell an equally convincing story as to why the momentum investor should earn a premium. While not considered a risk premium, it is often included in portfolio return regressions, and we shall therefore consider this factor in our analysis as well.

3.3 Fama Macbeth Regression

When testing the CAPM in cross-sectional tests, one encounters some statistical issues. The standard OLS regression framework assumes that observations are independent. However, cross sectional correlation amongst portfolios would result in less informative results than the OLS implies. The implications would result in the OLS standard errors becoming too low, and consequently, one would obtain to many false positives, also known as type 1 errors. Fama and Macbeth (1973), hereafter FMB, present a possible solution to this issue.

The cross sectional test in a FMB regression is split into 3 steps for in-sample beta estimation. Acharya and Pedersen (2005) construct cross sectional regressions using pre-estimated betas. While they do not implement the Fama and Macbeth approach, they point out that the FMB procedure would yield the exact same results. As we will present our results using the FMB procedure in this study, we will describe the procedure as applicable to our methodology after providing the framework in the three steps below:

- Step 1: Compute the time-series betas for each N portfolio.

$$E(r_{p,t}) - r_f = \alpha_p + \beta_p F_t + \epsilon_{t,p}, \qquad \forall_p \{1, ..., P\} \tag{16}$$

- Step 2: For each time period t, regress N returns on the N betas derived from step 1.

$$E(r_{p,t}) - r_f = \gamma_0 + \lambda \hat{\beta}_p + u \tag{17}$$

- Step 3: Compute averages to find Fama Macbeth estimators of λ and α obtained in step 2. From the cross sectional regressions in step 2, we get a times series of least squares alpha and lambda estimators: $[\hat{\alpha}_{t=1}, , ..., \hat{\alpha}_{t=T}]$, the risk premium is defined as

$$\hat{\lambda}_{\text{FMacB}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_t \tag{18}$$

$$SE(\hat{\lambda}_{\text{FMacB}}) = \frac{\sigma(\hat{\lambda}_t)}{\sqrt{T}} \tag{19}$$

In this paper, we use the liquidity risk beta formulas proposed by AP from equations 11 to 14 to complete step 1.

Step 2 can be a cumbersome one, as you need to run t cross sectional regressions. We can however enjoy the benefit of estimating only one liquidity beta through time, and do not have a rolling beta estimation window. This means that our betas are constant through time, and we can take a shortcut to step 3. Since we are applying a linear framework, the order in which we take the average and run the regression does not make a difference. The average of regression coefficients is the same as the regression coefficients of averages. We can therefore take the average off all portfolio returns considered in our analysis and run a single cross sectional regression of these averages on the betas estimated in step 1.

As such, for the pending cross sectional regressions presented in section 5.6.2, we estimate variations of the following regressions:

$$\begin{pmatrix} r_{p=1,avg} \\ r_{p=2,avg} \\ \dots \\ r_{p=P,avg} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\beta}_{p=1,1} & \hat{\beta}_{p=1,2} & \hat{\beta}_{p=1,3} & \hat{\beta}_{p=1,4} \\ 1 & \hat{\beta}_{p=2,1} & \hat{\beta}_{p=2,2} & \hat{\beta}_{p=2,3} & \hat{\beta}_{p=2,4} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \hat{\beta}_{p=P,1} & \hat{\beta}_{p=P,2} & \hat{\beta}_{p=P,3} & \hat{\beta}_{p=P,4} \end{pmatrix} \begin{pmatrix} \alpha \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} + \begin{pmatrix} e_{p=1} \\ e_{p=2} \\ \dots \\ e_{p=P} \end{pmatrix} \tag{20}$$

Where $r_{p=1,avg}$ is the average portfolio return, and $\lambda_1 - \lambda_4$ represents a market or liquidity risk premium. The cross sectional standard errors of each lambda is the same as the cross sectional standard errors, except they are missing a correction for the fact that the betas are generated regressors, a so called 'Shanken correction'.

4 Methodology

4.1 Data

The sample used in this paper comprises all stocks in the CRSP database. We employ daily price and volume data from August 1962 until December 2019. We select all common shares, denoted by CRSP share codes 10 and 11, listed on NYSE and AMEX. While Nasdaq stocks are included in GGR's study, our need for volume data to calculate illiquidity measures leads us to discard the companies listed on this exchange. Nasdaq stocks only have volume data since 1982 and it includes inter-dealer trades Anderson and Dyl (2005), which can lead to unreliable comparisons in illiquidity measures with stocks listed on the other exchanges. Therefore, our analysis is based only on stocks listed on NYSE and AMEX, which also allows us to apply the methodology from Acharya and Pedersen (2005) to analyze pairs trading returns as a compensation for taking liquidity risk.

Adjusted share prices are calculated using CRSP variables *PRC* and *CFACPR*, where *PRC* is the close price of day t for stock i, and *CFACPR* is the cumulative factor for adjusting prices for dividends and stock splits.

$$P_t^{i,adj} = \frac{PRC_t^i}{CFACPR_t^i} \tag{21}$$

Subsequently, we calculate daily stock returns as:

$$r_t^i = \frac{P_t^{i,adj}}{P_{t-1}^{i,adj}} - 1 \tag{22}$$

We filter out data with missing price information. Stock delistings are handled following the methodology of Acharya and Pedersen (2005). We maintain delisting returns of -100% and any CRSP delisting codes of 500 (reason unavailable), 520 (went to OTC), 551-573 and 580 (various reasons), 574 (bankruptcy) and 584 (does not meet exchange financial guidelines) are assigned a return of -30%. This figure is the average delisting return estimated by Shumway (1997), who examined the OTC returns of delisted stocks and found that a negative return of 30% was the average delisting return of all delisted companies included in the CRSP database. This adjustment functions as a mitigation against survivorship bias in our results.

We obtain the time series of the Fama-French 3-factors and the momentum factor from professor Kenneth French website[1].

4.2 Formation and trading of pairs

In this section, we explain how the formation and trading periods are defined, and how we calculate the returns, following the approach of GGR (1999; 2006) and Do and Faff(2010; 2012). In section 5, we explain the results obtained and compare them with those of the aforementioned authors.

Formation period

Every month, we form pairs over a 12 month look-back period. Subsequently, we trade the pairs over the following 6 months. Both the length of the formation period and the length of the trading period are arbitrarily chosen by GGR and Do and Faff. One could attempt to

[1] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

improve results by choosing an alternative formation and trading period, but this is not within the scope of our study. An illustration of how the pairs are formed and then traded can be found in figure 1.

Figure 1: Portfolio Construction

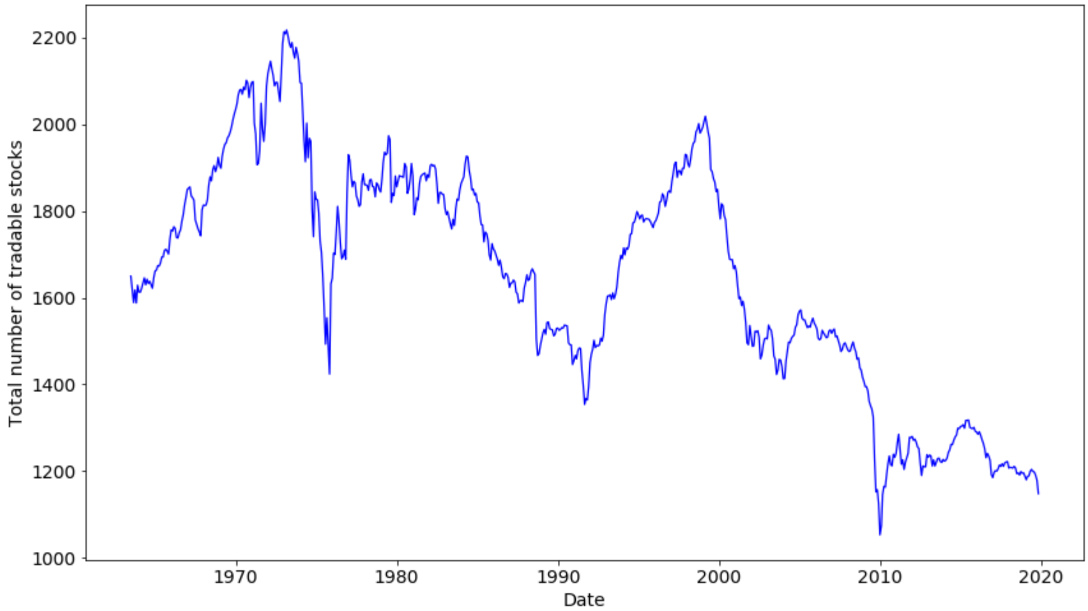
This figure illustrates the formation and trading periods of the strategy implemented in this study. Pairs are formed over 12 months and traded over the subsequent 6 months. The first portfolio of pairs starts trading at the beginning of month 13, and every month a new portfolio starts its trading period. At the end of month 18, the first portfolio stops trading and from then on, six different portfolios are traded every month.

We start each formation period by screening out all stocks with one or more days of zero volume trading, indicated by CRSP's variable *VOL*. We further apply prices filters in line with those used by Acharya and Pedersen (2005), screening out stocks with share prices below \$5 and above \$1000, since returns on low-price stocks are highly affected by the minimum tick size of \$1/8 for NYSE stocks effective until 1997. Figure 2 shows the monthly number of stocks living up to our criteria, and it can be seen that each month there are at least 1000 stocks that pass all the filters, and in some months this figures goes above 2000 stocks.

For each stock, we normalize our stock prices by constructing a total return index at the beginning of the formation period, until the end of the trading period. For a given period with N stocks in our sample, we obtain $N * (N-1)/2$ possible pairs. We divide by two because if stock A and B are be a pair, then stock B and A cannot be a pair. We calculate the sum of squared deviations between each pair based on the spread between the two normalized price series, denoted as $\tilde{P}^i$ and $\tilde{P}^j$ for stock i and j, respectively. Lastly, the pairs are ranked by their

Figure 2: Tradable stocks

This figure illustrates the number of stocks that fulfil all the requirements of data availability, minimum and maximum price. The highest number of tradable stocks is 2218 in March 1973, and the lowest is 1053 in January 2010, after the financial when many companies disappeared or got delisted from public markets.



SSD measure, and those with the lowest values are considered the *top n* pairs.

$$SSD_{i,j} = \sum_{t=1}^{T} \left(\tilde{P}_t^i - \tilde{P}_t^j \right)^2 \tag{23}$$

Trading period

We calculate the historical mean $\bar{P}$, and standard deviation $\hat{\sigma}$, for each pair spread during the formation period. Each pair consists of stock i and stock j. We then standardize our spreads during the trading period to obtain daily Z-scores.

$$Z_t^{i,j} = \frac{\left(\tilde{P}_t^i - \tilde{P}_t^j \right) - \bar{P}^{i,j}}{\hat{\sigma}^{i,j}} \tag{24}$$

If the spread moves more than two historical standard deviations away from its historical mean, we buy the underperforming stock, and short-sell the outperforming stock. We unwind our position when the two normalized price series cross again, when a stock is delisted or on the last day of the trading period, whichever occurs first.

The strategy is then based on the premise that stocks that historically moved together should continue to do so in the future. Should this argument hold, the pairs trader earns a profit when the stocks converge. GGR test a strategy based on executing a trade immediately after receiving a signal, and a strategy based on waiting one day before executing a trade. The latter being the more conservative method. The immediate signal trade implies that the trader can buy and sell the stocks at the respective close prices on the day of the signal. While it is possible with intra-day data, this study will only consider the more conservative approach of waiting a day before taking a position. The method of waiting one day was introduced by GGR to alleviate the concern of an upward bias in the returns induced by a potential bid-ask bounce or trading halt.

While the two standard deviation move may be appropriate for a majority of the pairs, some pairs may follow each other so closely that two standard deviation moves are rather negligible in magnitude. These scenarios are therefore not attractive enough for the pairs trader seeking compensation higher than the bid-ask bounce and transaction costs paid for the trade. This issue is addressed by Do and Faff (2010), who point out that pairs trading portfolios may include two share classes of the same company. To avoid trading pairs of the same company listed as separate share classes, we keep only one unique company identifier signified by CRSP's *PERMCO* variable. While this alleviates the problem to some extent, one should still be aware that sometimes even pairs of stocks of different companies may not be ideal for pairs trading when the trading signals are obtained with small price deviations that do not justify the transaction costs of entering the trade.

Figure 3 shows an example of the trading period of two stocks, consistent with figure 1 of GGR. The stocks represented are Kennecot and Uniroyal, and they form one of the top pairs selected by our algorithm for the first trading period. We obtain similar results to those of GGR, yet for us the pair was ranked 6th based on the SSD measure, while GGR finds it to be ranked 5th in that same period. However, the pattern of opening and closing positions is similar.

Figure 3: Daily normalized prices: Kennecot and Uniroyal (Pair 6)

This figure illustrates an example of the trading period of the pair formed by Kennecot and Uniyoral stocks. The full lines represent the normalized prices of each stock, while the dashed green lines represent the positions taken in the pair. When the normalized prices diverge by more than 2 historical standard deviations, a position in the pair is taken until the normalized prices converge or until the end of the trading period. In this example, we open and close positions several times during the trading period, so it is a profitable pair.

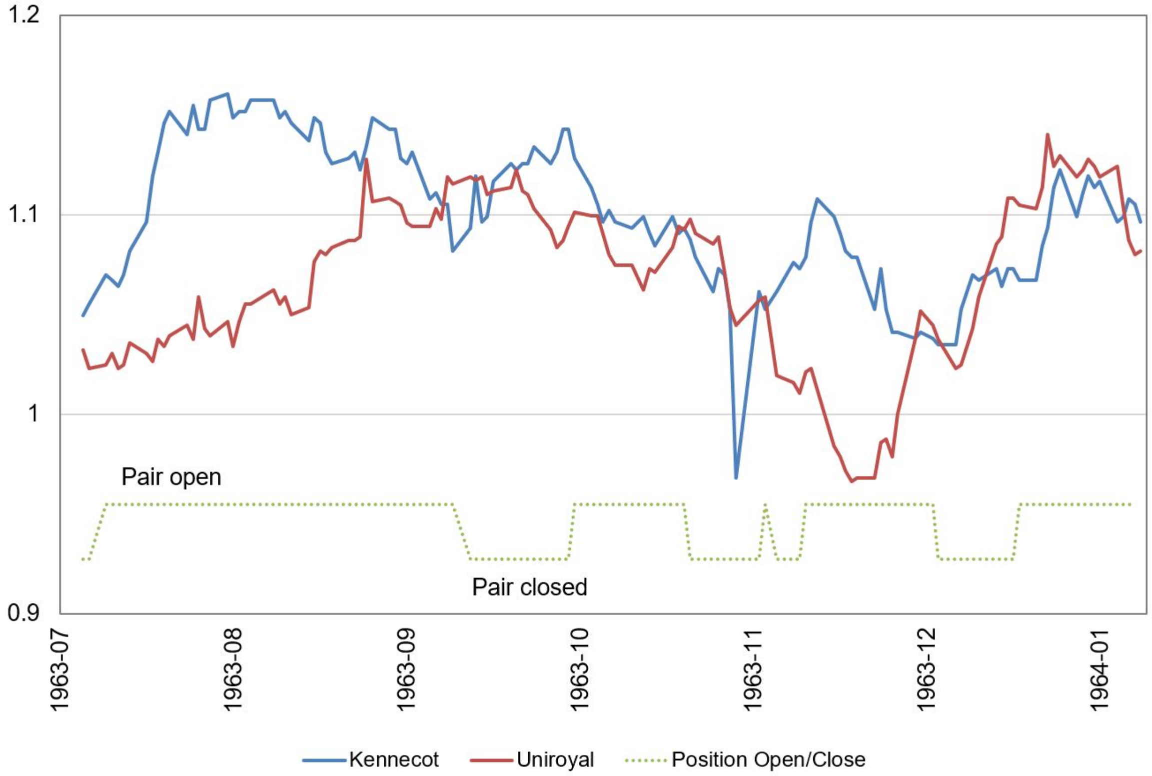
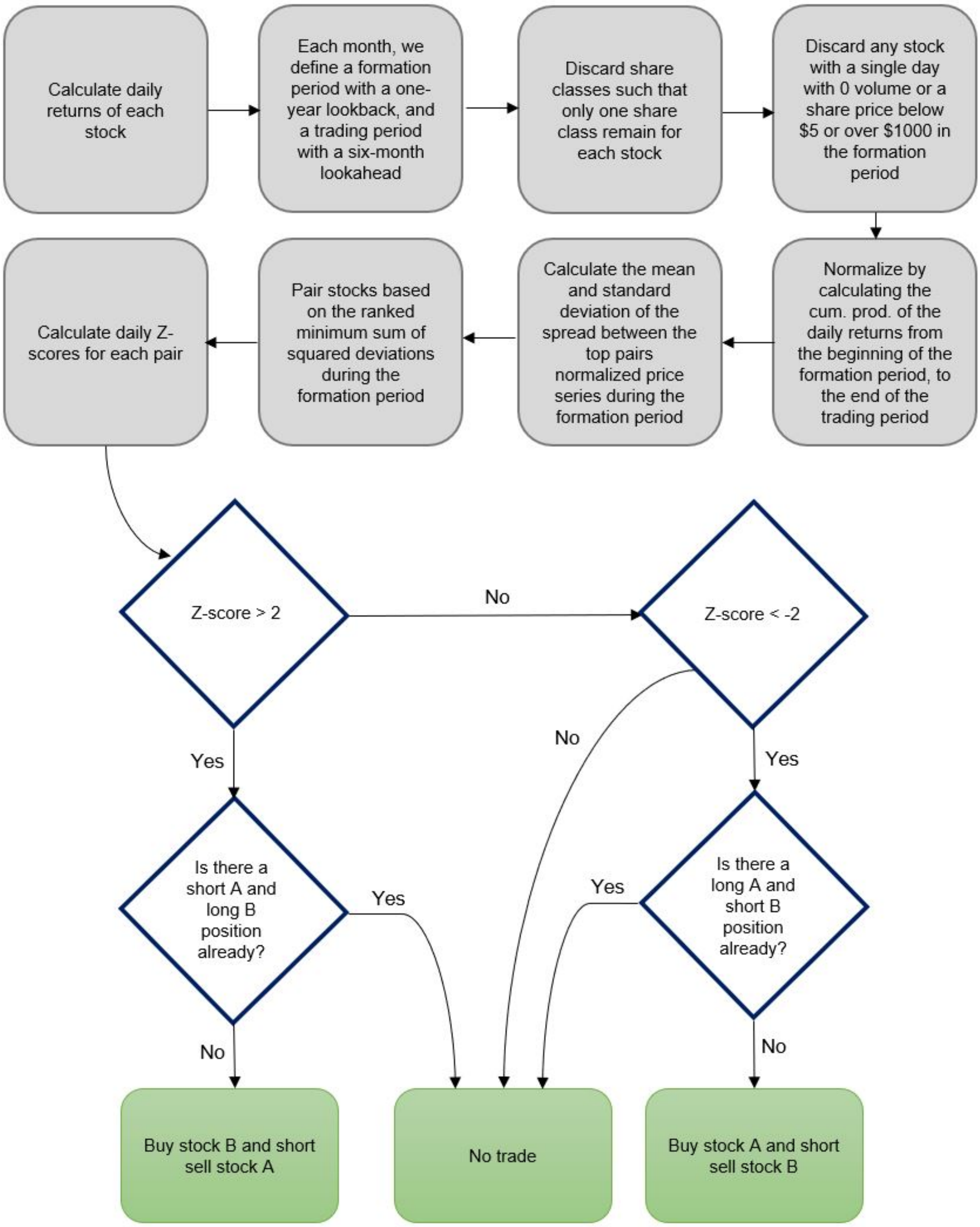


Figure 4 shows a flowchart explaining the trading strategy. It indicates how we form pairs, and whether or not we should place a trade each day for any given pair.

Figure 4: Flowchart of the trading strategy



4.2.1 Excess Return Computation

The strategy implemented in this thesis is trading intensive and it requires buying and selling stocks, sometimes every day. Therefore, the computation of the returns is a non-trivial issue, as there are positive and negative returns every day for all the different portfolios trading at the same time. We do not rebalance the long and short exposures of each pair despite that they may move significantly when positions are open. This could lead to higher-than-ideal exposures to overall market movements for a long-short strategy, but it reduces the impact of the trading costs. We will discuss this trade-off in further detail in section 6.

The daily return r, at time t of each pair P is calculated as the value-weighted returns of the long (L) and short (S) positions in the following way,

$$r_{P,t} = \frac{w_{L,t} \cdot r_{L,t} + w_{S,t} \cdot r_{S,t}}{w_{L,t} + w_{S,t}} \tag{25}$$

$$w_{i,t} = w_{i,t-1} \left(1 + r_{i,t-1}\right) = \left(1 + r_{i,1}\right) \cdots \left(1 + r_{i,t-1}\right) \tag{26}$$

where w is the weight of stock the position on the stock i which includes all the returns since the beginning of the trading period. On day 1 of the trading period, all weights are 1 since we consider long and short positions of 1 dollar. The returns on the short positions are considered with respect to the pairs trader, which means that the return is positive when the stock goes down in price, and vice-versa when the stock appreciates.

To make our results comparable to those of the existing literature, we follow the approach taken by GGR (1999; 2006) and Do and Faff (2010; 2012), who consider two different measures of excess return: the return on committed capital and on actual employed capital (or fully-invested return). The difference between the two approaches was already mentioned in section 2.4. Committed capital considers the number of pairs that are selected for trading, while fully-invested calculates the returns based on how many pairs are actually traded in the trading period. Both methods should give the same returns when all pairs have open positions but the the differences are evident when one (or more) pair never diverges enough to trigger a trading signal. Then, the committed capital approach still considers the cost of the money set aside for trading even though a trade is not executed, while the employed capital approach calculates returns based only on capital actually invested in trading positions.

The committed capital approach has the advantage of a simpler calculation of returns but it also implies a possible under-utilization of resources as there is a cost of opportunity for the pairs trader that has to commit capital to pairs that do not trade. The disadvantage of the committed capital is remedied in the fully-invested approach, which only takes into account the capital invested in active positions.

Related research

This study is not attempting to manipulate the original GGR recipe for generating pairs. While this recipe seems logical, one could question the robustness of the methodology applied, as it is by no means derived academically. Instead of challenging these choices ourselves, we present a non-exhaustive list of developments in academic literature related to statistical arbitrage and pairs trading.

GGR is by far the most cited paper in the pairs trading domain. The simplicity of the algorithm tested on U.S. equities has led to much related and more recent research in the area and on different markets and asset classes. Many apply more sophisticated methods of isolating co-moving stocks. We will therefore in this brief section attempt to break the different approaches into a few selected buckets to have an overview of various attempts to improve pairs trading results. Some honorable mentions of other approaches are:

- **Distance approach:** The sum of squared deviation as used in GGR's recipe for selecting pairs, and applied by Do and Faff as well. But other distance measures could also be used to identify co-moving stocks. This method benefits from a simplistic approach of generating signals through nonparametric thresholds.
- **Cointegration:** Another notable method is cointegration tests. During the formation period, a cointegration test is applied to obtain econometrically more reliable equilibrium relationships between co-moving stocks. These approaches differ in the formation period,

while most follow a GGR threshold rule in the trading period. Caldeira and Moura (2013) apply this approach on Sao Paulo stock exchange from 2005 to 2012. The modified strategy presents returns before transaction costs of 16.38% per year and a Sharpe ratio of 1.34.

- **Time series:** In this approach one ignores the formation period. Instead, emphasis is focused on the trading period, where one attempts to model the spread as a time series process, i.e. a mean reverting spread over time. Signals are thereby generated through predictions of the spread. Cummins and Bucca (2012) apply this approach on Energy futures and find annualized returns before transaction costs of over 18% per year.

- **Machine Learning and PCA:** These approaches have limited supporting literature. Machine learning combined with forecasting approaches has been studies by Huck (2009, 2010) on U.S. S&P 100 from 1992-2006. Huck takes multiple spreads and return forecasts based on bivariate information sets using an Electre III method to obtain rankings of all assets in terms of expected returns. He then buys the first ranked (under-valued) asset, and short the lowest ranked (over-valued) asset. The PCA approach has been implemented by Avellaneda and Lee (2010). They consider statistical arbitrage strategies in U.S. equities from 1997 to 2007, and generate their signals using PCA as well as sector ETFs. They work with the idiosyncratic component of stock returns and model the residuals as mean-reverting processes to generate their contrarian trading signals. The PCA approach generated a Sharpe ratio 1.44, while the ETF approach generated a Sharpe ratio of 1.1.

Many noteworthy attempts to improve the GGR algorithms are proposed by prior studies. Engelberg et al. (2009) propose closing positions if convergence does not occur after the first 10 days. They show that the profitability is highest during the first couple of days: 25 bps for the first day, and 13 bps for the second, while then slowly fading out over time.

We find that on average, we hold the pair for 36 trading days, or 1.6 months, and a median of 21 trading days, just below one month of trading. Table 1 shows that 5% of all signals converge within the first day. These positions will have a severe impact by the bid-ask bounce.

9% within two days, and 32.7% of pairs converge within the first 10 days.

Table 1: Average time to convergence

Days since signal	Pct of signals converged
1	4.98
2	9.14
3	12.94
4	16.43
5	19.61
6	22.58
7	25.29
8	27.87
9	30.31
10	32.70

We stick to the GGR methodology for the duration of our analysis, but make a note that the methodology is by no means the one and only method of obtaining abnormal returns. We note that the holding period clearly is another arbitrarily selected element to the algorithm, but we do not attempt to implement any improvements.

4.3 Top pairs robustness

One of the first choices to make when analyzing pairs trading is the number of pairs to trade. Gatev et al. (1999, 2006) present their results for top 5, top 20 and top 101-120 (pairs ranked 101 to 120 in lowest SSD). They find that the monthly average return and Sharpe ratios are the highest for the portfolio formed by the top 20 pairs. Comparing the metrics of the top 20 portfolio with the top 5 portfolio, there is not a significant difference in mean returns, but the most noticeable difference is in the standard deviation of returns (1.53% for top-20 vs. 2.10% for top 5). For the top 101-120 portfolio, the mean returns, Sharpe ratio, and standard deviation are similar to those for the top 20 pairs. GGR also reports the top 101-120 as the lowest SSD pairs may be so correlated that even the smallest price divergence could trigger a trading signal, and therefore enter several trades too early. In the case of Do and Faff (2010), they only present their results for the top 20 pairs.

We test the robustness of the arbitrary choice of the number of pairs to include in a portfolio by analyzing the results of selecting portfolios ranked lower than the best pairs. The results shown in this section are for the fully-invested and the committed capital returns, before and after transactions costs. Theoretically speaking, we should expect to observe an inverse U-shape for the Sharpe ratios, as there should be an optimum. However, given the number of pairs in each portfolio, we are subject to idiosyncratic noise.

In figure 5 we present scatter plots of the Sharpe ratios of four different combinations in the number of pairs to include in a portfolio. Each of the four subplots is formed by portfolios comprising 5, 10, 20 or 50 pairs. For example, sub-figure A, is formed by portfolio of top 5, top 6-10, top 11-15, etc., so there is no overlapping in the pairs included in each portfolio. The purpose of this representation is to show that even by varying the number of pairs, there is still a downward trend in Sharpe ratios by trading pairs with higher *SSD*. The dashed red line is the fitted line that minimizes the residuals of all the observations, and in all the different combinations considered, it has a negative slope. The four subplots are for the Sharpe ratios considering fully-invested returns, but the situation is similar for committed capital returns, presented in figure 21 in the Appendix.

In figure 6, we perform a similar analysis for the mean returns before and after transaction costs. In this case we show both fully-invested and committed capital returns in one figure. It is clear from the figure that forming portfolios of 5 pairs produces a noisier plot than for any of the other higher number of pairs. However, it is evident that for any of the options considered, the further away we go from the top pairs, in general we find lower mean returns. In all cases, the difference between the portfolios of top pairs and those of pairs close to rank 1000, is around 10 basis points per month.

Figure 7 presents the mean returns and Sharpe ratios for all possible portfolios of top pairs up to pair 120 [2]. The x-axis denotes the portfolio of the first pair up to pair n. For example, *0-0* represents the portfolio of just the top pair selected every month (we start counting at 0), so if the trader chooses to trade only the top pair, the mean return and Sharpe ratio will be determined by only that pair traded each month; likewise, *0-119* denotes all pairs up to the pairs ranked 119th, so the portfolio formed each month will contain 120 different pairs.

[2]We perform this analysis up to pair 120 due to computing power limitations as we are working with data frames of over 100 million data points. Each iteration requires the calculation of returns each day (from 1963-08 to 2019-12) for each pair (2 stocks each), for each of the six portfolios trading at the same time.

From sub-figure A, it is evident that the mean returns before and after transaction costs do not change significantly by adding more pairs to the portfolio. In terms of Sharpe ratios, portfolios with more pairs seem to perform better, probably due to similar mean returns and the benefit of a lower standard deviation by adding more stocks, but this benefit is reduced by higher transaction costs, as for any portfolio size the Sharpe ratio is around zero.

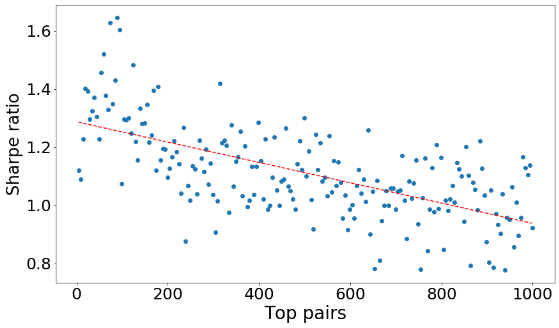
These findings suggest a decreasing profitability of the pairs trading portfolios when trading pairs ranked low in the SSD metric. For example, we can see in figure 5 that the Sharpe ratio for top 0-50 is smaller than for top 51-100, but then for top 101-150 the Sharpe ratio drops again to continue in an erratic pattern when we move further away from the best pairs. Do and Faff (2010) indicate that a low SSD does not imply that a pair is potentially more profitable than another pair with higher SSD, because the pairs trader benefits when the stocks performs a round trip (divergence and convergence), but if they just move together without crossing or they are too close substitutes, their divergences might not be enough to trigger trading signals. As theory suggests, we find a slight degree of an inverse U-shape amongst the top portfolios, whereas the remaining portfolios for top pairs 200 and forward to be relatively flat stable around an average Sharpe ratio. These findings suggest that the top portfolio may not be the optimal portfolio to trade, and one should instead focus on stocks further down the ranks.

As mentioned, we find some benefits from not trading only the “best” pairs. It would appear that one should include pairs of stocks that may not be so correlated during the formation period. These findings are in line with Do and Faff (2012) who suggest removing the pairs with the lowest SSD. However, these potential benefits come with the cost of higher transaction costs which translate in little to no improvement when including more pairs to the portfolio, and there is no clear evidence in our findings indicating that there is an optimal rank of the pairs to trade.

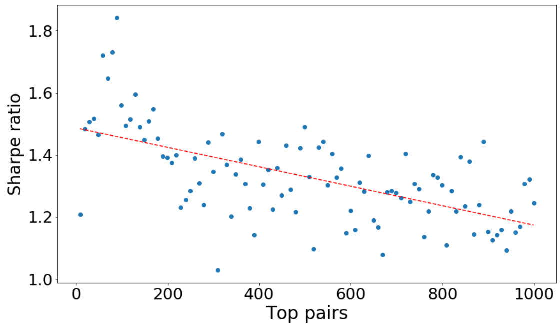
For these reasons, and in order to be able to compare our conclusions to those of the existing literature, in the rest of the paper we will only perform our analysis based on the top 20 pairs as Do and Faff. We expect the top 5 pairs, as used by GGR, to be subject to little diversification and too much idiosyncratic noise. Further research could be done in analyzing the optimal number of pairs and in which SSD ranks these pairs should be, but this procedure requires significant data mining which is not the purpose of our thesis.

Figure 5: Sharpe Ratios of Top-1000 pairs divided in sub-portfolios - Fully-Invested

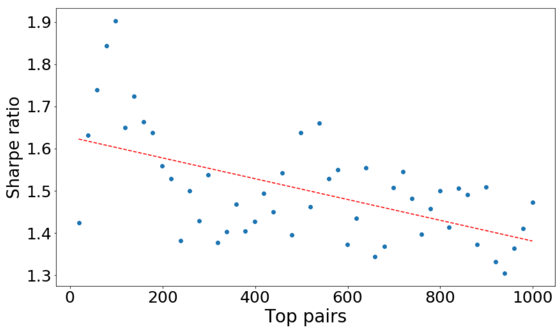
This figure shows the annualized Sharpe ratios of the fully-invested returns for different combinations of the number of pairs in each portfolio, up to the pair ranked 1000 in the *SSD* metric, with pair 0 being the pair with the lowest *SSD* during the formation period. Sub-figure A, shows the Sharpe ratios of portfolios comprised of 5 pairs each, with the left-most blue dot representing the metric for trading only the first 5 pairs with the lowest SSD, and the right-most blue dot indicating the Sharpe ratio of the portfolio which only trades pairs ranked between the 996th to 1000th in the SSD metric. The same explanation applies to the other sub-figures, which show the results of forming portfolios of 10, 20 and 50 pairs. The dashed red line is the fitted line that minimizes the residuals of all the observations.



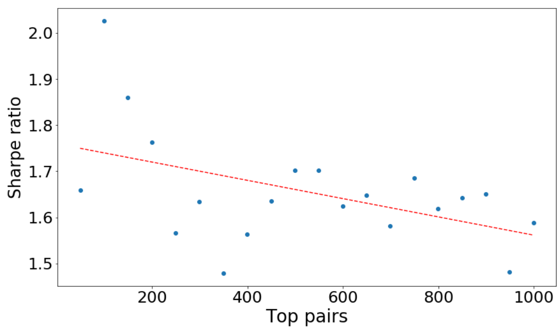
(a) Steps of 5 pairs



(b) Steps of 10 pairs



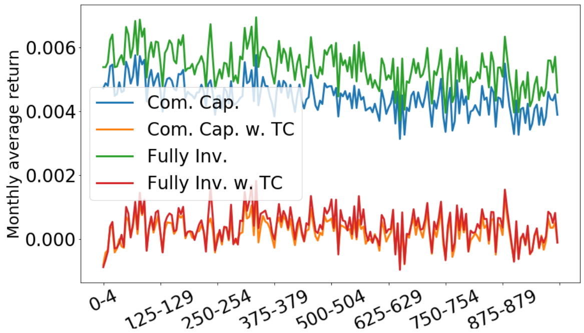
(c) Steps of 20 pairs



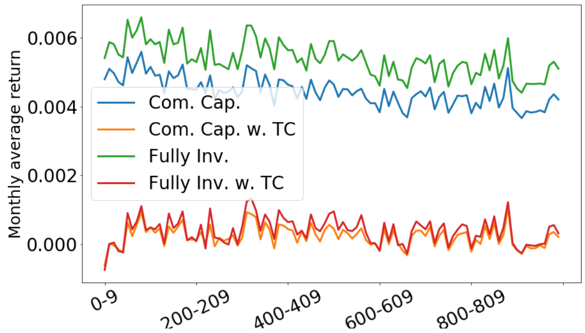
(d) Steps of 50 pairs

Figure 6: Mean Returns of Top-1000 pairs divided in sub-portfolios

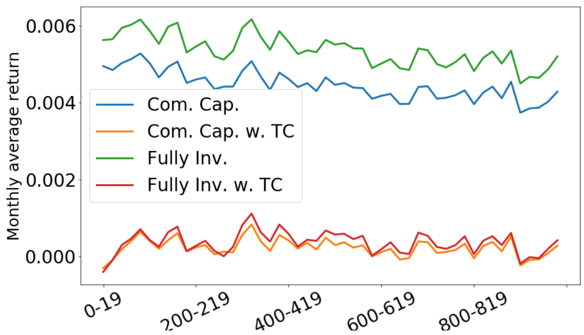
This figure shows the monthly mean returns of the fully-invested and committed capital approaches, before and after transaction costs, for different combinations of the number of pairs in each portfolio, up to the pair ranked 1000 in the SSD metric, with pair 1 being the pair with the lowest SSD during the formation period. Sub-figure A, shows the mean returns of portfolios comprised of 5 pairs each, for which to the left of the X-axis we find the portfolios of pairs with the lowest SSD (pairs 0-4), and in the right we find those pairs ranked lower in SSD metric. The same explanation applies to the other sub-figures, which show the results of forming portfolios of 10, 20 and 50 pairs.



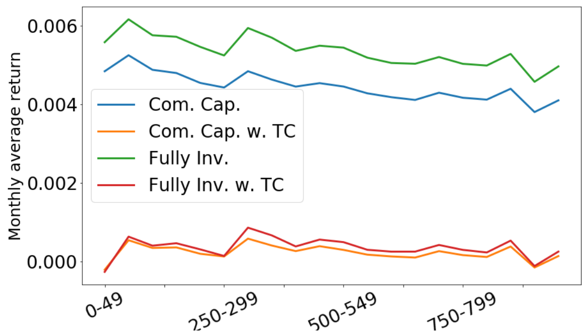
(a) Steps of 5 pairs



(b) Steps of 10 pairs



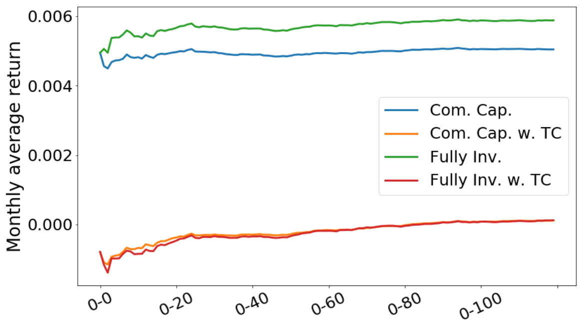
(c) Steps of 20 pairs



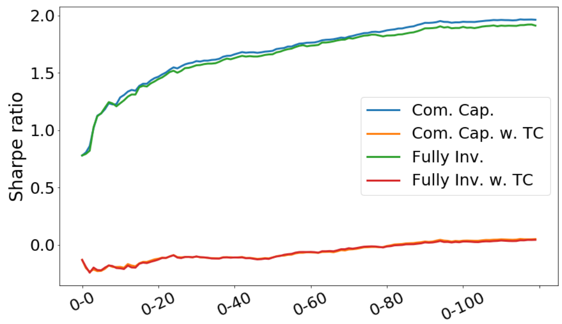
(d) Steps of 50 pairs

Figure 7: Means and Sharpe ratios for all combinations of top 120 pairs

This figure shows the monthly mean returns and the annualized Sharpe ratios of the fully-invested and committed capital approaches, before and after transaction costs, for portfolios of increasing number of pairs, up to the pair ranked 120 in the SSD metric, with pair 1 being the pair with the lowest SSD during the formation period. The x-axis indicates which pairs are included in the portfolio, so the further we go to the right of the x-axis, the more pairs are included. For example, *0-0* represents the portfolio of just the top pair selected every month (we start counting at 0), so if the trader chooses to trade only the top pair, the mean return and Sharpe ratio will be determined by only that pair traded each month; likewise, *0-119* denotes all pairs up to the pair ranked 119, so the portfolio formed each month contains 120 different pairs.



(a) Monthly mean returns



(b) Annual Sharpe ratios

4.4 Long/Short Excess returns

Given that the pairs trading strategy is not necessarily actively trading all days of the month, calculating excess returns on the short and the long leg is not as straight forward as simply subtracting (adding) the monthly equal weighted market return from the monthly return of long (short) leg of the pairs trading portfolio. This method is dubious as we may trade the long portfolio during the first 15 days of the month, but if the market goes down by 10% in the second half of the month, we are subtracting a full month return from a half month return, and our results would be biased.

The problem scenario above could be countered using daily equal-weighted market returns from CRSP, but here we have a new set of challenges. Canina et al. (1998) warn about the use of daily CRSP equal-weighted index returns. When comparing the mean return of the monthly rebalanced equal-weighted monthly return from CRSP, and the daily rebalanced equal-weighted accumulated monthly return, they find a monthly return difference of 0.43% or 6%

per year. Daily rebalancing in a market portfolio requires a fair share of transaction costs, that are not accounted for in the index, and replicating these returns are therefore next to impossible. Moreover, the authors find that daily autocorrelation and bid-ask bounces account for more than half the monthly variation between the two calculation methods. The daily equal weighted return is therefore a very noisy return series. We attempt to approximate the monthly rebalanced daily return by adjusting the daily equally-weighted return as follows.

We invest the same notional amount in both our long or short leg, as we do in the market. We then consider four points in time. At time A we have the start of the month, at time B we have the opening of a position, at time C we have a closing of the position and finally at time D we have end of month. Our excess return for a long leg position is then calculated as the daily equal weighted return between time B and time C, divided by the daily equal weighted return between time A and D, multiplied by the monthly rebalanced equal-weighted return provided by CRSP. With this approach, we scale down the daily rebalanced return to approximate the monthly rebalanced return. This method is naturally only valid if the monthly rebalanced and daily rebalanced return series are correlated on a monthly basis, and we find the correlation to be 0.9996. As such we can derive some approximated excess returns on our pairs trading long and short portfolio. We will discuss this method in section 6, and how it impacts our results.

Our analysis of the long and short leg will be following the GGR (2006) methodology. Performance is considered before and after transaction costs, but not in excess of a market return. Instead GGR consider it as excess of a monthly T-bill rate. While they do not explicitly state why they take returns in excess of the risk free rate, we assume that it breaks down to the same complications as stated above. Likewise, we will analyze how our long/short exposure has changed over time, to reflect upon whether daily rebalancing might be optimal given a potentially large exposure to market movements.

4.5 Transactions costs

Given that pairs trading is a trading intensive strategy, one must consider the costs incurred when implementing it. Nowadays, the costs of trading are much lower than they were in the 60's, 70's and 80's, but so are the returns of pairs trading, as found by Do and Faff (2010).

Therefore, in this section we analyze pairs trading by taking transaction costs into account. For that matter, we focus on Do and Faff (2012) who analyzed if pairs trading is still profitable after transaction costs, and in section 5.3 we show the results of applying these costs to the pairs trading strategy.

Commissions

To execute a trade, one must pay a commission that may include the compensation to the broker, exchange fees, among others. Usually these commissions depend on the size of the trade, so a simple way to calculate their impact in a trading strategy is to estimate a percentage of the total amount to be invested.

Do and Faff (2012) estimated the one-way commissions that an institutional investor would have paid for the period 1963-2009. They base their estimations on the findings of Jones (2002), who presents a time series of commissions paid by all investors from 1925 until 2000. These costs are most likely above the commissions paid by institutional traders who have access to better deals than retail investors, so Do and Faff apply a 20% discount to costs estimated by Jones, while citing other papers that support their estimations. The values reported range from 70 basis points (bps) in 1963 to as low as 9 bps in 2009, for an average of 34 bps for the entire period. As our study goes until 2019, we estimate the commissions for the years not covered by Do and Faff by linearly decreasing the commissions they estimate for 2009 (9 basis points), to reach 3 basis points in 2019.

Market impact

Market impact refers to the hypothetical effect in the markets produced by the application of a trading strategy. When there is an opportunity to make profits, most likely, market participants will identify it and subsequently try to exploit it. While doing so, the opportunity will gradually or rapidly fade away. In the case of a pairs trader, when he sees that a pair has diverged by a certain amount that justifies taking a position in the pair. It is likely that other traders will observe the divergence as well and the prices of the stocks will be affected by the new buy or sell orders. This may cause the depreciated stock to increase in price because of the increased demand (buy orders) and the “expensive” stock to lose value with the increased offer (sell orders). In a short amount of time, the arbitrage disappears.

Do and Faff estimate the market impact of pairs trading by measuring the price movements which occur after the divergence signals, defined as two standard deviations from the historical spread in the normalized price series. They analyze the spreads from one day prior to divergence until two days after to draw conclusions about how arbitrageurs affect prices when they seize to exploit these opportunities. Do and Faff argue that one should expect a gradual narrowing in the spreads after the divergence, because the long is expected to appreciate with the inflow of new buy orders, while the short leg is expected to fall in price given the new short selling orders. They report that, on average, the spread declines from 7.56% to 7.02% and 6.78% one and two days after divergence, respectively. These figures imply an average market impact of 26 basis points in the following two days after divergence signals for the entire period 1963-2009. In our thesis, we use Do and Faff’s estimation of market impact of 21 and 9 basis points for long positions before and after 1988, respectively. For short positions, we use their estimate of 10 and 7 basis points for the years before and after 1988, respectively.

Short selling fees

Pairs trading implies short-selling one of the stocks for each of the pairs in which the trader takes a position. These stocks have to be borrowed in order to be sold, and this has a cost for the trader, as the broker will normally charge a percentage fee of the amount borrowed. Do and Faff consider a constant 1% loan fee per year payable over the life of each trade. They mention that this fee might be conservative as it is slightly higher than those estimated in other studies. For our analysis we consider a 1% short selling fee in 1963 and a linearly decreasing fee to reach 20 basis points in 2019. This cost can be more accurately estimated but it also depends on the agreements between the trader and his broker, so it is hard to define a number that can be used by all traders. Additionally, one could also consider trading single stock futures, instead of the stock itself, which could be cheaper but it is not a product available for all stocks.

We assume that we can short-sell any stock but this may not necessarily be true in practice, especially during a crises, and particularly for small capitalization stocks. However, given that on average we take short positions in stocks belonging to deciles 7 to 8 (see figure 12), the short-selling constraints should not be a significant concern.

4.6 Barriers to arbitrage

We hypothesize that the pairs trader earns a liquidity premium when barriers to arbitrage are high. To get a better idea about whether this hypothesis holds, we attempt to isolate periods of time where the liquidity premium could be particularly high. When barriers to arbitrage are high, we hypothesize that market participants display a lack of risk willingness due to lower liquidity in the market or because they are unable to trade due to limited funding liquidity. When funding liquidity and market liquidity are low, we expect that the pairs trader requires a higher premium for providing liquidity in these particularly illiquid sub-periods, relative to the premium required during more stable sub-periods. Our results will therefore stand in contrast to those presented by Do and Faff and GGR.

In this section we will explain how market liquidity and funding liquidity are defined, and briefly introduce the findings of leading academic research in the area. Furthermore, we will explain how we approximate market and funding liquidity in our analysis.

4.6.1 Market liquidity

We attempt to isolate low market liquidity months by using the implied volatility index on the S&P 500 (VIX). One could make use of the Amihud (2002) measure as a proxy for market illiquidity by taking the average *ILLIQ* measure for each stock in each period. We hypothesize that when the VIX spikes, one should expect to obtain more signals due to an increased chance of divergence. This methodology follows that of Chung and Chuwonganant (2014), who find that VIX exerts a large market-wide impact on liquidity and on individual asset liquidity. The authors find that the effect of VIX on stock liquidity is greater than the combined effects of all other common determinants of stock liquidity. Therefore, we argue that a spiking VIX will likely result in lower risk appetite for arbitrageurs which could present an opportunity of a higher liquidity provision premiums for the pairs trader. We consider monthly percentage changes in the VIX, and test various monthly thresholds for these changes to see if large volatility shocks, both negative and positive, result in significant contemporaneous and subsequent gains for the pairs trader. Concretely, we consider percentage changes month over month at the following 5 percentile thresholds to capture drops in the VIX: 0.05, 0.10, 0.15, 0.20, and 0.25. Moreover, we consider 0.75, 0.80, 0.85, 0.90, and 0.95 to capture large surges in the VIX. We consider this method to be appropriate since it isolates the large shocks to the market liquidity, as opposed to selecting an arbitrary percentage threshold of the VIX. Selecting an specific threshold for the VIX could also be considered, but is left for further research.

The analysis considers both the lagged and non-lagged change in the VIX. With the lagged version, we hypothesize that a large spike in the VIX leads to an increased chance of divergence in our portfolios. We test whether this could result in our pairs converging in the following period, assuming the VIX does not continue to surge. A further increase in the VIX could, however, result in our pairs diverging even further. Profits are expected when the VIX reverse back to normal due to an increased risk appetite of arbitrageurs and increased market liquidity. With the non-lagged version, we attempt to detect any patterns as to how the pairs trader is affected by contemporaneous changes in the VIX. Using the same logic as for the lagged version, we expect the pairs trader makes profits when the VIX drops a lot, while he would lose money when the VIX surges, as pairs are expected to diverge. As we may be subject to our pairs trading portfolios not trading for weeks, we analyze whether we have any zero-trading days, and if so, we would have to adjust the VIX return to reflect the time we were actually in the market. We find a total of 19 days with no position in our pairs trading portfolios. This is out of a sample of 677 months. None of the zero-days occur during the same month. We therefore do not consider this an issue, and will leave it as an insignificant issue in our analysis.

The higher probability of divergence with an increasing VIX is countered by a simultaneous increment in the correlation amongst securities. One could suggest further research to pair our approach with a further disaggregation of the VIX by considering an index of implied correlation instead. An argument could be made for pairs trading to be more profitable when implied correlations are low, as one should expect to receive more trading signals.

The VIX time series starts from 1990, and our analysis will thus be reflecting the period of time, that by Do and Faff is deemed particularly unattractive to the pairs trader, as it generates negative pairs trading returns after transaction costs.

4.6.2 Funding liquidity

When markets are already illiquid, they are highly sensitive to changes in funding conditions. Ibbotson (1999) finds that the median market betas of brokers and speculators are in excess of 1. This serves as evidence of why they find that when markets are dropping, capital constraints are more likely, and it explains why liquidity dry-ups often coincide with drops in the market. Brunnermeier and Pedersen (2007) (BP) find that speculator funding is a key driver of the market liquidity. Their findings are based on evidence of two separate but mutually reinforcing liquidity spirals. BP first define a ‘margin spiral’, which occurs if margins goes up when market illiquidity increases. A margin spiral could be explained with the following scenario: markets are dropping, and market participants lose wealth; when market participants lose wealth, market liquidity dries up. Given Ibbotson’s findings, this subsequently leads to a tightening of speculators’ funding constraints. As such, we have a resulting margin spiral. BP then introduce the ‘loss spiral’. A loss spiral occurs when speculators hold positions that exhibit negative correlation with customers’ demand shocks. When funding shocks occur, market illiquidity increases, speculators suffer losses on their positions, they are forced to sell more and they push prices even further down. This increases market illiquidity, and we then have a ‘loss spiral’. The margin spiral and loss spiral reinforce each other.

BP consider the accessibility of capital and determinants of margin requirement. Both of which affect market participants’ risk willingness. They find that in times of liquidity crisis, margins increase, and exogenous shocks to speculator capital lead to a reduction in market liquidity. Furthermore they find that the relation between speculator capital and market liquidity is non-linear. When speculators are far from their constraints, the impact on market liquidity is limited. When they are close to their capital constraint, the impact on market liquidity is substantial. These effects are found to be greater for illiquid securities.

The findings of BP collide with our hypothesis of pairs trading profits. In times of low funding liquidity, liquidity spirals should in theory increase the liquidity provision premium for the pairs trader. In the same way as the VIX functions as a proxy for the market liquidity, we use changes in the yield spread of Eurodollars over T-bills (TED spread) to isolate periods with large shocks to funding liquidity. There is no unified way of measuring funding liquidity, but we follow the methodology of Asness, Moskowitz and Pedersen (2013) who uses, amongst others, the TED-spread as a proxy for funding liquidity.

The TED spread is available from 1986 and forward. It will therefore have slightly more observations than the VIX-based analysis. Thresholds for the TED spread are selected in the same way as for the VIX. Resulting thresholds are shown in table 2, where we see that the VIX and TED spread changes span rather symmetrically from -6.56% to 6.37% and -0.3% to 0.29% respectively. While we do not expect to see much better results using the 0.25 and 0.75 percentiles, this method does allow for more visually pleasing presentation of the incremental changes and put the more interesting edge cases of 0.05 and 0.95 into perspective.

Table 2: VIX TED thresholds

Percentile	Δ TED pct	Δ VIX pct
0.05	−0.30	−6.56
0.10	−0.20	−4.10
0.15	−0.14	−3.24
0.20	−0.11	−2.75
0.25	−0.08	−2.19
0.75	0.06	1.80
0.80	0.09	2.31
0.85	0.12	3.50
0.90	0.19	4.83
0.95	0.29	6.37

The VIX and TED spread change is calculated from end of month *t-1* to end of month *t* (a full month change). Therefore the pairs trading portfolio must be active throughout the entire month, to actually reflect returns affected by these changes in the VIX. When checking for non-active days in our portfolio, we find a total of 18 days with no positions in our 677 months sample.

As introduced in section 4.4, both the excess mean returns and Sharpe ratios will be provided for both the long and the short leg - only after transaction costs - to test whether either side drive returns during these volatile sub-periods, and if the strategy actually generates substantial returns to offset the transaction costs, that during the full period nullified our pairs trading returns. Furthermore, 5-factor regressions for each percentile are presented for the net pairs trading returns after transaction costs, as well as for the long and short leg after transaction costs. Regressions for the net pairs trading portfolios are specified as:

$$r_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2(Mkt - RF_t) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t} \tag{27}$$

where $r_{net,t}$ is the monthly fully-invested return at time t.

We introduce the *IML* factor to the regression as it should tell us whether the hypothesis of whether or not the pairs trader is compensated for taking liquidity risk. We replicate the *IML* factor following the methodology of Amihud (2014) as explain in section 2.2.2. We can compare our results to the sub-period of 7/1981 - 2012, and we obtain a slightly higher mean return of 0.38% relative to 0.289%. A regression of the *IML* factor on the Carhart four-factor model results in an R^2 of 0.653 compared to the 0.67 reported by Amihud. The same highly significant negative loading on the market with coefficients of -0.2191 compared to -0.226, for *SMB* we find a positive loading of 0.748 compared to 0.769. For *HML* we find 0.306 while Amihud reports 0.321, and finally for momentum we find a negative loading of %0.042 while Amihud reports -0.029%. In summary, we come very close to a replicating factor, and we will implement it for future regressions throughout this study. As explained in section 2.2.2, one could argue that the *SMB* and *IML* factors capture much of the same variation. Amihud tests for this issue, by estimating regression models with *IML* as the dependent variable, and the Carthart four-factor model as the explanatory variables. He finds that *SMB* does explain some of the returns in the *IML* factor, but not all. Amihud therefore argues that the two factors do capture something different. We report the factor correlations below to inform the reader of a slight degree of multicollinearity introduced by including both of these two variables in the regressions. We find a correlation of 0.62 between the *IML* and *SMB* factor.

4.7 Liquidity Adjusted CAPM

Acharya and Pedersen (2005) present empirical evidence consistent with flight to liquidity and the pricing of liquidity risk. The authors form portfolios based on the Amihud illiquidity

Table 3: Factor Correlations

	IML	Mkt-RF	SMB	HML	RF	Mom
IML	1	-0.17	0.62	0.29	0.03	-0.12
Mkt-RF	-0.17	1	0.29	-0.25	-0.08	-0.14
SMB	0.62	0.29	1	-0.19	-0.04	-0.01
HML	0.29	-0.25	-0.19	1	0.08	-0.2
RF	0.03	-0.08	-0.04	0.08	1	0.06
Mom	-0.12	-0.14	-0.01	-0.2	0.06	1

measure; however, this methodology is not suitable in a pairs trading framework. Our initial hypothesis is that the pairs trader provides liquidity in the illiquid stocks, while simultaneously taking liquidity in the liquid stocks. But our portfolio construction is set up in a way that pairs cannot be formed to express this hypothesis in a trade. We therefore base our analysis on portfolios formed by pairs within each SIC industry as done by GGR (2006) and Do and Faff (2010). This method alleviates the problem, and allows for pairing across liquidity levels. While GGR and Do and Faff solely consider four sectors (Utilities, Transportation, Financials and Industrials), we will break the SIC sectors into 10 additional pieces for two purposes: Firstly, we need a larger sample size to run more robust cross sectional regressions. Secondly, from the perspective of the pairs trader, more sectors should result in better results, as you separate stocks into clusters that move closer together.

In this section, we will present our modified replication of their methodology to test whether the pairs trader is compensated for taking liquidity risk, and at what points in time this compensation is the highest.

4.7.1 Acharya and Pedersen replication

In order to construct a liquidity CAPM model, we follow the methodology of Acharya and Pedersen (2005) as described in section 2.3, but with slight modifications to fit the pairs trading portfolios into their framework.

Their market portfolio is formed each month and consists of the sample of stocks with a price above \$5 and below \$1000, and consider only stocks with valid data for volume and returns for at least 15 days in that month. For both returns and illiquidity, their reported results use an equal weighting to define portfolio returns and illiquidity. This method is also used in Amihud (2002). The argument for this choice is that it allows for compensation of an overestimated liquidity level using value-weighting, since stocks are more liquid than many other asset classes represented in the market portfolio, i.e. real estate or corporate bonds. After calculating the Amihud measure for each stock following equation 1, we normalize the illiquidity measure to obtain a stationary measure, c_t^i by

$$c_t^i = min(0.25 + 0.30 ILLIQ_t^i P_{t-1}^M, 30.00) \tag{28}$$

where P_{t-1}^M is the ratio of the total market capitalization of the market portfolio in the prior month relative to the total market capitalization of the market portfolio at the starting point of our analysis in August 1963. This adjustment helps alleviate the stationarity problem as it accounts for inflation. We use the same coefficients 0.25 and 0.3 as AP, as we are considering the exact same sample. These coefficients are selected such that the cross sectional distribution of c_t^i, for portfolios based on size-deciles, has an approximately same level and variance as the half-spread. The half-spread is defined as the difference between the current share price and the midpoint between the bid and the ask. AP base these numbers on Chalmers and Kadlec (1998). The cap of 30% is in place to cleanse for extreme outliers in the ILLIQ measure. This method of normalization puts the illiquidity on a scale corresponding to the cost of a trade.

To obtain portfolio illiquidities we use the sum of the equal-weighted normalized illiquidity

$$c_t^p = \sum_{i \text{ in } p} w_t^{ip} c_t^i \tag{29}$$

To compute the liquidity betas as presented in section 2.3, equation 11 to 14, we estimate the illiquidity innovations by first defining the un-normalized liquidity truncated for outliers as

$$\overline{ILLIQ_t^p} = \sum_{i \text{ in } p} w_t^{ip} \min\left(ILLIQ_t^i, \frac{30.00 - 0.25}{0.30 P_{t-1}^M}\right) \tag{30}$$

The illiquidity innovations for the market and the portfolios can then be predicted by defining an autoregressive process.

$$\begin{aligned}\left(0.25 + 0.30\overline{ILLIQ}_t^M P_{t-1}^M\right) =& a_0 + a_1 \left(0.25 + 0.30\overline{LLIQ}_{t-1}^M P_{t-1}^M\right) \\ &+ a_2 \left(0.25 + 0.30\overline{ILLIQ}_{t-2}^M P_{t-1}^M\right) \\ &+ u_t\end{aligned} \tag{31}$$

The error term from this regression correspond to the market illiquidity innovation series: $u_t = c_t^M - E_{t-1}(c_t^M)$. Innovations in returns and illiquidities are then estimated for each portfolio as, $r_t^p - E_{t-1}(r_t^p)$ and $c_t^p - E_{t-1}(c_t^p)$ respectively, where the E_{t-1} term is estimated as a prediction from a second-order autoregressive model. This method of calculating the liquidity and return surprises functions to alleviate the issue of persistence in liquidity, as presented in section 2.3.

For the market illiquidity series, we estimate an R^2 of 75.3% from the AR(2) process using the same sample period as AP. This is close to AP's estimate of R^2 of 78%. We further estimate the standard deviation of market illiquidity innovations to be 0.13% relative to AP's 0.17%. To test for the continued persistence in liquidity, we check the autocorrelation of the illiquidity innovations and find -0.0368% relative to AP's -0.03. An Augmented Dickey Fuller (ADF) test on the equal weighted illiquidities return a ADF statistic of -2.856 and a p-value of 0.051, so we have a slight degree of stationarity. However, for the illiquidity innovations we find an ADF statistic of -5.88 and a p-value of < 0.0001, we then reject the hypothesis of a unit root in the time series of illiquidity innovations. We present a plot of the market illiquidity innovations in the appendix, as a comparison to AP's Fig. 1. In comparison to the AR(2), we estimate an AR(1) and AR(3). We find that the AR(1) performs worse in terms of resulting autocorrelation of -0.24 compared to AP's -0.29. The AR(3) model results in very limited improvement in the explanatory power of the model, with an R^2 of 75.9%, the R^2 is not reported by AP, but only mentioned to produce little improvement.

Equipped with monthly sector portfolio returns, our illiquidity and return innovations for the market portfolio and the pairs trading portfolios, we estimate the liquidity betas for each portfolio using equations 11 to 14. For each sector portfolio we find the average of monthly illiquidity measures, $E(c_t^i)$, by taking the average of the entire time-series of illiquidity observations.

To examine how liquidity risk affects expected returns, we use a Fama and Macbeth (1973) method, as described in section 3.3. We obtain point estimates in a cross sectional regression that take the pre-estimation of the beta into account. Since the betas are estimated using our entire sample period, they are constant through time. We can therefore enjoy the benefit of Fama and Macbeth's approach by not having to estimate t number of cross sectional regressions, where t is the number of estimation periods in our sample. Instead, we can simply take the mean return for each sector and regress them on the sector specific betas. We present 8 regressions specified in the same way as AP, by considering special cases of the relation presented in section 2.3.

$$E(r^p_{sec,t}) = \alpha + \kappa E(c^p_t) + \lambda_1\beta^p_1 + \lambda_2\beta^p_2 + \lambda_3\beta^p_3 + \lambda_4\beta^p_4 + \lambda\beta^p_{net} \tag{32}$$

Since the liquidity level, $E(c^i_t)$, is an average that does not scale with time, we need to adjust the measure from being paid monthly, to reflect the actual holding period. The holding period is not necessarily one month. In table 4, we find an average holding period in months for each of our sector portfolios.

Table 4: SIC sectors average holding period

Sector	Avg. holding period
Utilities	1.62
Financials	1.88
Transport	2.26
Industrials	1.91
Technology	2.13
Wholesale	2.22
Retail	2.12
Real Estate	2.45
Mining	2.13
Construction	2.32
Consumer Goods	2.13
Materials	2.02
Consumer Services	2.12
Communications	2.05

From table 4, it is evident that the holding period is smallest for the utilities sector and highest for the real estate sector. On an overall average, we are holding pairs for 2.0972 months.

To reflect this as a holding period, we scale our overall $E(c_t^i)$ by $\kappa = 1/2.0972 = 0.4814$.

In an attempt to examine the effect of the liquidity risk $(\beta_2, \beta_3, \beta_4)$, isolated from the liquidity level, $E(c_t^i)$, and the market risk (β_1). Eight regression models are specified as follows:

- **Model 1**:
$$E(r_t^p) - \frac{1}{AHP_{sec}} E(c_t^p) = \alpha + \lambda \beta_{net}^p \tag{33}$$

  In this model, we treat κ as a constant of 1, scaled to adjust for the average holding period for each sector, denoted, AHP_{sec}. This allows us to specify the dependent variable net of the average illiquidity costs.

- **Model 2**:
$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda \beta_{net}^p \tag{34}$$

  In this model, we treat κ as a free parameter.

- **Model 3**:
$$E(r_t^p) = \alpha + \lambda_1 \beta_1^p \tag{35}$$

  This model model represents the baseline capital asset pricing model (CAPM). As such, we treat κ as 0.

- **Model 4**:
$$E(r_t^p) - \frac{1}{AHP_{sec}} E(c_t^p) = \alpha + \lambda_1 \beta_1^p + \lambda \beta_{net}^p \tag{36}$$

  This model again treats κ as a constant of 1, but includes both the market risk premium and the net beta.

- **Model 5**:
$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda \beta_{net}^p \tag{37}$$

  This model considers κ a free parameter, while again including both the market risk premium and the net beta.

- **Model 6**:

$$E(r_t^p) = \alpha + \lambda_1 \beta_1^p + \lambda \beta_{net}^p \tag{38}$$

In this model we set κ to zero, while preserving the market beta and the net beta.

- **Model 7**:

$$E(r_t^p) - \frac{1}{AHP_{sec}} E(c_t^p) = \alpha + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p \tag{39}$$

In this model we introduce the liquidity risk premias and remove the model restriction that $\lambda_1 = \lambda_2 = -\lambda_3 = -\lambda_4$, as given by β_{net}. We note that this model is subject a high degree of multicollinearity as can be seen in the results section, table 19. κ is treated like a constant of value 1.

- **Model 8**:

$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p \tag{40}$$

Finally, we consider each of the liquidity risk betas and the market beta, while keeping κ as a free parameter.

5 Empirical Results

In this section we present our empirical results. The section consists of four parts. First, we show our replication of the seminal papers on pairs trading by GGR (2006) and Do and Faff (2010, 2011), we denote this part: 'Replication'. Second, we present our results concerning what the pairs traders portfolio is loading on, we denote this part 'Loadings'. Third, we present our results from liquidity shock sub-period returns, we denote this third part 'Liquidity Shocks', and finally we present our liquidity CAPM regression model, denoted 'Liquidity Adjusted CAPM'.

PART I - Replication

5.1 Implementation of GGR / Do and Faff methodology

In this section we present our replicated results following the GGR (1999, 2006) pairs trading algorithm. This section functions as a validation of our methodology, before moving on to testing our hypothesis.

GGR (1999; 2006) were the first authors to publish an analysis of pairs trading. Do and Faff (2010; 2012) extended GGR's research until 2009, following the same methodology as GGR, but they included sub-periods and the performance of the strategy when taking transaction costs into account. We validate our pairs formation and subsequent trading methodology, by replicating the results of Do and Faff (2010, 2012) and showing its extension with data until December 2019. Since we are utilizing the CRSP database, which is not a point-in-time database, we can expect some differences between our results and those from the literature on the same topic. Additionally, we are excluding stocks listed in Nasdaq for the reasons explained in section 4.1.

Results from the replication

We implement the pairs trading strategy explained by GGR and Do and Faff and obtain similar results. One can observe the same patterns as those found by GGR and Do and Faff: early sub-periods of high profitability and diminishing returns through the last decades.

In table 5, we summarize the results from implementing the trading strategy following the methodology explained in section 4.2. All the figures presented are for the fully-invested approach except for the last line, which reports the monthly mean return of the committed capital approach. It is evident from the table that pairs trading profitability has decreased significantly over the decades, with a mean return after 2003 of around zero. In the period 1963-1988, for employed capital we obtain an average monthly return of 1.05%, and 0.91% for committed capital, compared to 0.86% and 0.85%, respectively, reported by Do and Faff for the same period. In the next period (1989-2002), the mean returns go down dramatically to about one-third of those in the previous period. We find mean monthly returns of 0.38% for fully-invested and 0.33% for committed capital, while Do and Faff find 0.38% and 0.36%, respectively. For the last period included in Do and Faff's paper, 2003-2009, they find mean monthly returns of 0.24% for both approaches, while we find almost zero in both cases. Finally, taking into account the period 1963-2009 covered by Do and Faff, they find mean monthly returns of 0.62% and 0.61%, while we find 0.68% and 0.60% (figures not reported in the table), for fully-invested and committed capital, respectively. The low returns are also confirmed in the period 2010-2019, when the returns we find for both approaches are almost zero.

Table 5: Monthly returns of top 20 pairs

This table provides the summary statistics of our implementation of GGR and Do and Faff's methodology from 1963 to 2019. All of the numbers shown are Fully-Invested returns, except for the last row, which presents the monthly mean returns of the Committed Capital approach. The Sharpe ratio is annualized.

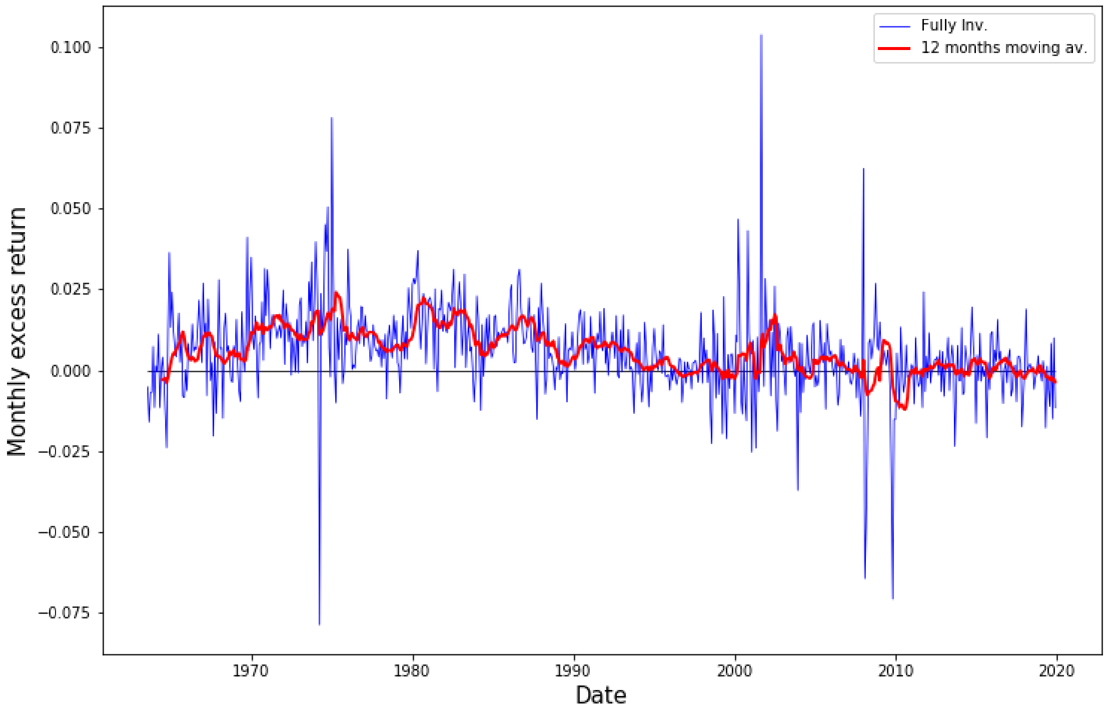
	1963-1988	1989-2002	2003-2009	2010-2019	Full period
Mean return (employed)	0.0105	0.0038	−0.0002	−0.0001	0.0056
Median	0.0099	0.0027	0.0021	0.0004	0.0047
Std	0.0131	0.0131	0.0168	0.0080	0.0137
Sharpe Ratio	2.7693	0.9990	−0.0415	−0.0325	1.4242
Min	−0.0788	−0.0253	−0.0708	−0.0236	−0.0788
Max	0.0780	0.1037	0.0624	0.0242	0.1037
Kurtosis	8.6023	20.0429	7.8961	0.9594	9.5750
Skewness	−0.4623	2.8069	−1.4240	−0.1167	0.1389
Obs. with return below 0	0.1475	0.3512	0.3929	0.4583	0.2836
Monthly serial correlation	0.2387	0.0382	0.1897	−0.1686	0.2493
Mean return (committed)	0.0091	0.0033	0.0003	0.0000	0.0050

We also find similar results for the fraction of months with negative returns, which is evident from the table that they have a great impact in the returns as in the first sub-period the faction is below 15%, ending in 46% for the last decade. Do and Faff find 11% of months negative returns in 1963-1988, 32% in 1989-2002 and 47% in 2003-2009. In terms of Sharpe ratios, we also get results close to those reported by Do and Faff for the periods 1963-1988 and 1989-2002 (2.76 and 1 vs. 2.94 and 1.2). We observe a difference in the period 2003-2009 when we get a negative Sharpe ratio of -0.0415 (influenced by a negative mean return of -0.0002) but Do and Faff find a positive Sharpe ratio of 0.8.

To further validate our approach, in figure 8, we recreate figure 1 from Do and Faff (2012), extended until 2019. We find the spikes in the monthly excess returns in the same months and for similar magnitudes as those reported by Do and Faff. The downward trend of pairs trading profitability is evident by looking at the higher frequency (as reported in Table 5) and magnitude of negative returns in the more recent years compared to the returns at the beginning of the period analyzed. It can be seen from the twelve months moving average, that the early periods were signified by high average return whereas since the late 90's, the returns have oscillated around zero with two major spikes doing the dot-com bubble and the financial crisis of 2008.

Figure 8: Monthly Fully-Invested returns of the portfolio of Top 20 pairs

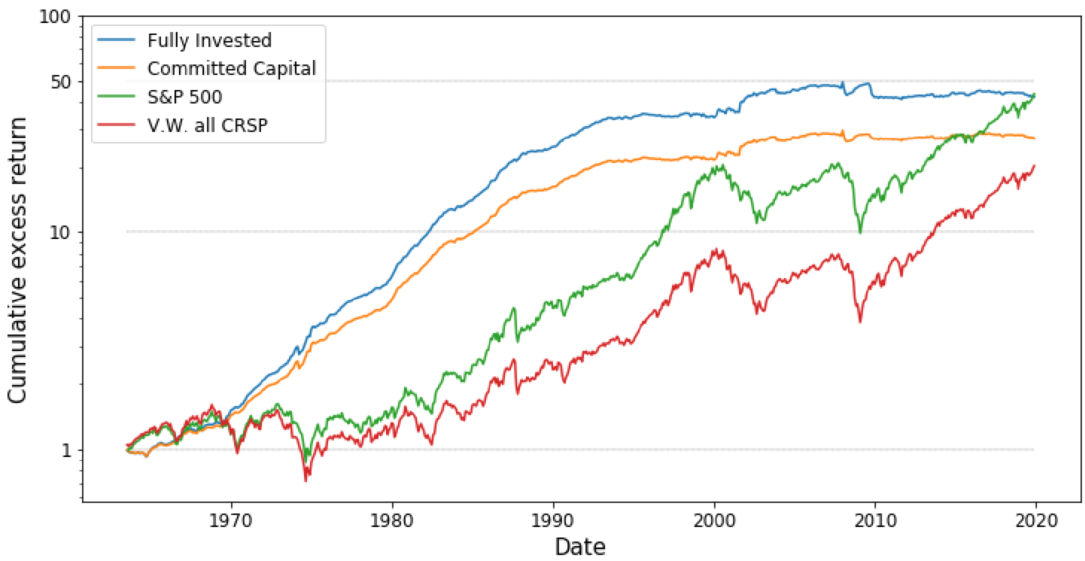
This figure shows the monthly returns time series of the Top-20 pairs from August 1963 until December 2019, together with the 12 months moving average of the monthly returns.



Additionally, in figure 9 we recreate figure 3 from Gatev et al. (2006) and extend it until 2019, by calculating the cumulative return of pairs trading compared to the SP 500 and the value-weighted all CRSP index. We get almost exactly the same as GGR even though we are not including Nasdaq stocks and we apply some extra filters to our stock selection process. It is clear from the figure, and consistent with table 5 and figure 8, that the volatility of pairs trading returns is remarkably low compared to the overall market, and that the returns have flattened out in the last decades. Some possible explanations for the lower returns in the recent years, as pointed out by GGR and Do and Faff, are the higher popularity of statistical arbitrage strategies in the hedge funds industry, and the worsening of arbitrage risks (unexpected disruption in the relative relationship between pairs stocks and higher irrational trading as indicated by Do and Faff).

Figure 9: Cumulative returns of Top-20 pairs

This figure shows the cumulative excess returns of both the Fully-Invested and the Committed Capital approaches, together with cumulative returns of the SP 500 index and the value-weighted all CRSP index. The y-axis is in logarithmic scale.



In summary, we find similar results to those reported by GGR and Do and Faff, with small differences that could be generated by our exclusion of Nasdaq stocks, the application of price filters for reasons explained in section 4.1, and the use of data from CRSP retrieved several years later.

Risk-adjusted returns of pairs trading

To explore the systematic risk exposure of pairs trading, we regress the fully-invested monthly returns on the 3 factors of Fama and French (1993), including the momentum factor and the illiquid minus liquid factor (*IML*, Amihud (2014)). We perform this analysis to get an initial understanding of what drives pairs trading returns.

Even though we open a position with 1 dollar in the long-leg and 1 dollar in short-leg, this does not mean that the strategy is market neutral as each stock can have different betas. Additionally, given that we do not implement any rebalancing, as our long and short positions change in value, the overall market exposure will be affected by market movements to some extent. Therefore, we expect to see some beta exposure in our portfolios.

In terms of value and growth stocks, if we consider long periods, there might not be a significant exposure to the *HML* factor. However, if the analysis is limited to periods of crises when investors prefer to own more liquid stocks to reduce their risk, we can expect to get signals to buy the higher risk assets (value) that investors want to get rid of. Regarding the exposure of pairs trading to the *SMB* factor, as small companies tend to be more volatile than the large ones, it is unlikely that we find pairs with low *SDD* of companies with a significant difference in size. Then, we do not expect to see a significant loading to the *SMB* factor.

If the strategy implemented in this paper systematically buys stocks with low liquidity, and sells the liquid counterpart, we expect to see a positive loading on the *IML* factor. Finally, as pairs trading bets on price reversals, it is essentially a contrarian investment strategy, so we expect to have a negative loading on the momentum factor.

In table 6 we present the results of the regressions for 4 different sub-periods and the entire period 1963-2019, before transaction costs.

We find that for the early sub-periods, we find significant alpha of 1% and 0.4% per month, in the two more recent sub-periods we find close to 0% alpha. The regressions exhibit positive a large loading on *IML* in the early periods, but this effect disappear in later periods. One should beware of the inherent multicollinearity between the *SMB* and *IML* factors. As we see they change sign in 2003-2009. We find a large market loading in the early period, and a large negative loading on the market in 2003-2009. These findings speak to the fallacy of the choice not to rebalance in order to limit transaction costs. In general we observe an expected negative momentum loading, however, only 2003-2009 is significant. The *HML* is significant for 2003-2009 as well. All factors are found significant for the sub-period capturing the financial crisis of 2008, and we find a much better fit for that sub-period, than for the rest.

Table 6: Factors - Replication of Do and Faff - Fully-Invested

This table presents the following 5-factor regression model

$$r_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2 (Mkt - RF)_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t}$$

Where $r_{net,t}$ is the monthly fully-invested return, before of transaction costs, on pairs trading portfolio. The table has parameter estimates and t-Statistics reported in parentheses below. The R^2 and R^2_{adj} are provided below. The t-statistics are computed using Newey-West standard errors with six lags. Factor loadings are presented for the Do and Faff sub-periods.

	1963-1988	1989-2002	2003-2009	2010-2019
Alpha	0.0097	0.004	-0.0005	-0.0003
	(9.33)	(3.31)	(-0.25)	(-0.56)
IML	0.149	0.0038	-0.24	-0.0273
	(3.87)	(0.1)	(-2.16)	(-0.46)
Mkt-RF	0.0655	-0.0202	-0.2715	0.0343
	(3.06)	(-0.43)	(-3.65)	(1.47)
SMB	-0.0915	-0.0661	0.4644	0.0012
	(-2.56)	(-1.36)	(3.5)	(0.03)
HML	0.0003	-0.0404	0.1163	0.0014
	(0.01)	(-0.78)	(2.09)	(0.05)
Mom	-0.0281	0.0014	-0.1082	-0.0267
	(-1.25)	(0.05)	(-2.22)	(-0.94)
R2	0.1024	0.0308	0.3139	0.0516
$R2_{adj}$	0.0874	0.0009	0.2699	0.0101

5.2 Long/Short disaggregation

In this section we look at the profitability of long and short positions separately, which naturally add up to the returns of pairs trading as we invest the same amounts on each leg. In order to obtain the hypothetical returns of investing only in the long or only on the short positions, we have to consider a gross investment of 2 dollars on each leg such that it can be compared with the total gross investment of the pair, following GGR (2006). Then, the returns attributable to just investing long or short would twice as much as their sole contribution to the pair.

We disaggregate our pairs portfolio into its long and short component to examine whether the pairs trader is simply exploiting mean reversion. This would be evident in the data, if returns on the long and short side are symmetric, as a signal would be generated by either a divergence by stock A or B with equal probability. The analysis further allows us to understand what drives our returns. We are asking the question of whether risk-adjusted returns are driven by our long or our short leg. Since we are assuming the possibility of taking short positions in any stock (with no zero-volume day during the formation period) is possible, we are naturally interested in whether the short leg is driving our returns. If so, one should examine short-restrictions in more detail. If instead, we find that the long leg is driving returns, it is interesting to consider if short positions should rather be taken in the market index rather than the stock due to a cost minimization objective.

In table 7 we present the monthly average returns of the long and short positions by different sub-periods chosen by how GGR and Do and Faff present their findings. In all sub-periods considered, the average long returns for fully-invested portfolios are positive and over 1% per month except for the period 2003-2009 which is around 0.78%, and we find an average monthly return of 1.33% for the full period 1963-2019. Comparing our results with those reported by Gatev et al. (2006) for the period 1963-2019, we find almost no difference with the authors as they report monthly average returns of 1.33% while we find them to be 1.34%. The returns have been more stable in recent years as their volatility has been decreasing, and they experienced less extreme minimum and maximum observations.

For short positions, we find positive monthly average returns only for the early years considering 1963-2002. All other sub-periods present negative returns for our short-portfolio, with increasing magnitudes and frequency of months with losing positions. Compared to Gatev et al. (2006), we find lower monthly average returns of 0.28% for the period 1963-2019, while they report 0.44% for the same period.

Figure 10 shows the gross exposure to long and short positions separately for all the 6 overlapping portfolios comprised of the top 20 pairs. As each position is initiated with 1 dollar in the long leg and 1 dollar in the short leg of the pair but no rebalancing is applied during the trading period, the pairs traded can end up having a long or short net exposure. This

Table 7: Returns of Long and Short positions

Summary statistics on the disaggregation of the pairs trading portfolios returns into its long and short components. Panel A shows the summary statistics for the long leg. Panel B shows the summary statistics for the short leg. Sub-periods are based on Do and Faff sub-periods, as well as GGR (2006) who perform the same analysis from 1963-2002. All of the numbers shown are Fully-Invested returns, except for the last row, which presents the monthly mean returns of the Committed Capital approach. The Sharpe ratio is annualized.

(a) **LONG** positions

	1963-1988	1989-2002	1963-2002	2003-2009	2010-2019	Full period
Mean return (fully inv.)	0.0141	0.0120	0.0134	0.0078	0.0169	0.0133
Median	0.0090	0.0173	0.0107	0.0086	0.0194	0.0124
Std	0.0612	0.0619	0.0614	0.0373	0.0539	0.0576
Sharpe Ratio	0.7989	0.6740	0.7550	0.7224	1.0888	0.8005
Min	−0.3528	−0.1499	−0.3528	−0.1325	−0.1120	−0.3528
Max	0.3927	0.1818	0.3927	0.0985	0.1642	0.3927
Kurtosis	8.4635	0.2768	5.4487	2.6895	0.0741	5.1108
Skewness	0.3425	−0.1189	0.1749	−0.7236	−0.2618	0.1036
Obs. below 0	0.4197	0.3393	0.3911	0.3929	0.3333	0.3811
Monthly serial corr.	0.1018	0.0094	0.0688	0.2697	−0.2600	0.0300
Mean return (committed)	0.0123	0.0145	0.0116	0.0071	0.0150	0.0117

(b) **SHORT** positions

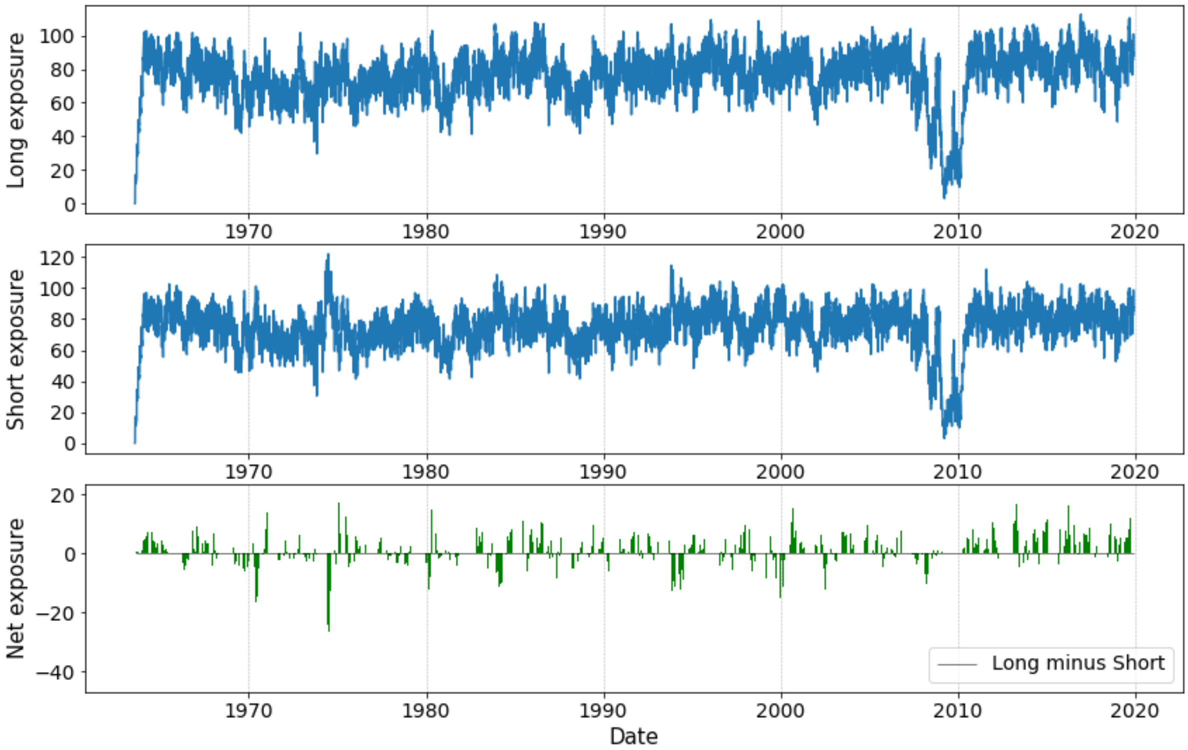
	1963-1988	1989-2002	1963-2002	2003-2009	2010-2019	Full period
Mean return (fully inv.)	0.0069	−0.0045	0.0028	−0.0082	−0.0171	−0.0021
Median	0.0064	−0.0067	0.0016	−0.0057	−0.0246	−0.0056
Std	0.0532	0.0574	0.0549	0.0254	0.0537	0.0525
Sharpe Ratio	0.4473	−0.2700	0.1791	−1.1182	−1.1014	−0.1359
Min	−0.2366	−0.1446	−0.2366	−0.0741	−0.1252	−0.2366
Max	0.1953	0.1625	0.1953	0.0496	0.1093	0.1953
Kurtosis	2.1189	0.1769	1.1988	0.4479	−0.1394	1.1361
Skewness	0.0238	0.3475	0.1310	−0.3096	0.5218	0.2396
Obs. below 0	0.4492	0.5595	0.4884	0.6786	0.7000	0.5495
Monthly serial corr.	0.1106	−0.0378	0.0621	0.2585	−0.2791	0.0254
Mean return (committed)	0.0059	−0.0038	0.0024	−0.0066	−0.0150	−0.0018

is presented in the bottom part of the figure which takes the difference in the gross exposure between long and short positions at each point in time. There is no clear pattern of a positive or negative exposure throughout the entire period but the strategy can end up having significant market exposure as the net long/short exposure moves away from the equilibrium.

In table 8 we present the risk-adjusted returns in excess of the 30-day Treasury bill returns, following the methodology of GGR. The risk adjustment factors include the Amihud *IML* factor, the Fama and French three-factor model and finally the Jegadeesh and Titman momentum

Figure 10: Long-Short exposure

Total long and short exposure for the 6 overlapping portfolios of top 20 pairs. The bottom part shows the difference between long and short total exposures. Each position is initiated with 1 dollar in both long and short positions, so a positive Net Exposure indicates that the total gross amount on long positions exceeds the amount on short positions, and vice versa.



factor.

Given the limitation on short opportunities on most stocks in the early period of our analysis, we expect to see significant risk-adjusted returns to be primarily driven by the short leg. In recent years, as short opportunities are less scarce, we expect a less negative alpha. Our exposure to the market is naturally expected to be given as a strongly positive loading on both legs, as the short is considered from a long perspective. The momentum factor loading is expected to be negative as we are following a contrarian strategy and could therefore load negatively on momentum. The *IML* factor exposure is expected to be positive given our hypothesis of loading on illiquid stocks, while shorting the liquid counterparts. We note that this may be affected by some degree of multicollinearity between the *SMB* factor and the *IML* factor as they capture some of the same variation. Small stocks are naturally more illiquid, and therefore, as we expect to by illiquid, our long side should load positively on the *SMB* factor, while our short leg should be in the liquid counterpart, and as such, we should expect a negative sign. For the *HML* factor, we cannot see any systematic reason why we should load on value over growth on average, as the pairs trader does not take book values into account. One could speculate whether a positive sign could be expected due to the strategy buying the cheap stock and selling the expensive stock should lead to a positive tilt towards value stocks.

In line with the findings from GGR who regress on a reversal factor instead of a liquidity factor, in table 8 we observe that the short leg is, in fact, driving much of the abnormal returns. GGR speculate that since the short positions are generating all the risk-adjusted excess return, and the short leg contains stocks that have increased in value relative to the counterpart before opening the trade, it is unlikely that the pairs trading returns are driven by a reward for unrealized bankruptcy risk. We find that the long leg is losing money during the period analyzed by GGR. While they find 24 bps positive risk-adjusted excess returns, it is insignificant. As mentioned, a difference is to be expected as they are including Nasdaq in their analysis. Our analysis finds similar statistics as those found by GGR, and close to our hypothesized signs. We naturally load heavily on the market on both the long and short portfolio. The *IML* factor is significant for all periods but the 2003-2009 sub-period. So in more recent years, the *IML* factor does not seem to capture returns. For the long leg it is a positive loading and for the short it is negative, it indicates as we hypothesized that the pairs trader is long the illiquid stocks and short the liquid counterpart. We do however find that the *SMB* loading is negative for the long leg, indicating that we are loading on large capitalization stocks. One can again speculate that this is an effect of the inherent positive correlation between the *IML* and *SMB* factors. We see that in 2003-2009, the signs change for the two as well. For the *HML* factor we do see strong positive significance. One could speculate that this is an effect of the element of buying cheap and selling expensive in the pairs trading algorithm, but it is unclear. The momentum factor is only significant for the long leg in the sub-period 1963-1988, where the loading is, as expected, negative. We finally note that the R^2 appears to be dropping over time. This could indicate that pairs trading results are not captured by the same factors today as they were decades ago, or merely that the transaction costs are lower today, and liquidity in the market is far greater than in the past. One could therefore argue that periods of time where the pairs trader provide liquidity, to a degree that deserves a premium, are more scarce today than 50 years ago.

Given our results, we will now account for transaction costs to determine if the abnormal excess returns on the short leg could be an effect of ignored short restrictions.

Table 8: Risk-adjusted returns of Long and Short positions

Risk-adjusted return of long and short positions for top 20 fully-invested returns.

(a) LONG positions

	1963-1988	1989-2002	1963-2002	2003-2009	Full Period
Alpha	-0.0022	-0.0027	-0.0029	0.0011	0.0025
	(-1.34)	(-1.12)	(-2.1)	(0.47)	(1.34)
IML	0.2251	0.4238	0.3228	-0.1474	0.645
	(2.55)	(3.71)	(4.29)	(-1.24)	(4.32)
Mkt-RF	0.5318	0.4221	0.5003	-0.0647	0.4041
	(10.93)	(7.1)	(11.65)	(-0.79)	(6.53)
SMB	-0.2476	-0.4464	-0.3207	0.3767	-0.5579
	(-3.41)	(-5.19)	(-5.37)	(2.56)	(-4.19)
HML	0.3849	0.2445	0.3539	0.053	-0.0962
	(6.33)	(2.63)	(6.67)	(0.64)	(-1.22)
Mom	-0.1157	0.0532	-0.0152	-0.0702	0.0564
	(-2.39)	(1.35)	(-0.45)	(-1.14)	(0.71)
R2	0.5395	0.344	0.4459	0.1656	0.2854
$R2_{adj}$	0.5318	0.3237	0.44	0.1121	0.254

(b) SHORT positions

	1963-1988	1989-2002	1963-2002	2003-2009	Full Period
Alpha	-0.0119	-0.0067	-0.0105	0.0015	0.0028
	(-7.55)	(-3.45)	(-8.25)	(0.99)	(1.56)
IML	0.0761	0.42	0.1992	0.0926	0.6723
	(0.93)	(3.73)	(2.77)	(1.4)	(5.59)
Mkt-RF	0.4663	0.4423	0.4529	0.2068	0.3697
	(10.39)	(8.37)	(11.63)	(3.73)	(5.73)
SMB	-0.1561	-0.3803	-0.2316	-0.0877	-0.5591
	(-2.44)	(-4.07)	(-4.0)	(-0.98)	(-4.2)
HML	0.3846	0.2849	0.3667	-0.0633	-0.0976
	(6.3)	(3.18)	(6.93)	(-1.08)	(-1.4)
Mom	-0.0875	0.0518	-0.0007	0.0381	0.0831
	(-1.86)	(1.36)	(-0.02)	(1.61)	(0.95)
R2	0.5353	0.3944	0.4514	0.2903	0.2703
$R2_{adj}$	0.5275	0.3757	0.4455	0.2448	0.2383

5.3 Transaction costs

In the previous sections, we presented the results of implementing pairs trading before taking into account any transaction costs, but these costs cannot be neglected as the strategy can be potentially be trading 240 stocks every month (20 pairs x 2 stocks/pair x 6 overlapping portfolios). Furthermore, we found that the risk-adjusted returns were primarily driven by the short-leg.

We apply the transaction costs explained in section 4.5. They cover trading commissions, market impact and short-selling fees. We obtain similar results (reported in table 9) to those of Do and Faff (2012). They show that after transaction costs, pairs trading is not profitable, consistent with our own findings. Even though they do not report the returns by different periods, we find that for all periods considered, both the mean and the median monthly returns are around zero, with slightly negative returns for the entire period 1963-2019.

Given these results, one can argue that even though pairs trading shows attractive results before transaction costs for the early periods, it could also be because these potential returns were not possible to exploit given the costs they would implicate. Do and Faff (2012) suggest that pairs trading could still be profitable after transaction costs if the pairing is done within well defined industry groups or by combining the SSD ranking with a sort by the number of crossings in prices during the formation period. However, the purpose of our thesis is not to try to improve the strategy, which has been extensively attained in the existing literature, but to explain the rationale behind the returns of pairs trading.

Despite our findings, more extensive research on the effect of transaction costs in pairs trading is needed as we are applying the same costs to all long and short positions, no matter the size or liquidity of the stocks in the portfolio. It could be the case that we are overestimating the costs of trading as our portfolios on average contain stocks that belong to the 8th decile in market capitalization (being 10 the stocks with the highest market cap.)(see section 5.4.2), for which one can expect lower transaction costs than for lower size decile stocks.

Table 9: Mean returns - Net of Transaction Costs

This table presents the summary statistics of our implementation of GGR and Do and Faff's methodology from 1963 to 2019. All of the numbers shown are for Fully-Invested returns of the top 20 pairs net of transaction costs, except for the last row, which indicates the monthly mean returns of Committed Capital net of transaction costs. The Sharpe ratio is annualized.

	1963-1988	1989-2002	2003-2009	2010-2019	Full Period
Mean return (employed)	−0.0004	0.0009	−0.0026	−0.0013	−0.0005
Median	−0.0002	0.0002	0.0001	−0.0011	−0.0003
Std	0.0128	0.0129	0.0167	0.0080	0.0127
Sharpe Ratio	−0.0993	0.2303	−0.5421	−0.5654	−0.1386
Min	−0.0986	−0.0272	−0.0736	−0.0250	−0.0986
Max	0.0612	0.0998	0.0574	0.0229	0.0998
Kurtosis	12.0797	21.1214	8.2203	0.9494	14.7203
Skewness	−1.2090	2.9752	−1.6602	−0.1295	−0.2810
Obs. with return below 0	0.5082	0.4881	0.5000	0.5750	0.5140
Monthly serial correlation	0.2522	0.0173	0.1973	−0.1712	0.1591
Mean return (committed)	−0.0004	0.0007	−0.0016	−0.0011	−0.0004

PART II - Loadings

5.4 Sector, size and illiquidity loadings

Now that we have validated our methodology by comparing our results to those of GGR and Do and Faff, we attempt to answer our research questions stated in the introduction. We start with whether the pairs trader is providing liquidity or not, and how it has changed over time.

In order to obtain more details regarding what goes into the pairs trader's portfolio, we consider what risk exposures and loadings the pairs trading strategy is subject to. In this section, we examine to which sectors the stocks traded belong, their market capitalization decile and their Amihud illiquidity measure decile. This will allow us provide an answer the research question of whether the pairs trader is actually providing liquidity or not, and how the level of liquidity provision has changed over time. We disaggregate the analysis into how our long and short legs are loading on each category. In section 5.6.1, we analyze the performance of pairs trading when the pairing process is restricted to stocks of the same sector, but here we will focus on unrestricted pairs to provide more information about how stocks are paired when applying the methodology presented by GGR and Do and Faff.

5.4.1 Sector Loadings

We define 14 different sectors based on the variable *HSICIG* from CRSP and we map each of the codes with the definitions of sectors from SIC website (See appendix table 21 for reference). In table 10 we present the fraction of long and short positions taken in each of the sectors considering the total number of positions taken over the period 1963-2019, and in the right-most column we show the fraction of pairs traded for which both the long and the short positions are in stocks of the same sector. The vast majority of the positions are taken in stocks that below to the Utilities sector (74% of long positions and 77% of short positions), which is in line with Gatev et al. (2006) who finds that 71% of the positions are taken in Utilities, despite that our analysis includes 17 years more of data. Given these figures for longs and shorts, we observe that 68.58% of all positions are taken in pairs formed by two Utilities stocks. We hypothesize that it is an effect of these companies have a rather stable demand as their products have low differentiation and are consumed with little correlation with the swings in the economy. As such, stocks in the Utilities sector should have lower betas than the rest of the market.

The second sector in terms of fraction of positions taken is Financials, with 10% and 7% for long and short positions, respectively. We find that a total of 6.15% of all positions are in pairs formed by two Financials stocks which is a significant portion considering that Utilities comprise most of the positions. For Financials, there is a larger room for differentiation than for Utilities, but they are still commonly affected by macroeconomic factors such as interest rates or unemployment shocks. In contrast with GGR, who find that 13% of all positions are in Financials, we find that only a total of 6.15% of all trades in pairs formed by two Financials stocks. This difference is most likely caused by our broader segregation of stocks into 14 sectors, compared to the four sectors considered by GGR.

In third place comes the Communications sector, representing 4.47% for all long positions and 3.09% of all short positions, but they are not always traded together in the same pair, as only 1.42% of all positions are in pairs of two Communication stocks. The rest of the positions are spread across the other 11 sectors, with a total of 78.35% of all trades done over stocks of the same sector.

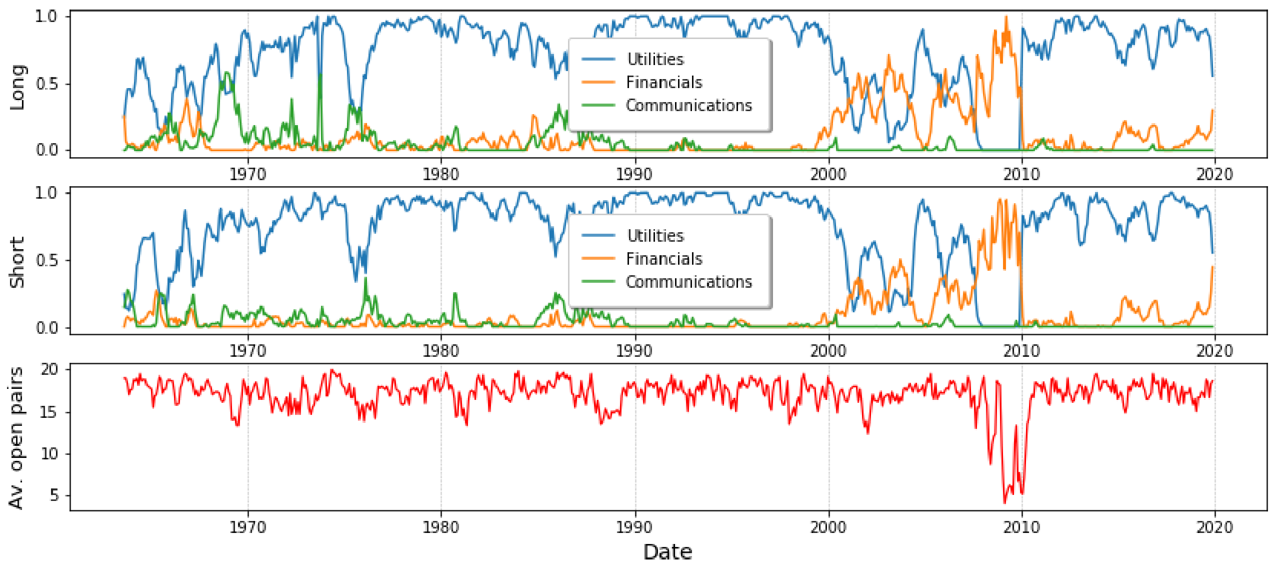
Table 10: Composition of pairs by SIC sectors

This table presents in the columns in the middle, the fraction of long and short positions in each of the sectors considering the total number of positions taken in the period 1963-2019. The right-most column shows the fraction of pairs traded for which both the long and the short positions represent stocks that belong to the same sector.

Sector	Long	Short	Same sector pairs
Utilities	0.7405	0.7700	0.6858
Financials	0.0994	0.0702	0.0615
Communications	0.0447	0.0309	0.0142
Materials	0.0299	0.0272	0.0074
Industrials	0.0157	0.0248	0.0045
Mining	0.0167	0.0245	0.0030
Transport	0.0167	0.0057	0.0028
Consumer Goods	0.0081	0.0171	0.0022
Consumer Services	0.0114	0.0095	0.0010
Wholesale	0.0020	0.0049	0.0004
Retail	0.0042	0.0059	0.0003
Technology	0.0041	0.0036	0.0001
Real Estate	0.0046	0.0053	0.0001
Construction	0.0020	0.0006	0.0001
Total	1.0000	1.0000	0.7835

Figure 11: Sectors Loadings

This figure presents the fraction of long (upper part) and short positions (middle part) in each of the sectors considering the total number of positions taken each month during the period 1963-2019, for the top 20 portfolio. To ease the visualization, only Utilities, Financial and Communications sectors are shown as they represent the vast majority of all trades. The bottom part of the image shows the average number of pairs with at least one day of open positions each month.



5.4.2 Size loading

To obtain information about the market capitalization of the companies behind the stocks chosen for trading, we rank stocks in deciles based on their market capitalization at the beginning of each month. If the pairs trader systematically buys small capitalization stocks over large ones, we can expect some loading on the SMB factor by Fama and French, but also on Amihud's *IML* factor.

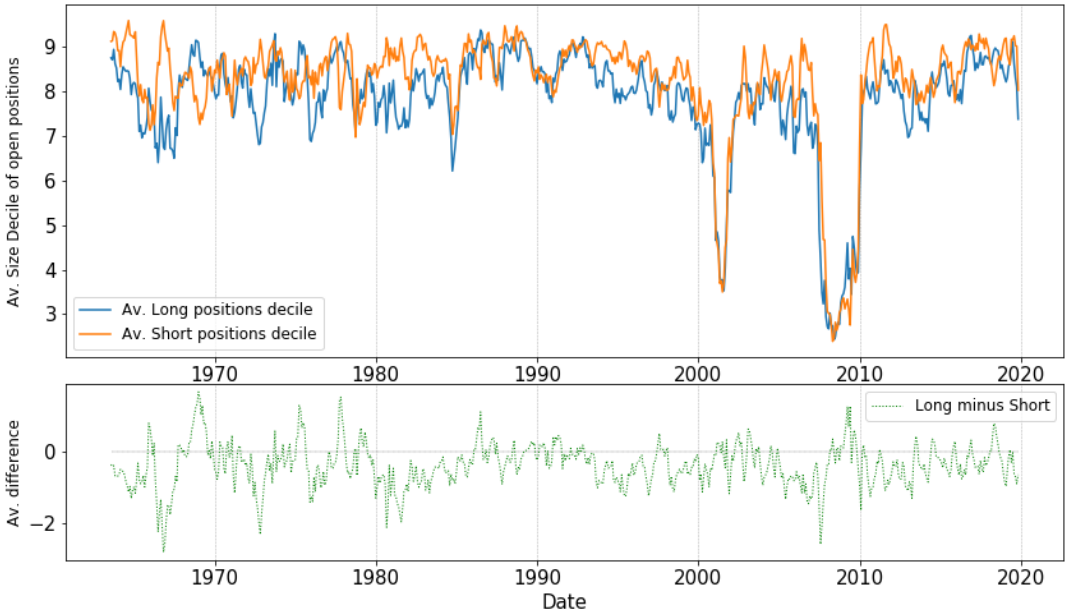
In figure 12 we present the average size deciles of all pairs traded each month, and the lower part shows the difference between the averages for long and short positions. We can see that the lines tend to move together with some temporal noticeable divergences and that there is a tendency for taking long positions in stocks with lower size than the stocks shorted. This finding is expected to some extent, because when two companies have similar risk factors and their prices move together closely, it is likely that they are of similar sizes. Therefore, when there occurs a divergence in their prices, one of the stocks increases in market capitalization, while the other decreases. As a result, one could argue that this could cause the paired stocks to end up located in different size deciles.

Our results indicate that, for the entire period 1963-2019, we take long positions on stocks with an average size deciles of 7.81 (being 10 the highest market capitalization), while the average decile of the short positions is 8.2, resulting in an average size decile for each pair of 8.01 and an average difference in long and short deciles of 0.39. Additionally, we find that only 29% of the pairs include stocks from the same size decile but this number might be so informative as there could be several pairs of stocks from almost the same size which are on the limits of each decile. Then, if we consider the pairs for which the stocks have no more than one decile of difference, we find that 65% of pairs are formed with stocks of similar market capitalization.

If we only consider the period covered by Gatev et al. (2006), 1963-2002, our results indicate that the strategy we implement has an average difference in long and short size deciles loading of 0.39, than that reported by GGR (0.97). They do not report long and short separately, but they point out that the average size decile of their pairs is 7.29, while we find a this figure to be 8.20 in our analysis. Hence, our implementation of pairs trading, on average, finds more homogeneity in the composition of the pairs based on size, compared to GGR. These differences may arise because we are not using Nasdaq stocks and we apply additional price filters in the formation period.

Figure 12: Average Market Cap. Decile of open positions

The upper part of this figure shows the average market capitalization decile of the long and short positions open at each point in time for the top 20 portfolio. Decile 10 represent the stocks with the highest market capitalization, while decile 1 comprise the smallest companies. The lower part indicates the difference over time between the long and short positions average market capitalization decile.



5.4.3 ILLIQ loading

To get an indication of whether a shock to market or funding liquidity leads to the pairs trader providing liquidity in the more illiquid stocks while shorting the more liquid counterpart, we examine the average Amihud *ILLIQ* decile loading of the strategy. *ILLIQ* can take extreme values which makes it hard to compare results. Therefore, to avoid that outliers drive our results, we rank stocks based on their *ILLIQ* measure each month and assign them to *ILLIQ* deciles.

In figure 13 we present the average monthly *ILLIQ* deciles of all pairs traded each month, separated by long and short positions, and the lower part shows the difference between the

ILLIQ deciles average for long and short positions. We can see that the lines tend to move together with some temporal noticeable divergences, and that there is a tendency for taking long positions in stocks with higher *ILLIQ* than the stocks shorted, as shown in the lower part of the figure. This figure looks as the inverted figure 12, and this not surprising as small cap. stock are usually more illiquid than large cap.

Our results indicate that for the entire period 1963-2019, we take long positions on stocks with an average *ILLIQ* decile of 3.41 (being 10 the decile formed by the most illiquid stocks), while the average *ILLIQ* decile of the short positions is 3.01, resulting in an average *ILLIQ* decile for each pair of 3.21 and an average difference in long and short deciles of 0.40. Additionally, we find that only 27% of the pairs include stocks from the same *ILLIQ*, and when we consider the pairs for which the stocks have no more than one *ILLIQ* decile of difference, we find that 62% of pairs are formed with stocks of similar liquidity.

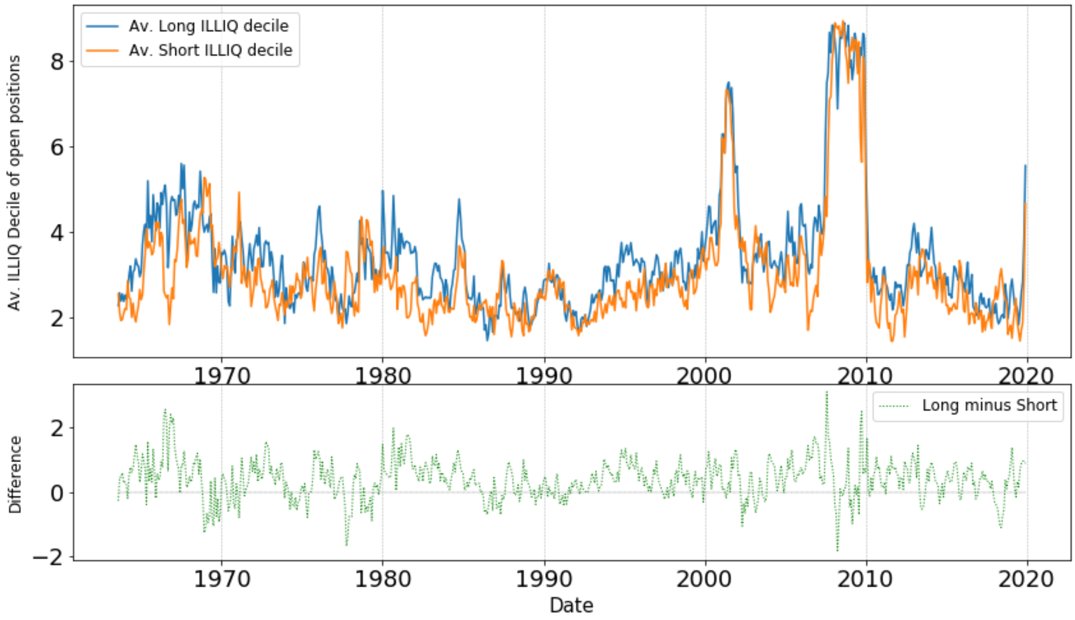
In conclusion, there is a tendency in the strategy to take long positions in stocks that are more illiquid than the stocks shorted, but this does not necessarily mean that the pairs trader only buys the extremely illiquid stocks and sells the most liquid ones as the difference between long minus short *ILLIQ* deciles seldom exceeds a full decile. These results could to some extent be argued as expected because the GGR algorithm only trades the pairs with the lowest SSD during the formation period, and if one of the stocks is extremely illiquid while the other stock is highly liquid, there could be a significant difference in their price series movements which would give the pair a low ranking.

Exposure during the financial crisis

The composition of the portfolios during the financial crisis of 2007-2008 deserves to be analyzed separately from all the other years considered in this paper, as it was a period with severe changes to the exposure levels for the pairs trader. The effects of the financial crisis were devastating for the stock market, with several companies going bankrupt and other companies carrying out major corporate restructurings to survive. Considering that the strategy implemented here does not take into account the broad market conditions but just the relationship in prices between pairs of stocks, this period of major disruption affects not only the composition

Figure 13: Average *ILLIQ* Decile of open positions

The upper part of this figure shows the average decile of the long and short positions for the top 20 portfolio, taken each month and sorted by the Amihud *ILLIQ* measure of the stocks. Decile 10 includes the most illiquid stocks, while decile 1 is formed by the most liquid stocks. The lower part indicates the difference over time between the two lines from the upper part of the figure.



of the portfolios, but also the number of trades carried out.

In figure 11 we show from 1963 until 2019, the fraction of positions taken in the long and short legs for three of the sectors which account for most of the trades executed: Utilities, Financials and Communications. Utilities (blue bars) on average comprise most of the positions, but during the financial crisis we see a surge in trading of Financials (orange bars) and a drop in Utilities. By looking at figure 14, which shows in stacked bars format, the same information as in figure 11 but just for the period 2007-2010, we can see that most of the long and short positions during the crisis are taken in Financials stocks. The bottom part of figure 11 and the solid lines in figure 14 represent the number of pairs with at least one position averaged across the six portfolios trading at the same time. In figure 11 we see that this number oscillates between 14 and 20 for all years except during the financial crisis.

It can be observed in figure 14 that at the beginning of the financial crisis (around mid-2007) there is a sudden increase in the average number of pairs traded from 14 to over 18. Then, in early 2008 there is a dramatic drop in the average number of pairs traded to an average of around 8, after the Dow Jones index starts its downward trend after hitting hits its all-time-high in October 2007 and the U.S. reports a contraction of the GDP in the fourth quarter of 2007. In the second part of 2008, with the failure of Bears Stearns and later the bankruptcy of Lehman Brothers among other financial institutions, there is a new spike in the number of positions, led by Financials, followed by a sudden drop in the number of positions in 2009, which stay relatively low during the entire year.

A possible explanation to these dramatic changes in the composition of the portfolios and the frequency of trading could be that the "normal" relationships in prices measured during the formation periods are completely broken apart in the trading period as a consequence of the increased turbulence in the markets. Therefore, pairs that traded in a relatively low range of normalized prices differences, may not do so when the crisis hits. This mechanism both presents a problem and an opportunity for the pairs trader. The opportunity for the pairs trader consists of an increase number of signals for the already active portfolios. These new signals may entail an increased liquidity premium, as barriers to arbitrage are high. The problem is for the period to come. The period of market turbulence is now part of the new formation period. This results in a new high threshold of price differences for the coming trading periods, so if the volatility does not stay high, there will be fewer trading signals.

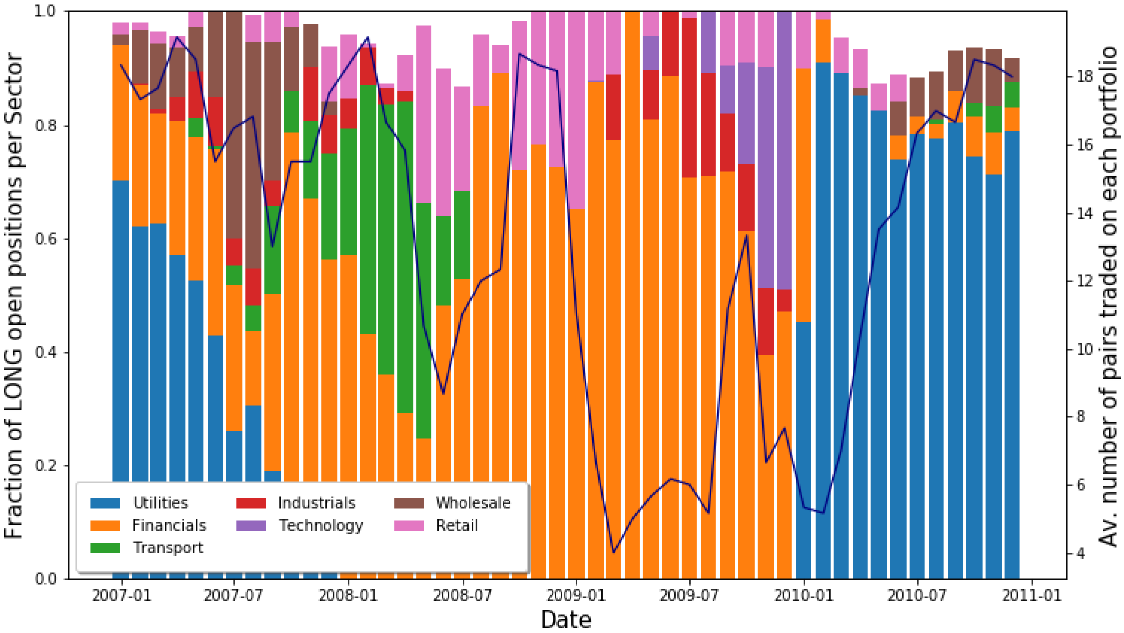
In a relatively stable market, one would expect the pairs trader to short stocks that have gone up in price and therefore moved higher in the size deciles. In terms of the size of the companies traded during the financial crisis, we see an interesting pattern. Figure 12 exhibits a large drop in the average size decile of both the long and short positions. This is probably the consequence of the high loading on financial stocks, which all suffered losses during this period.

In summary we find some evidence of liquidity provision on average when considering the average loading on Amihud's *ILLIQ* deciles. But given that the net exposure to illiquidity is limited, we do not expect to observe a significant liquidity premium on average.

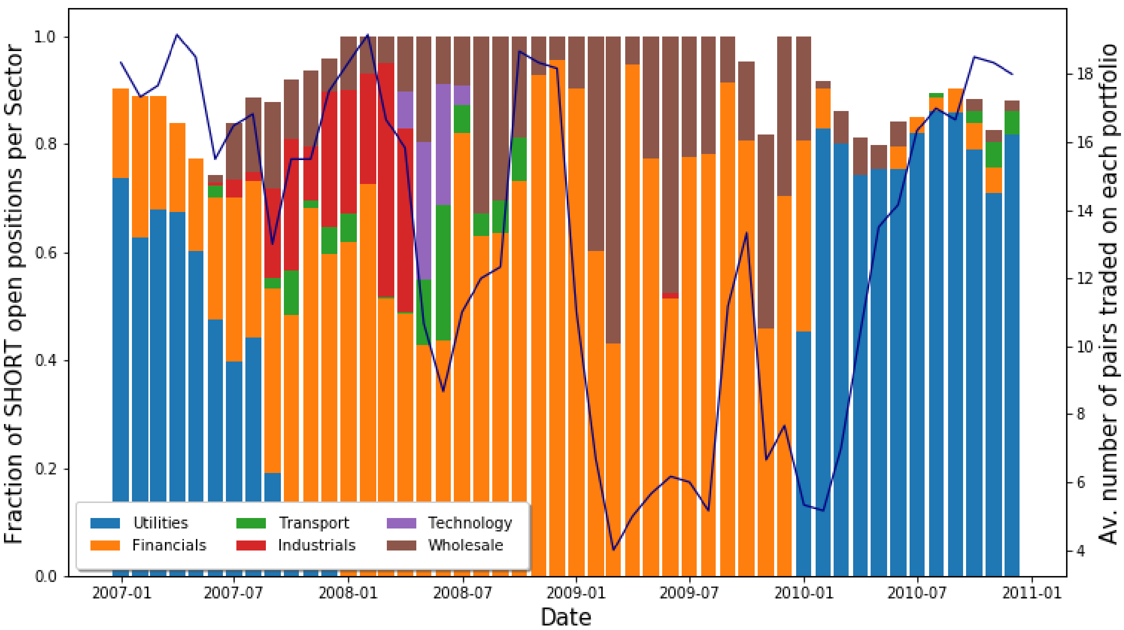
Figure 14: Fraction of open positions by sectors during the financial crisis

This figure presents the composition by sectors of all positions taken each month from 2007 until 2010. The horizontal bars show which fraction of all positions is taken in each sector. To avoid filling up the figure with too much information, only the sectors that for any of the months considered represent at least 25% of all positions, are displayed in the figure. The upper sub-figure is for long positions, while the lower sub-figure is for short positions. The solid line in each sub-figure indicates the number of pairs with at least one open position each month, and its magnitude is indicated in the Y-axis to the right.

(a) LONG positions



(b) SHORT positions



PART III - Liquidity Shocks

5.5 VIX and TED sub-periods

After getting a better understanding of what the pairs trader keeps in his/her portfolio, we now move our focus to the research question about whether specific sub-periods characterized by high barriers to arbitrage offer the pairs trader enough profitable opportunities to generate abnormal returns. Nagel (2012) finds indications of implied volatility having predictive power when explaining returns in investment strategies characterized as liquidity providing. Under the assumption that the pairs trader is liquidity providing, we should expect to see improved returns during these sub-periods. Given our results from section 5.4.3 it is not definitively evident that the pairs trader in a GGR framework is providing a significant amount of liquidity.

In table 11, we present raw summary statistics on fully-invested returns net of transaction costs during large changes in the VIX. Panel A presents results with no lag, meaning, we consider month pairs trading portfolio returns with contemporaneous changes in the VIX. We observe that the pairs trader seems to benefit for large spikes, both in the negative and positive direction. However, neither are statistically significant. We do however note that the sample size is limited. We do see, as hypothesized, that the pairs trader makes money, during periods with decreasing volatility and increasing market liquidity. Seemingly, the pairs trader generates some returns, but just 21 bps net of transaction costs, which is also not statistically significant. Contrary to our hypothesis, average returns seem more lucrative during months where the VIX increased a lot throughout the month with 46 bps net of transaction costs. We note that the distribution of our 17 monthly returns is positively skewed with a median below 0, and a max. of 9.98%. Seemingly, the returns are skewed primarily driven by positive outliers. We speculate that large changes in the VIX do not necessarily mean that the VIX ends the month at the intra-month high, and therefore market liquidity may have stabilized since the top, and as such, the pairs trader has already reaped the liquidity premium during the period in which we observe the month over month spike in the VIX. Furthermore, we observe that only the far edge cases appear to provide a raw payoff for the pairs trader. As such, the results are improved from considering the full period, yet the returns are insignificant at an alpha level of 5% and of limited scale.

Considering the lagged returns in panel B, we observe more or less the same scale of returns for the non-lagged VIX drops, as we see for the lagged VIX surges: 24 bps after transaction costs during market downturns. One could argue that following a large surge in the VIX, a large drop is more likely. Therefore, we speculate that the two sections capture some of the same information, and that is why returns are of similar scale.

As we are considering monthly changes, one could perhaps argue that a weekly disaggregation (we expect daily returns to be too noisy) could be an interesting analysis to make, as one short term volatility shock may rebound within the same month. As such, we do lose some information about the exact period of time when the returns are high for the pairs trader by considering month-over-month changes in the VIX and TED spread, but weekly changes are outside the scope of this paper and suggested for further research.

Table 11: Δ VIX - Net Portfolio

This table reports summary statistics for large changes to the market liquidity, approximated by changes in the VIX index. Returns are given as fully-invested returns after transaction costs. Results are given for VIX changes with no lag in Panel A, and lagged VIX changes in Panel B. The column names indicate the percentile of changes in the VIX. For example, percentile 0.95 means that we consider months in which the change in the VIX was more extreme than the 95th percentile.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Panel A:	No lag									
Mean	0.0021	-0.0013	-0.0004	-0.0007	-0.0001	-0.0005	-0.0004	0.0007	0.0014	0.0046
Median	0.0022	0.0012	0.0012	0.0015	0.0013	-0.0011	-0.001	-0.0011	-0.0013	-0.0016
Std	0.0079	0.0149	0.0132	0.0125	0.0119	0.018	0.0184	0.0208	0.023	0.0273
T-stat	1.12	-0.51	-0.25	-0.48	-0.1	-0.24	-0.17	0.25	0.36	0.69
Kurtosis	1.99	16.18	17.3	15.64	15.88	11.85	13.28	10.27	9.9	10.03
Skewness	0.62	-3.39	-3.3	-2.99	-2.82	2.41	2.76	2.49	2.4	2.86
Min	-0.0127	-0.0736	-0.0736	-0.0736	-0.0736	-0.0392	-0.0387	-0.0387	-0.0387	-0.0244
Max	0.0229	0.0229	0.0229	0.0229	0.0229	0.0998	0.0998	0.0998	0.0998	0.0998
Obs	18	36	54	72	90	88	70	52	34	17
Panel B:	Lagged									
Mean	0.0004	-0.0021	-0.0021	-0.0016	-0.0011	0.0011	0.0005	-0.0025	-0.0019	0.0024
Median	-0.0011	-0.0013	-0.0013	-0.0011	0.0001	0.0005	0.0003	0	0	0.002
Std	0.0102	0.0108	0.0094	0.0097	0.009	0.0187	0.0205	0.0178	0.0171	0.0122
T-stat	0.16	-1.19	-1.63	-1.41	-1.12	0.55	0.21	-1.02	-0.66	0.8
Kurtosis	0.86	0.66	1.25	0.73	1.04	12.45	10.77	7.03	8.53	-0.43
Skewness	0.75	-0.07	-0.06	-0.32	-0.44	0.47	0.51	-2.18	-2.19	-0.22
Min	-0.0178	-0.0272	-0.0272	-0.0272	-0.0272	-0.0736	-0.0736	-0.0736	-0.0736	-0.022
Max	0.0258	0.0258	0.0258	0.0258	0.0258	0.0998	0.0998	0.0229	0.0229	0.0229
Obs	19	37	55	73	91	88	70	52	34	17

If we turn our attention to the surges and drops in funding liquidity, approximated by the TED spread, we observe a very similar picture. Table 12, Panel A, shows a mean return of 20-27 bps in both edge cases. Unlike the VIX, less extreme changes in the TED spread seem to generate on average the same returns as the extreme thresholds, and given the larger sample, we observe a higher degree of statistical significance, yet 20 bps is probably not of any economical importance.

Panel B considers the lagged changes to the TED spread. Noteworthy is that for observations when funding liquidity eases in the prior month, the pairs trader generates an average of 47 bps. This amount is not statistically significant given the sample size and the nature of the move, and the median is a mere 25 bps. Once again, as with the VIX, the conclusion must be that the GGR-replicating pairs trader is getting abnormal returns after transaction costs. However, the returns are higher for sub-periods with liquidity shocks relative to the Do and Faff's sub-periods as well as for the full period as presented in table 5. Given our observations in section 5.4.3, one could argue, as done by Do and Faff, that trading the top pairs may not be beneficial as they move too closely together,

To extend the analysis of raw returns during market and funding liquidity shocks, we now consider factor regressions during the same period, to get an indication of whether the returns are driven by exposure to common factors.

Before estimating our regressions, we state our expectation. When the change in VIX is exceeding our lower bound thresholds of percentiles 0.05 to 0.25, we expect to observe a positive loading on the *IML* factor. As markets become more liquids, the pairs are expected to converge, whereas when the market volatility increases a lot; the opposite is expected as pairs are expected to diverge as volatility increases. As stated in the previous section, our results are affected by considering monthly returns rather than at a higher frequency. From our analysis of our long short exposure, it is evident that we are not exactly market neutral, and some degree of exposure to the market is to be expected since we do not daily rebalance our positions, as the shocks to market liquidity intensifies, we should expect a larger and larger impact by our minor net market exposure.

For the **SMB** factor, we expect some degree of correlation with the IML factor. From the

Table 12: Δ TED - Net Portfolio

This table reports summary statistics for large changes to the funding liquidity, approximated by changes in the TED spread. Returns are given as fully-invested returns after transaction costs. Results are given for TED changes with no lag in Panel A, and lagged TED changes in Panel B. The column names indicate the percentile of changes in the TED spread. For example, percentile 0.95 means that we consider months in which the change in the TED spread was more extreme than the 95th percentile.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Panel A:	No lag									
Mean	0.002	0.0028	0.0022	0.002	0.0016	-0.0001	0.0009	0.0013	0.003	0.0002
Median	0.0027	0.001	0.001	0.0009	0.0008	-0.0008	-0.0006	0.0008	0.0018	0.0022
Std	0.0102	0.0133	0.0118	0.0112	0.0103	0.0178	0.0121	0.0116	0.0125	0.012
T-stat	0.88	1.36	1.45	1.61	1.58	-0.05	0.66	0.93	1.54	0.07
Kurtosis	-0.26	6.4	7.35	6.69	7.95	11.45	2.2	1.85	1.67	-0.78
Skewness	-0.03	1.47	1.45	1.42	1.54	1.09	0.73	0.54	0.6	-0.27
Min	-0.0161	-0.0272	-0.0272	-0.0272	-0.0272	-0.0686	-0.0244	-0.0244	-0.0234	-0.0234
Max	0.0219	0.0574	0.0574	0.0574	0.0574	0.0998	0.0441	0.0441	0.0441	0.0196
Obs	20	41	61	81	101	104	82	65	41	20
Panel B:	Lagged									
Mean	0.0047	0.0007	0.0013	0.0011	0.001	-0.0004	0.0007	0.0006	0.0013	0.0009
Median	0.0025	0.0012	0.0012	0.0012	0.0008	-0.0002	0.0008	0.0009	0.0014	0.0011
Std	0.0151	0.0166	0.0155	0.0144	0.0131	0.0138	0.0094	0.0099	0.0095	0.01
T-stat	1.46	0.29	0.68	0.68	0.77	-0.32	0.63	0.49	0.89	0.41
Kurtosis	6.73	9.37	8.78	9.04	10.84	8.98	0	-0.05	0.23	0.54
Skewness	1.96	-0.88	-0.53	-0.52	-0.52	-1.63	-0.29	-0.36	-0.35	-0.47
Min	-0.0168	-0.0686	-0.0686	-0.0686	-0.0686	-0.0736	-0.0234	-0.0234	-0.0234	-0.0234
Max	0.0574	0.0574	0.0574	0.0574	0.0574	0.0441	0.0196	0.0196	0.0196	0.0182
Obs	22	43	62	81	102	104	82	65	41	20

analysis of size decile exposure, we find that we are loading on small capitalization stocks when markets are dropping, but both on the long and short side. As we are tilting towards holding the smaller stocks, we expect to observe a positive exposure, but this is under the assumption of an equal net long/short exposure of zero.

For the *HML* value factor, we refer to Akbas et al. (2012). They study the time-variation of liquidity risk of value and growth stocks. Their findings suggest that value stocks tend to have higher liquidity betas when markets go down a lot, relative to lower liquidity betas when markets go up a lot. They find the opposite pattern for growth stocks. They argue that across business cycles, investors should substitute lower-risk assets for higher risk assets, which is why selling pressure should increase in value stocks during market downturns. As selling pressure increases, liquidity dries up, and the liquidity premium on value increases. While liquidity dries up in value, investors may seek the lower-risk growth stocks, or simply seek out other low-risk asset classes. This suggests an asymmetric effect in the transition on liquidity, as investors sell their value stocks more aggressively than their holdings of growth stocks. Regardless, the authors find that liquidity worsens more in value relative to growth when markets go down, and the opposite when markets go up. Their findings are therefore consistent with flight to liquidity. Consistent with this research, Petkova et al. (2005) find that value (growth) stocks tend to have higher market betas than growth (value) stocks during bad (good) times. So when markets drop a lot, value is expected to drop more than growth, again consistent with flight to liquidity. We should therefore expect to see a negative loading on the HML factor when markets increase, as these findings would suggest that the pairs trader buys growth stocks in good times, and value stocks in bad times. So our loading on HML should be positive when implied market volatility spikes.

For the momentum factor, *Mom*, on average we expect a negative loading, as the pairs trader is a natural contrarian to the momentum stocks. Pastor and Stambaugh (2003) find that liquidity risk is positively related to momentum in individual stocks on the US market. Asness et al. (2013) find that this relation holds for multiple markets and asset classes divergence is expected to increase. As divergence increases, momentum gains profits, and the pairs trader's losses increase. Daniel et al. (1998) presents a behavioral model to explain momentum, in which they argue that speculators overreact to private information due to overconfidence, and together with self-attribution bias in their reaction given the following publication of information triggers return continuation. Thereby explaining momentum returns. As a consequence, as argued by Avramov et al. (2013), when confidence is high it leads to excesssive trading, meaning liquidity is high. When confidence is low, it leads to low trading, meaning liquidity is low. When we observe VIX spikes, confidence is low and one should for this reason expect momentum returns to drop. Given our hypothesis that the pairs trader receives his signals when the VIX spikes, it implies that we expect diverge. Therefore, the pairs trader should also loose money. With this reasoning, we expect a positive loading on momentum when the VIX spikes, while we should observe the naturally negative loading on momentum during any other sub-period. For shocks to the funding liquidity, we do not see any arguments as to why our expected exposure levels should be any different, as Brunnermeier and Pedersen (2007) point to market liquidity and funding liquidity reinforcing each other.

In table 13 we consider a factor regression on the net portfolio when markets are particularly volatile and we observe shocks to the market liquidity. When VIX goes down a lot, we find, contrary to our hypothesis, significant negative returns. We find a highly significant positive coefficient for the market. This is a potential effect of our choice to not daily rebalance our portfolios. As we do not daily rebalance, a net exposure could develop which would lead to a significant net long or short exposure in times of large volatility. In this case, as VIX is dropping, we should expect markets to increase, and as we have a positive coefficient, it would seem that our exposure is net positive. One could argue from figure 10 that it would seem as if the pairs trader on average is net long.

We also find a significant negative coefficient for the *HML* factor when markets are increasing. These findings are consistent with the findings from Akbas et al. (2012), as explained above. Their findings could provide an explanation for what we observe, under the hypothesis that we are buying the more illiquid stocks. As value has lower liquidity betas when markets go up, while growth has higher liquidity betas, one should expect a tilt towards buying growth stocks and selling value. This should result in a negative *HML* exposure when the VIX is low, and a positive *HML* exposure when the VIX is high. This is precisely what we find in our data in the edge cases for the VIX changes.

For the *Mom* factor, we find negative insignificant loadings when the VIX is dropping. Interestingly, in line with the expectation, when the VIX increases a lot, we observe a highly significant positive exposure to momentum stocks.

When considering the lagged VIX change regression in table 14, the alpha is still economically insignificant, and only for percentile threshold 0.1 do we observe a significant, yet negative, alpha for the pairs trader. Moreover, the explanatory power of our models is very low. For a big decrease in the VIX index, we observe some a negative loading on momentum, as well as a negative loading on the *HML*. While the factor loadings are in line with expectations, the negative alpha is not. Further research could look into the alpha decay for the pairs trader during these market liquidity shocks at a more frequent evaluation period. One could speculate whether the alpha is only generated in very short periods of time after market liquidity

Table 13: Δ VIX - Net Portfolio

This table reports the results from the following 5-factor OLS regression:

$$r^{ex}_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2(Mkt - RF_t) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t}$$

Where $r^{ex}_{net,t}$ is the monthly fully-invested excess return, net of transaction costs, on pairs trading portfolio. The table has parameter estimates and t-Statistics reported in parentheses below. The R^2 and R^2_{adj} are also provided. The t-statistics are computed using Newey-West standard errors with six lags. Results are given for multiple percentile thresholds for month over month changes in the VIX index.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Const	-0.0074	-0.0029	-0.0027	-0.0028	-0.0007	-0.0028	-0.0024	-0.0022	-0.006	0.0026
	(-3.58)	(-1.84)	(-1.55)	(-1.7)	(-0.42)	(-2.19)	(-1.24)	(-1.02)	(-1.59)	(0.36)
IML	0.0004	0.0376	0.0542	0.0231	0.0264	-0.1308	-0.1133	-0.1034	0.0133	-0.0921
	(0.01)	(0.77)	(1.5)	(0.69)	(0.7)	(-1.96)	(-1.76)	(-0.81)	(0.12)	(-0.82)
Mkt-RF	0.1427	0.028	0.0526	0.0525	0.0107	-0.1064	-0.0792	-0.083	-0.0322	0.2486
	(3.53)	(0.39)	(0.94)	(0.97)	(0.21)	(-1.86)	(-1.3)	(-0.84)	(-0.34)	(1.28)
SMB	0.0271	-0.0009	-0.0049	-0.0054	-0.0032	0.0305	0.0233	-0.0086	-0.1912	-0.5216
	(0.21)	(-0.01)	(-0.06)	(-0.11)	(-0.07)	(0.24)	(0.17)	(-0.03)	(-0.63)	(-2.24)
HML	-0.1454	-0.0287	-0.0329	-0.0071	0.0097	0.0568	0.0389	0.0442	-0.0017	0.1456
	(-2.59)	(-0.55)	(-0.86)	(-0.18)	(0.31)	(1.21)	(0.81)	(0.65)	(-0.03)	(1.53)
Mom	-0.0053	-0.0203	-0.0147	-0.0104	-0.0192	-0.0127	-0.0078	-0.0234	0.0745	0.461
	(-0.34)	(-1.52)	(-1.28)	(-0.86)	(-1.39)	(-0.17)	(-0.09)	(-0.21)	(0.88)	(2.71)
R2	0.42	0.03	0.04	0.02	0.02	0.08	0.05	0.06	0.07	0.56
$R2_{adj}$	0.18	-0.13	-0.06	-0.05	-0.04	0.03	-0.02	-0.04	-0.1	0.37

increases. That would explain our results, as the pairs trader is not expected to earn money if all the diverged pairs have converged already.

For a large spike in the VIX, we observe zero alpha and a strong negative loading on the *IML* factor as well as a strong positive loading on the *SMB* factor. While the two are highly correlated, we are sceptical about the sign, and the effects of leaving one out. Both are kept in the analysis as Amihud (2014) reasonably points to the conclusion that the two factors capture something different, but our results are likely affected by a slight degree of multicollinearity given the 0.62 correlation between *IML* and *SMB*. Factor correlations were presented in section 4.6 in table 3. Again, in the edge cases concerning the 0.05 and 0.95 percentile thresholds, the sample size is limited to around 20 observations, which in turn also does raise an issue with concluding concretely based on the resulting output.

Turning our attention to shocks in the funding liquidity, we are now observing slightly more uplifting risk adjusted returns. When the changes in the TED spread falls in the category of being lower than the 10th percentile, the pairs trader earns an mean return of up to 42 bps.

Table 14: Δ VIX Lagged - Net Portfolio

This table reports the results from the following 5-factor OLS regression:

$$r^{ex}_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2(Mkt - RF_t) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t}$$

Where $r^{ex}_{net,t}$ is the monthly fully-invested excess return, net of transaction costs, on pairs trading portfolio. The table has parameter estimates and t-Statistics reported in parentheses below. The R^2 and R^2_{adj} are also provided. The t-statistic of the mean is computed using Newey-West standard errors with six lags. Results are given for multiple percentile thresholds for month over month lagged changes in the VIX index.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Const	0.0004	-0.0021	-0.0019	-0.0014	-0.0008	0.0012	0.0001	-0.0024	-0.0019	-0.0006
	(0.11)	(-2.27)	(-1.71)	(-1.16)	(-0.74)	(0.75)	(0.06)	(-1.49)	(-1.0)	(-0.26)
IML	0.0628	-0.0053	-0.0259	-0.0375	0.0012	-0.1484	-0.1887	-0.2549	-0.1223	-0.3344
	(0.86)	(-0.09)	(-0.75)	(-1.05)	(0.04)	(-1.6)	(-1.68)	(-2.05)	(-1.5)	(-4.61)
Mkt-RF	-0.0669	0.0411	0.0281	0.0367	0.0301	-0.0689	-0.067	-0.0512	-0.041	-0.0228
	(-0.69)	(1.26)	(1.0)	(1.33)	(1.16)	(-1.65)	(-1.63)	(-0.99)	(-0.62)	(-0.79)
SMB	-0.0087	-0.0167	0.0076	0.001	-0.0514	0.1179	0.1532	0.2712	0.2505	0.2631
	(-0.06)	(-0.17)	(0.12)	(0.01)	(-1.1)	(1.07)	(1.26)	(2.34)	(1.92)	(3.19)
HML	0.056	-0.0911	-0.0784	-0.08	-0.0663	0.089	0.0858	-0.0255	-0.0551	-0.077
	(0.59)	(-3.31)	(-2.33)	(-1.89)	(-1.86)	(1.88)	(1.36)	(-0.52)	(-1.16)	(-1.21)
Mom	-0.0823	-0.0325	-0.0308	-0.0093	-0.0207	0.0254	0.0642	0.0169	0.0576	0.0771
	(-5.78)	(-1.17)	(-1.36)	(-0.66)	(-2.33)	(0.46)	(1.07)	(0.29)	(0.85)	(1.77)
R2	0.17	0.22	0.23	0.18	0.12	0.04	0.05	0.1	0.11	0.46
$R2_{adj}$	-0.15	0.1	0.15	0.11	0.06	-0.02	-0.03	-0.0	-0.05	0.21

We quickly note that the R^2 for all models are very low, and the model does not seem to explain returns generated by the pairs trader, however little the pairs trader might make.

We find a strong negative loading on the *SMB* factor when funding liquidity eases, suggesting a loading on the large capitalization segment. When markets drop, we tend to have a stronger tilt on our long leg towards lower capitalization stocks relative to our short leg; we find a difference of around 2 deciles during the 2008 financial crisis. This effect wore off quickly and in mid to late 2009, we end up being net long about 1 decile difference in large-cap. As the loadings were volatile during this liquidity shock period, it is hard to say in general, whether this is conclusively why we observe what we do in terms of the *SMB* tilt. But it would serve as a possible explanation for the 2008 case, where we observe a a rebound in funding liquidity, lower barriers to arbitrage, we momentarily start trading more large capitalization stocks, and therefore we get a positive significant exposure. Further research is needed to attempt to formalize this hypothesis on other sub-periods than just based on the financial crisis of 2008. Our momentum exposure is consistently negative, and only in unique cases statistically significant.

This is consistent with our natural tilt towards a negative momentum loading. The *IML* factor is found to be statistically significant in times of large spikes to the TED-spread. This is in line with our expectation that we tilt towards illiquid stocks when markets drop. We simultaneously and contradictory negative tilt towards the *SMB* factor, which we again must point out, as it could be a function multicollinearity in our regression model.

Table 15: Δ TED - Net Portfolio

This table reports the results from the following 5-factor OLS regression:

$$r^{ex}_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2 (Mkt - RF_t) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t}$$

Where $r^{ex}_{net,t}$ is the monthly fully-invested excess return, net of transaction costs, on pairs trading portfolio. The table has parameter estimates and t-Statistics reported in parentheses below. The R^2 and R^2_{adj} are also provided. The t-statistic of the mean is computed using Newey-West standard errors with six lags. Results are given for multiple percentile thresholds for month over month changes in the TED spread.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Const	0.0019	0.0042	0.0029	0.0023	0.0018	-0.0004	0.0006	0.0011	0.0025	0.0018
	(0.86)	(1.35)	(1.14)	(0.94)	(0.94)	(-0.3)	(0.41)	(0.89)	(1.84)	(1.04)
IML	0.0509	-0.0907	-0.0362	-0.0409	-0.0159	-0.0123	0.1557	0.1662	0.2026	0.1008
	(1.57)	(-0.74)	(-0.36)	(-0.42)	(-0.22)	(-0.11)	(2.04)	(1.96)	(2.07)	(1.19)
Mkt-RF	0.0094	-0.039	-0.0228	-0.0063	-0.0021	0.0117	0.0767	0.0881	0.0958	0.1278
	(0.21)	(-0.59)	(-0.41)	(-0.11)	(-0.04)	(0.28)	(1.95)	(2.2)	(2.27)	(3.92)
SMB	-0.1426	-0.0221	-0.0502	-0.0147	-0.0361	-0.0179	-0.1251	-0.162	-0.2622	-0.1194
	(-3.94)	(-0.26)	(-0.57)	(-0.17)	(-0.6)	(-0.15)	(-1.63)	(-1.82)	(-2.29)	(-0.84)
HML	-0.0748	-0.0469	-0.0696	-0.0264	-0.025	0.1452	0.1083	0.057	0.0325	0.207
	(-1.26)	(-0.77)	(-1.05)	(-0.45)	(-0.63)	(2.38)	(1.42)	(1.05)	(0.59)	(2.16)
Mom	-0.0235	-0.0831	-0.0585	-0.0436	-0.029	-0.0091	-0.0998	-0.0764	-0.0418	-0.0016
	(-1.02)	(-1.58)	(-1.21)	(-1.3)	(-1.33)	(-0.11)	(-2.56)	(-1.68)	(-1.08)	(-0.04)
R2	0.25	0.19	0.12	0.05	0.04	0.06	0.25	0.19	0.22	0.28
$R2_{adj}$	-0.02	0.07	0.04	-0.01	-0.01	0.01	0.2	0.13	0.11	0.03

For the lagged TED-spread changes, we observe highly significant mean returns of 70 bps when the TED-spread was down significantly in the previous month. This could be an effect of a small sample size, and we saw in the summary statistics for this lagged TED-spread, that indeed, the outliers in this edge case are affecting the mean return. We are observing the same pattern with regards to factor loadings as seen for the lagged VIX. The key difference is now that the *IML* and *SMB* factor loading have changed sign. We observe the same odd pattern of loading positively on momentum when the funding liquidity tightened in the past month. Moreover, we seem to load on growth stocks when the change in the TED-spread is more extreme than the 90th percentile. This is inconsistent with the findings of the Akbas et al. (2012)

Table 16: Δ TED Lagged - Net Portfolio

This table reports the results from the following 5-factor OLS regression:

$$r^{ex}_{net,t} = \alpha_i + \beta_1 IML_t + \beta_2(Mkt - RF_t) + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 MOM_t + \varepsilon_{i,t}$$

Where $r^{ex}_{net,t}$ is the monthly fully-invested excess return, net of transaction costs, on pairs trading portfolio. The table has parameter estimates and t-Statistics reported in parentheses below. The R^2 and R^2_{adj} are also provided. The t-statistic of the mean is computed using Newey-West standard errors with six lags. Results are given for multiple percentile thresholds for month over month lagged changes in the TED spread.

Percentile	<0.05	<0.10	<0.15	<0.20	<0.25	>0.75	>0.80	>0.85	>0.90	>0.95
Const	0.007	0.0029	0.0033	0.0025	0.0016	-0.0006	-0.0004	-0.0005	0.0005	-0.0002
	(2.9)	(1.3)	(1.78)	(1.72)	(1.48)	(-0.57)	(-0.38)	(-0.46)	(0.32)	(-0.07)
IML	-0.217	-0.2079	-0.1708	-0.1102	-0.0631	0.0071	0.0004	0.0006	0.1477	0.387
	(-1.48)	(-1.05)	(-1.17)	(-0.95)	(-0.66)	(0.1)	(0.01)	(0.01)	(2.36)	(4.07)
Mkt-RF	-0.0654	-0.0573	-0.0628	-0.0458	-0.036	0.0004	0.0343	0.038	0.0329	0.0067
	(-0.92)	(-0.8)	(-1.05)	(-0.88)	(-0.82)	(0.01)	(1.3)	(1.28)	(0.97)	(0.3)
SMB	0.1942	0.205	0.1684	0.143	0.0976	-0.0407	-0.0804	-0.0909	-0.1758	-0.2995
	(1.61)	(1.33)	(1.36)	(1.42)	(1.13)	(-0.53)	(-1.39)	(-1.36)	(-4.04)	(-3.86)
HML	-0.0342	-0.0076	0.0696	0.0243	0.0069	-0.0075	-0.0214	-0.0278	-0.1067	-0.2461
	(-0.43)	(-0.13)	(1.11)	(0.54)	(0.18)	(-0.2)	(-0.72)	(-0.6)	(-1.85)	(-3.35)
Mom	-0.1807	-0.1595	-0.1285	-0.1335	-0.1012	-0.006	0.0427	0.0368	0.025	0.0466
	(-3.09)	(-1.83)	(-1.91)	(-2.37)	(-2.23)	(-0.2)	(1.91)	(1.47)	(0.59)	(1.67)
R2	0.29	0.18	0.15	0.15	0.11	0.01	0.11	0.13	0.14	0.34
$R2_{adj}$	0.06	0.07	0.07	0.1	0.06	-0.04	0.05	0.06	0.02	0.11

Conclusively we do not observe much evidence of a significant liquidity premium for the pairs trader. We observe a higher risk adjusted return and raw return during sub-periods with market liquidity shocks and funding liquidity shocks relative to the full-period. Yet not a statistically significant return or a raw return of a scale worth pursuing for a statistical arbitrage hedge fund.

5.5.1 Robustness of top pairs performance

In this section, we will continue looking at the periods in which we observe large positive and negative shocks to market and funding liquidity measured as the changes in the VIX index and changes in the TED spread, respectively. Given our earlier observations that the top portfolio may not allow for the optimal amount of liquidity provision as initially hypothesized, we test how robust the results presented in the two earlier sections are to the choice of top pairs portfolios. We consider the top 120 pairs, divided into six portfolios of 20 pairs each, resulting in a portfolio consisting of pairs 0-19, 20-39, 40-59, 60-79, 80-99 and finally 100-119. We will denote this portfolio construction method as 'Top Pair 120/20'. Top 120 in steps of 20. Since we observed a slight degree of an inverse U-shaped in the Sharpe ratio amongst the top portfolios it offers an indication that one should consider portfolios further down the ranks to perform if not equally well, then better than the top 20.

Results in the following figures are all net of transaction costs, and presented for the net portfolio. In total, six figures are presented. Each of the figures presents four separate 3D-plots. On the X-axis we compare the Top pair 120/20 portfolios. This portfolio construction will give insights as to whether the top ranked pairs outperform the lower ranks. Percentile thresholds of changes in the VIX and TED spread, surges and drops, are represented on the Y-axis in percentile form. The Z-axis presents monthly average returns, monthly standard deviation and annualized Sharpe ratios for the net portfolio after transaction costs.

We present both results with contemporaneous changes in the VIX and TED spread, as well as lagged changes. Contemporaneous returns are considered to understand the immediate impact on pairs trading returns, whereas the lagged version presents potential investment opportunities given information about the past month change in the TED spread and the VIX.

Figure 15 **(a)** presents the mean return for the net portfolio given a surge in the VIX. When the VIX spikes, we observe a monthly mean return of the top 20 portfolio of 46 bps after transaction costs, while the 5 other top portfolios generate zero or negative returns. At the same time, in figure 15 **(c)**, we observe that the top portfolio, regardless of being formed based on the lowest *SSD*, seems to have a higher standard deviation during implied volatility spikes relative to the 5 lower ranked portfolios. The resulting sharpe ratio in figure 15 **(e)** estimates an annualized Sharpe ratio of 0.5, during the largest surges in the VIX. These results seem to indicate that only the top pair portfolio 0-19 perform better and better the higher the VIX spike. While profits during spikes in the VIX are contrary to our hypothesis, as previously mentioned, one could still make an argument for the pairs trader earning the liquidity premium relatively fast after the VIX surge eases off. In example, if the VIX jumps 10% by the midpoint of a given month, and drops down 5% by the end of the month, the VIX may look like it's surging, even while the liquidity premium may have been harvested already.

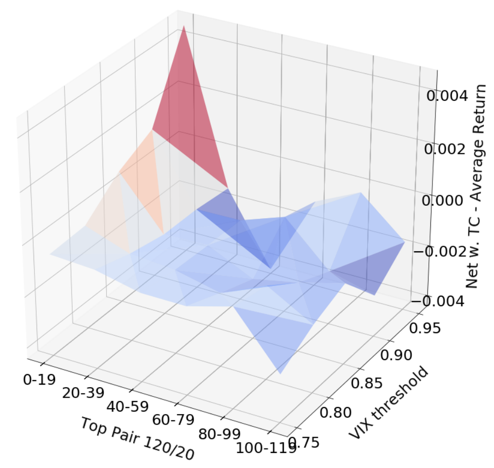
In figures 15 **(b), (d), (f)**, we consider drops in the VIX. We observe a more consistently good performance for all portfolios, with the top mean return being around 40 bps for the 80-99 portfolio. During these extreme sub-periods, Sharpe ratios vary between 0.6 and 1.6, and we do not observe any clear pattern in the standard deviation as to which portfolio is more volatile. We note that the sharpe ratio drops to around 0-0.5 when considering the 0.1 percentile threshold. These findings could suggest that the pairs trader is earning a slight liquidity premium, when the market liquidity increase again, consistent with our initial hypothesis.

If we consider lagged changes in the VIX, in figure 16, we observe an almost identical pattern in the mean return and Sharpe ratio as present in the contemporaneous drops in the VIX. These two sub-periods are considered to be very likely to overlap, as a large drop in the VIX, often follow a large increase in the VIX. As such, we should expect to see a degree of resemblance in our data.

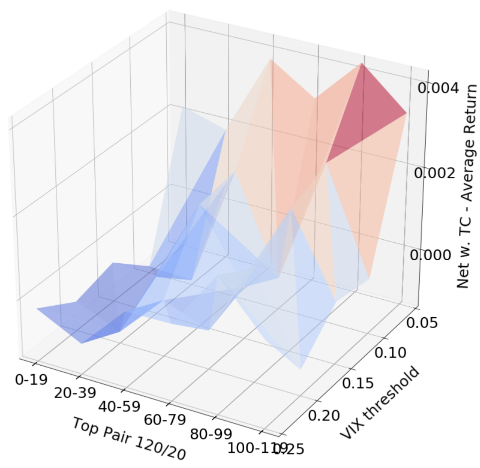
For the lagged drops in the VIX, we observe zero to 40 bps negative subsequent mean returns depending on which top portfolio is considered. These results are not directly linked to our hypothesis, as the pairs trader is not hypothesized to enter a lot of trades after a large drop in the VIX. We are therefore not drawing any conclusions from this section.

In summary, when the VIX increase, it is a hypothesized indicator of market liquidity distress. We find that in those sub-periods, regardless of looking at lagged changes or contemporaneous changes to our portfolio returns, the pairs trader earns a small premium. This can be due to a fast intra-month market liquidity rebounce allowing the pairs trader to earn a quick liquidity premium, or simply a noisy estimate based on too few observations in sample. Similarly the pairs trader earns a small premium when the VIX drops during the same period. As such, we find some, yet weak evidence of a liquidity risk premium in times with shocks to the market liquidity.

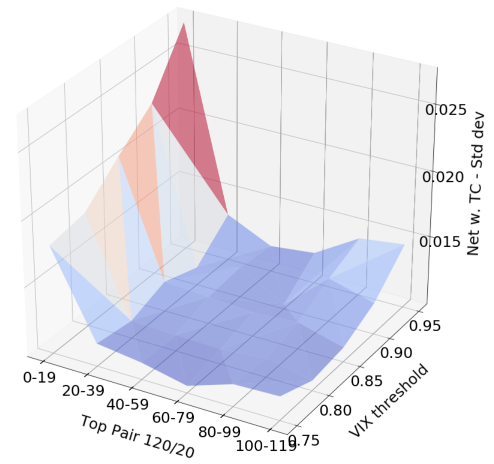
Figure 15: ΔVIX - Net portfolio - Mean, Std, Sharpe w. TC



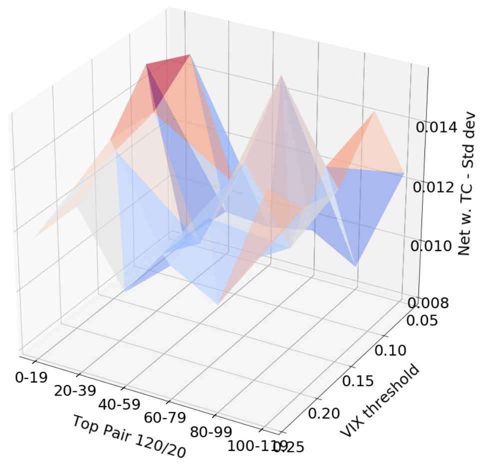
(a) VIX surge - Net portfolio - Mean



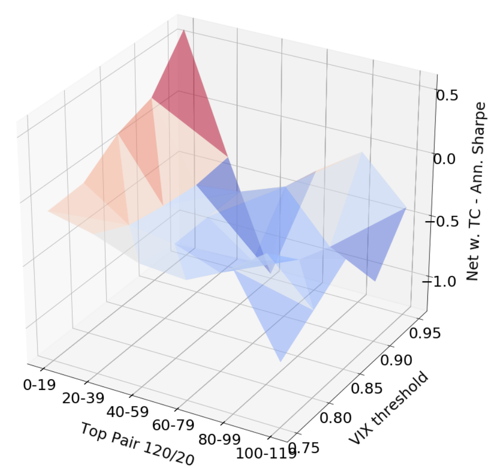
(b) VIX drop - Net portfolio - Mean



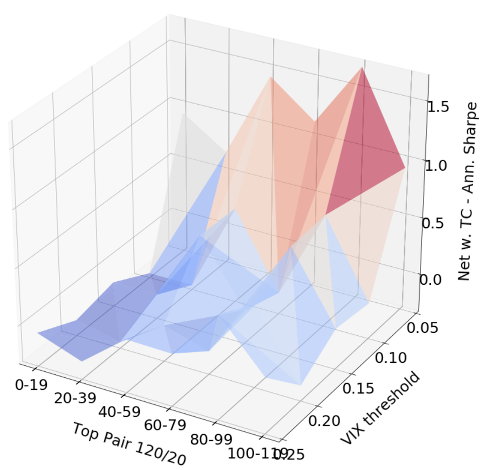
(c) VIX surge - Net portfolio - Std



(d) VIX drop - Net portfolio - Std

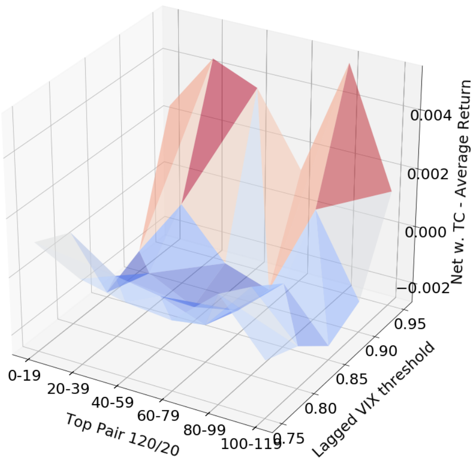


(e) VIX surge - Net portfolio - Sharpe

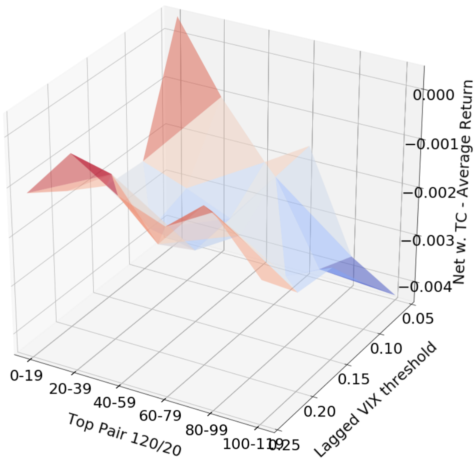


(f) VIX drop - Net portfolio - Sharpe

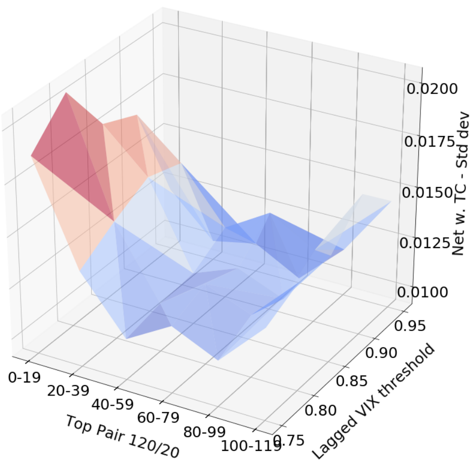
Figure 16: ΔVIX Lagged - Net portfolio - Mean, Std, Sharpe w. TC



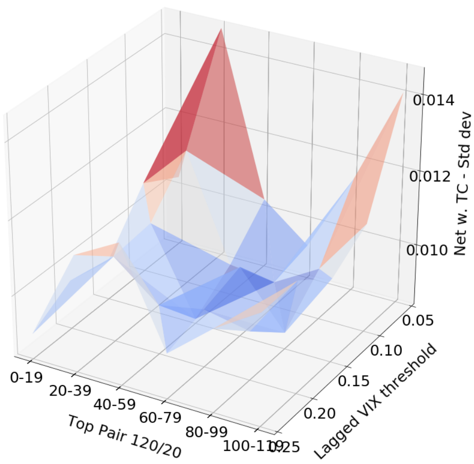
(a) Lagged VIX surge - Net portfolio - Mean



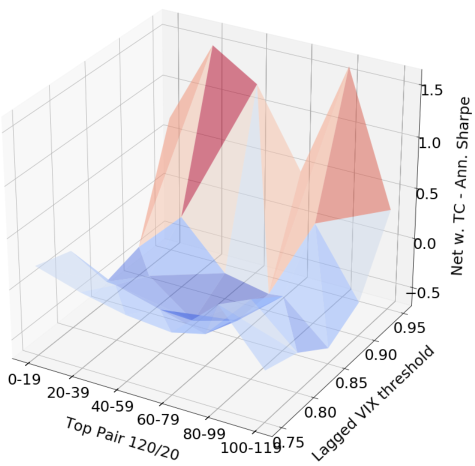
(b) Lagged VIX drop - Net portfolio - Mean



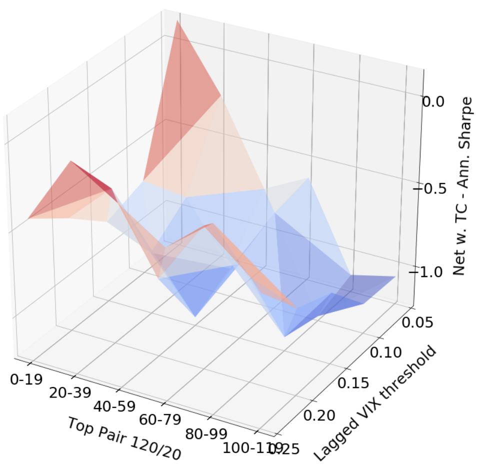
(c) Lagged VIX surge - Net portfolio - Std



(d) Lagged VIX drop - Net portfolio - Std



(e) Lagged VIX surge - Net portfolio - Sharpe

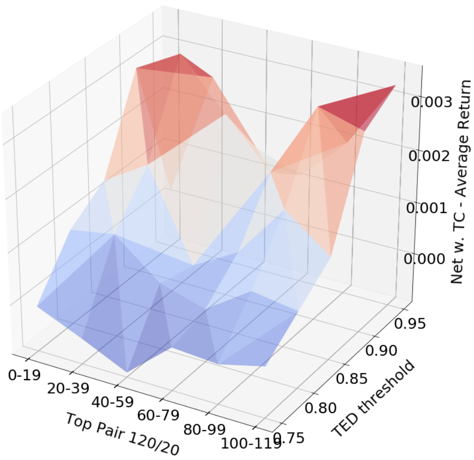


(f) Lagged VIX drop - Net portfolio - Sharpe

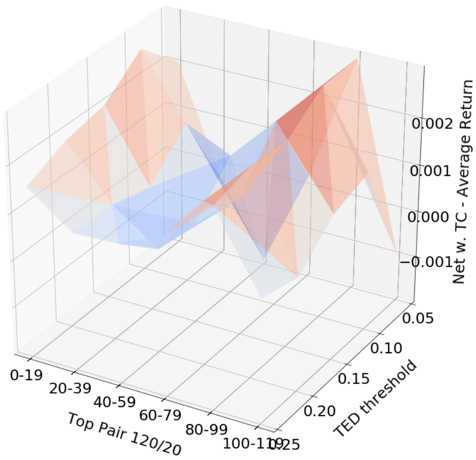
For the funding liquidity, figure 17 shows the non-lagged TED-spread thresholds. We again find some evidence of a small, yet consistently higher risk adjusted return for the pairs trader in sub-periods with high (numerically) positive shocks to the funding liquidity. Our hypothesis will have to reflect the same as the one entertained for the VIX surges. Contrary to the VIX, we do not find any clear pattern of an increasing risk adjusted return with the scale of the drop in the TED-spread. However, the pattern is not entirely clear and is of the same magnitude of the TED surges, so we cannot definitively say that market liquidity matters over and above the funding liquidity in terms of determining the lucrative sub-periods for the pairs trader.

Considering figure 18, which shows the lagged TED-spread thresholds, the mean returns do not display any clear patterns other than the two top portfolios seem to perform well following a month with a large drop in funding liquidity. One could speculate whether funding liquidity takes longer to ease than market liquidity, and as such a premium is earned later after a big drop in the TED-spread. But once again, and in summary, we cannot reach any definitive conclusion of a liquidity premium. We can only observe that the returns are higher, yet not to an statistically nor economically meaningful degree worth pursuing for the pairs trader unless the strategy is paired with another strategy as a hedge to liquidity risk. We will discuss this in more detail in section 6. To reach a more definitive conclusion about whether these returns are compensation for a liquidity premium, we will employ a Liquidity CAPM regression on these particular sub-periods to get a clearer insight as to what component of liquidity risk may be driving the slightly higher returns.

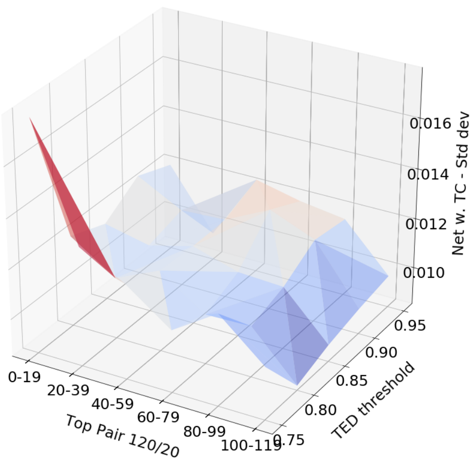
Figure 17: ΔTED - Net portfolio - Mean, Std, Sharpe w. TC



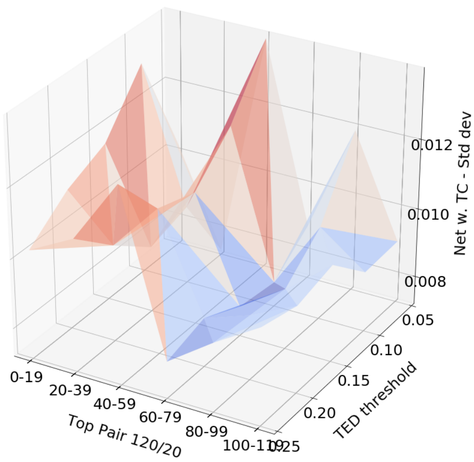
(a) TED surge - Net portfolio - Mean



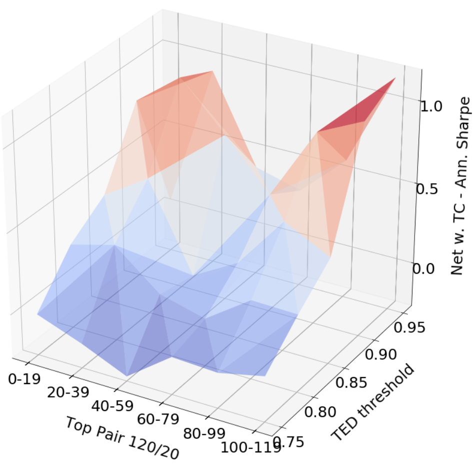
(b) TED drop - Net portfolio - Mean



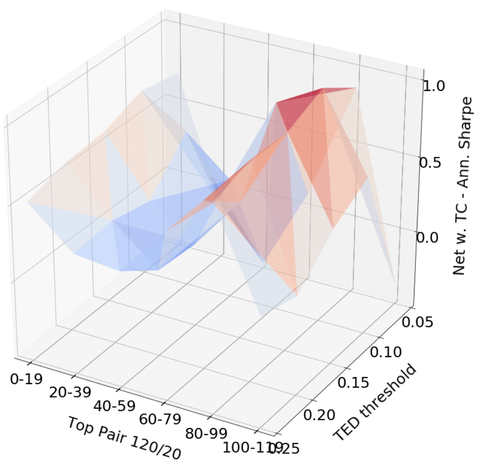
(c) TED surge - Net portfolio - Std



(d) TED drop - Net portfolio - Std

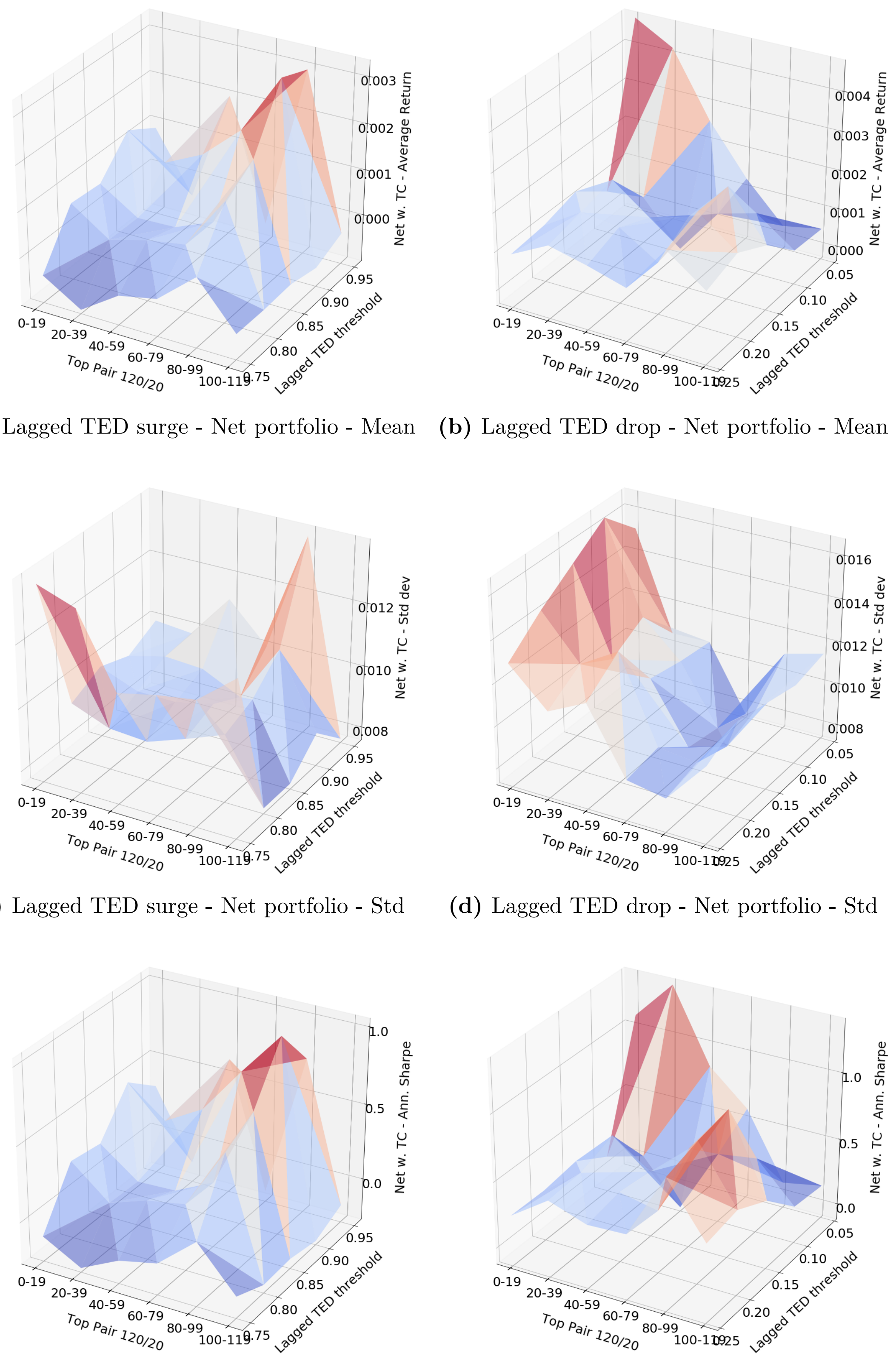


(e) TED surge - Net portfolio - Sharpe



(f) TED drop - Net portfolio - Sharpe

Figure 18: ∆TED Lagged - Net portfolio - Mean, Std, Sharpe w. TC



(a) Lagged TED surge - Net portfolio - Mean **(b)** Lagged TED drop - Net portfolio - Mean

(c) Lagged TED surge - Net portfolio - Std **(d)** Lagged TED drop - Net portfolio - Std

(e) Lagged TED surge - Net portfolio - Sharpe **(f)** Lagged TED drop - Net portfolio - Sharpe

PART IV - Liquidity Adjusted CAPM

5.6 Acharya and Pedersen - LCAPM

Before we can obtain cross sectional regression results, we have to make a choice about how we construct portfolios. As introduced in sections 4.7.1, we make use of the sector portfolios, rather than the AP approach of pairing stocks based on their Amihud *ILLIQ* measure. We justify this choice by arguing that sectors are more suitable to test our hypothesis as they allow for pairing across liquidity levels. We are therefore not forcing our conclusion on our data, but merely providing a better framework in which we can test whether it holds for the simple GGR algorithm.

5.6.1 Sector portfolio performance

We begin by examining some overall sector statistics to see how each sector has performed historically. This section briefly presents sector cumulative returns plots and summary statistics, before and after transaction costs.

Figure 19 displays the cumulative returns before transaction costs for sector portfolios consisting of the top 20 pairs in each sector. We observe a decreasing profitability trend in all sectors. From the unrestricted pairing portfolio we saw that the sector loading was concentrated primarily in Utilities. This loading seems to be beneficial in the early years of the sample period as Utilities is the best performing sector, together with Real Estate. However, from the mid 1990's, shortly before the publication of GGR's findings, Utilities' return flattened out. We speculate that this could be an effect of the popularization of the pairs trading strategy, as Utilities are the main output from the strategy when using the SSD measure to select pairs. The Utilities' cumulative return series draw a clear parallel to the unrestricted pairs cumulative return chart in figure 9. Both return series only exhibit slight deviations during 2000 and 2008 when the sector loading turns from Utilities to Financials. The Real Estate sector portfolio appears much more volatile yet profitable up until 2008. We observe that most sectors suffer slight losses during the financial crisis, yet the Real Estate sector together with Financials, Construction and Communications suffer particularly great losses during the financial crisis, and Financials suffer losses a great deal of time before that.

Figure 19: Cumulative returns for top 20 pairs in sector portfolios

This figure illustrates the cumulative product of fully-invested excess returns before transaction costs for each sector for a sample period running from 1963 - 2019. Each portfolio represents returns for top 20 pairs.

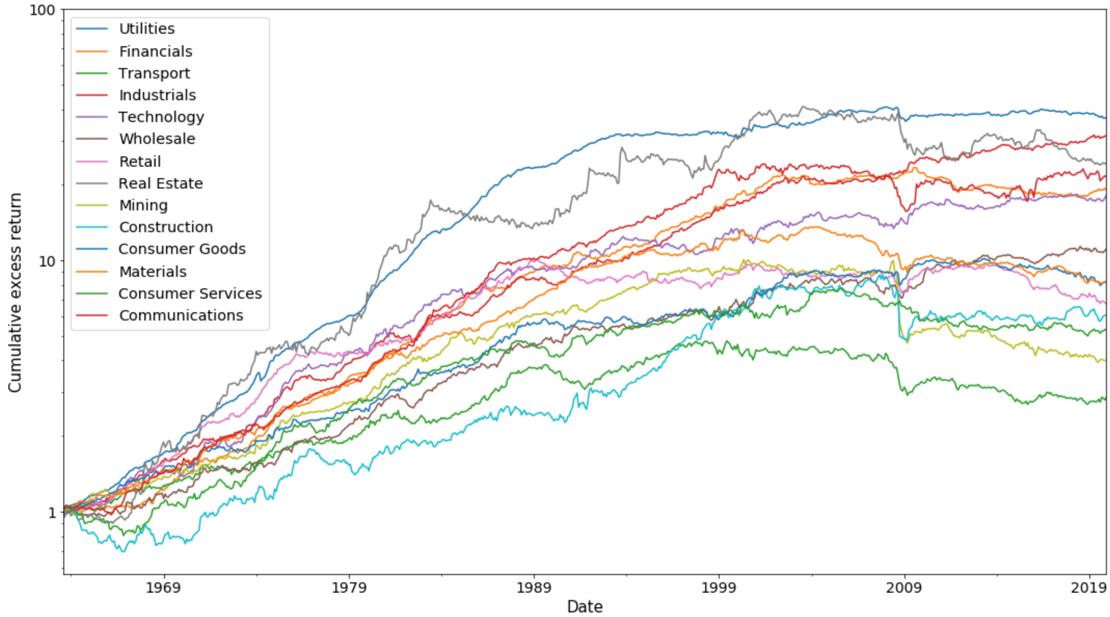


Table 17 reports the average sector mean fully-invested returns, t-Statistics, standard deviation, skewness, kurtosis, min, max and ratio of monthly fully-invested returns below 0. In Panel A, statistics are provided before transaction costs, and in Panel B they are after transaction costs. All statistics are given for the top 20 portfolio in each sector. We quickly note that the mean return post transaction costs are all around zero and insignificant. While Real Estate and Communications are close to being significant at an alpha 5% level with mean returns of 32 and 19 bps per month, respectively. Relative to the full period for the one-portfolio, we observe a 6 bps improvement per month by trading only stocks from either the Utilities or Communications sector.

5.6.2 Liquidity-adjusted CAPM cross sections

In this final section of our analysis, we will consider how the pairs trader is loading on each of Acharya and Pedersen's liquidity risk betas. While we find limited returns for the pairs

Table 17: Summary statistics - Sector performance

This table reports the average sector mean fully-invested returns, *t*-Statistics, standard deviation, skewness, kurtosis, min, max and ratio of monthly fully-invested returns below 0. In Panel A, statistics are provided before transaction costs, and in Panel B the same statistics are provided but after transaction costs. All statistics are given for the top 20 portfolio in each sector.

Sector	Mean	t-Stat	SD	Skew	Kurt	Min	Max	Obs < 0%
Panel A: Pre-TC								
Utilities	0.0062	12.89	0.0126	-0.28	5.24	-0.0804	0.0646	0.3
Financials	0.0052	9.01	0.0151	0.29	3.03	-0.0725	0.0848	0.35
Transport	0.0024	2.59	0.0237	-0.04	1.09	-0.0853	0.0867	0.46
Industrials	0.0059	10.08	0.0152	0.26	1.55	-0.0623	0.0801	0.36
Technology	0.0054	6.38	0.0218	0.36	1.94	-0.0737	0.1113	0.4
Wholesale	0.0046	5.53	0.0216	0.29	1.6	-0.0596	0.1018	0.41
Retail	0.0036	4.86	0.0191	0.26	3.96	-0.07	0.1364	0.42
Real Estate	0.0069	3.92	0.0455	1.2	6.33	-0.1444	0.3287	0.46
Mining	0.0028	2.98	0.0247	-2.12	21.28	-0.2542	0.0874	0.43
Construction	0.0046	3.08	0.0387	-1.92	28.96	-0.4475	0.1968	0.44
Cons. Goods	0.0038	5.52	0.0181	0.11	2.72	-0.104	0.0702	0.42
Materials	0.0038	6.08	0.0163	0.25	3.21	-0.0657	0.0947	0.41
Cons. Services	0.0033	3.86	0.0219	-0.02	2.61	-0.1019	0.0907	0.44
Communications	0.0062	6.16	0.0262	1.93	22.3	-0.1034	0.2925	0.37
Panel B: Post-TC								
Utilities	0.0004	0.9	0.0112	-0.91	11.88	-0.0993	0.0625	0.49
Financials	0.0005	0.81	0.0145	0.19	3.37	-0.0737	0.082	0.51
Transport	-0.0016	-1.8	0.0235	-0.02	1.1	-0.088	0.0811	0.52
Industrials	0.0009	1.57	0.0149	0.16	1.33	-0.0673	0.0652	0.49
Technology	0.001	1.21	0.0213	0.26	2.12	-0.085	0.1101	0.5
Wholesale	0.0007	0.85	0.0214	0.25	1.66	-0.0658	0.0983	0.49
Retail	-0.0009	-1.23	0.0184	0.33	4.78	-0.0715	0.134	0.52
Real Estate	0.0032	1.86	0.0448	1.17	6.51	-0.1494	0.3243	0.49
Mining	-0.0014	-1.51	0.0243	-2.21	22.05	-0.2561	0.0834	0.54
Construction	0.0009	0.57	0.0386	-1.94	28.76	-0.4496	0.1944	0.48
Cons. Goods	-0.0006	-0.92	0.0177	0.06	2.65	-0.106	0.0654	0.52
Materials	-0.0009	-1.41	0.0157	0.35	3.77	-0.067	0.0919	0.55
Cons. Services	-0.001	-1.2	0.0213	-0.17	2.7	-0.1036	0.0876	0.54
Communications	0.0019	1.89	0.0258	2.18	24.77	-0.1051	0.2907	0.46

trader, we see a slight increase in net returns after transactions costs when considering periods of high shocks to market and funding liquidity. We limit our cross-sectional regressions to consider the full-period of returns to obtain more robust beta estimates. We justify this choice by pointing to our findings that the pairs trader is on average net long a more illiquid stock, than a liquid counterpart, as it was noted in section 5.4.3. We also consider cross-sections during our hypothesized lucrative sub-periods with either a high spike or drop in the VIX or TED spread signified by percentile thresholds 0.1 and 0.9. These thresholds are selected to obtain a larger sample size, but maintain the desired large impact on the market- and funding liquidity, the analysis has been done with different thresholds, of 0.05 and 0.15, and both additional thresholds produce results (not reported). These results are provided in the Appendix. We use fully-invested returns net of transaction costs to preserve space.

We follow Acharya and Pedersen's methodology and provide results for 8 separate regressions, all of which are defined in section 4.7.1.

Table 20 presents the average liquidity costs, c, the market beta, the three liquidity risk betas as well as the standard deviation of the *ILLIQ* innovations and net beta for each sector. The table is sorted based on the liquidity level. We observe that the lowest liquidity costs are found in the Utilities sector and the highest are, by far, in the Real Estate sector. Moreover, we find evidence of more illiquid sectors having higher liquidity risk. In particular, an illiquid sector that has a high value of c, also tends to display commonality in liquidity with the market, captured by β_2. Moreover, we find a slight pattern that the liquidity sensitivity to market returns, β_4 tends to be more negative for illiquid sectors. We do not observe much evidence of a relation between the liquidity level and return sensitivity to market liquidity, β_3. We further observe an almost monotonically increasing net beta, and, naturally, the standard deviation of *ILLIQ* innovations are increasing with the liquidity level, as they are bounded by 0. These results are in line with the findings of AP, and they are consistent with the notion of flight to liquidity. We do note that the Financials sector appears to be an anomaly with respect to the rest, by having a relatively large illiquidity level, a positive return sensitivity to market liquidity and liquidity sensitivity to market returns. Compared to earlier results, we note that the Utilities sector, which is dominating our pairs trading one-portfolio, is the least illiquid sector of all sectors presented, and paradoxically the best performing sector as presented in our summary statistics. This could function as an indication of why our liquidity premium appears to be insignificant, as we are mostly trading a highly liquid sector. We should therefore note that the implications of our findings in this section mainly apply to funds with strict sector exposure limit, assuming that they do not daily rebalance, as the mean return across portfolios do not necessarily reflect the return series generated by the simple GGR algorithm pairs trader.

Table 18: Sector Betas - Full-period

This table reports the average liquidity costs, C, the market beta, the three liquidity risk betas as well as the standard deviation of the *ILLIQ* innovations and net beta for each sector. The table is sorted based on the liquidity level, C. Betas are computed using all monthly return and illiquidity observations for each sector portfolio and for an equal-weighted market portfolio.

sector	c	b_1	b_2	b_3	b_4	std inno	b_{net}
Utilities	0.0031	0.0042	0.0423	-0.004	-0.0086	0.0265	0.0591
Communications	0.0048	0.0141	0.0941	-0.0137	-0.0272	0.1722	0.1491
Materials	0.0053	0.0076	0.1366	-0.0084	-0.0145	0.183	0.167
Mining	0.0054	0.011	0.1137	-0.0123	-0.0103	0.1891	0.1473
Consumer Goods	0.0076	0.0063	0.2625	-0.0076	-0.042	0.2467	0.3184
Industrials	0.008	0.0048	0.2257	0.0017	-0.0695	0.2623	0.2983
Transport	0.0081	0.0055	0.3882	-0.0022	-0.0587	0.4143	0.4547
Retail	0.0088	0.0058	0.298	-0.0104	-0.0386	0.2982	0.3528
Wholesale	0.0102	0.0061	0.5198	-0.0133	-0.0598	0.3725	0.599
Financials	0.0104	0.003	0.0886	0.0016	0.002	0.4765	0.088
Consumer Services	0.0125	0.0052	0.5254	-0.0078	-0.1318	0.4585	0.6702
Technology	0.0128	0.0056	0.6159	-0.0056	-0.1036	0.4283	0.7307
Construction	0.0134	0.019	0.8619	-0.0224	-0.1104	0.6028	1.0137
Real Estate	0.0265	0.0107	1.6598	-0.0133	-0.1988	1.0473	1.8827

As AP find strong collinearity amongst their liquidity based portfolios, as well as for individual stocks, we present in table 19 a correlation table reflecting this issue. The collinearity of our liquidity risk measures are confirmed and of the same scale as found by AP. The task of empirically distinguishing between the effects of illiquidity and each of the individual liquidity betas is therefore difficult, as noted by AP.

Table 19: Beta Correlations - Full-period

This table reports the correlations of β_1, β_2, β_3, and β_4 for the 14 sectors formed from 1963 - 2019.

	b_1	b_2	b_3	b_4
b_1	1.0	0.36	-0.85	-0.26
b_2	0.36	1.0	-0.44	-0.92
b_3	-0.85	-0.44	1.0	0.33
b_4	-0.26	-0.92	0.33	1.0

We now turn our attention to examining the liquidity risk betas affect on our sector portfolio expected returns.

Table 20 presents the eight regression models introduced in section 4.7.1. We start by comparing the CAPM model (model 3) to a liquidity-adjusted CAPM (LCAPM) with one degree of freedom (model 1). The LCAPM calculates a net beta, as a function of the market beta and the three liquidity betas, as given in equation 15. By estimating this net beta, we obtain a model with only one risk premium, λ. We find that while the LCAPM is a better fit, both models are very poor fits for the sector portfolio returns, with R^2_{adj} down to 0.027 for LCAPM and 0.011 for the CAPM. Neither the beta for the LCAPM nor the market beta are significant. Given that the returns are after transaction costs, it is no surprise that the constant is insignificant as well. While this alone does not constitute a test for the liquidity risk effect of our liquidity betas over the liquidity level and market risk, we include both the market beta, β_1, the liquidity level, $E(c)$, and the net beta in our regression models in model 4, 5 and 6.

It is evident that as soon as we include the liquidity level as a free parameter, we find strong significance for the κ parameter in model 2 and but not only closely significant to an alpha %5 level for model 5. R^2_{adj} amounts to 0.803 and 0.787 for model 2 and 5, respectively. κ is estimated to be negative, as also found by AP, even though the model implies κ should be positive. Given this finding, we, as done by AP, estimate model 6 with $\kappa = 0$. For this model we find strong significance for the beta net. So we find some weak evidence of the liquidity risk mattering slightly, yet it would seem to that the liquidity level, $E(c)$, matters over the liquidity risk.

While we find a positive market beta coefficient for models with both the market beta and the net beta as independent variables, we note that it does not necessarily reflect a positive risk premium on the market, as the net beta contains the market beta. However, given the low magnitude of the negative net beta, even after subtracting the net beta, we still obtain a positive market risk premium. As illustrated for model 4 below:

$$\begin{aligned} E\left(r_t^p - r_t^f\right) &= 0.00073 + 0.4818 E\left(c_t^p\right) + 0.02433 \beta_1^p - 0.00136 \beta_{net}^p \\ &= 0.00073 + 0.4818 E\left(c_t^p\right) + 0.02297 \beta_1^p - 0.00136 \left(\beta_2^p - \beta_3^p - \beta_4^p\right) \end{aligned} \tag{41}$$

In model 7 and 8 we allow for a risk premium for each separate liquidity beta, and we thus estimate three liquidity risk premiums. We note that none of the models are significant, but that they all suggests a negative loading on the liquidity risk premiums. Given the multicollinearity problems, separating these effects is difficult.

In summary, while our estimated sector liquidity betas are consistent with flight to liquidity, we find little to no evidence that liquidity risk matters over the liquidity level and market risk for pairs trading returns in a sector portfolio setup. This conclusion holds for the full period considered in this section, as well for the cross sectional regressions based on changes in the VIX and TED-spread sub-periods presented in the appendix. In summary, we do not find any evidence of the pairs trader being compensated for taking liquidity risk. In the following section, we discuss why our results are not as convincing as first hypothesized.

Table 20: Liquidity-adjusted CAPM - Full period - Net TC

This table reports 8 cross sectional regression models explaining 14 sector portfolio excess returns by a liquidity-adjusted CAPM model. The data used in the regression reflect months where lagged monthly changes in the VIX is below the 0.1 percentile. Each model consider special cases of the relation

$$E(r_{sec,t}) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p + \lambda \beta_{net}^p$$

The table presents parameter estimates and t-Statistics reported in parentheses below. The t-statistics are calculated using a Fama Macbeth framework, that takes pre-estimation of the betas into account. R^2 and R^2_{adj} are provided in the last two columns.

Model	Constant	E(c)	b_1	b_2	b_3	b_4	b_{net}	R^2	R^2_{adj}
Model 1	0.00088	0.4814					-0.00128	0.202	0.136
	(1.76)						(-1.74)		
Model 2	0.00068	-0.52992					0.00043	0.833	0.803
	(0.64)	(-2.48)					(0.17)		
Model 3	-0.00358		-0.09113					0.017	-0.065
	(-2.0)		(-0.45)						
Model 4	0.00073	0.4814	0.02433				-0.00136	0.207	0.063
	(0.96)		(0.27)				(-1.66)		
Model 5	0.00019	-0.47843	0.04426				-0.0003	0.836	0.787
	(0.11)	(-1.86)	(0.39)				(-0.09)		
Model 6	-0.00248		0.15015				-0.00601	0.78	0.739
	(-2.75)		(1.4)				(-6.17)		
Model 7	0.00094	0.4814	-0.09027	-0.00124	-0.09269	0.00408		0.263	-0.064
	(1.06)		(-0.53)	(-0.47)	(-0.77)	(0.22)			
Model 8	0.00016	-0.41948	-0.08395	-0.00057	-0.11353	0.00659		0.853	0.761
	(0.08)	(-1.41)	(-0.46)	(-0.11)	(-0.83)	(0.32)			

6 Discussion

In this section we will challenge the methodology used in this study, and suggest other approaches in an attempt to further the understanding of where pairs trading return drivers. Each section is ranked given its importance, with the top section being the one we consider the most important.

Modifying the formation algorithm

Our initial hypothesis assumes that the pairing across liquidity levels occurs frequently enough for it to have an impact on returns. Our results indicate that this assumption was overly optimistic. We do observe a slight tilt towards buying the illiquid stock while short selling a more liquid counterpart. But while this pattern is uplifting in relation to our initial hypothesis, the scale to which we are net-long illiquidity during liquidity shock sub-periods does not yield any promising results. This questions whether the simplicity of GGR's algorithm is unsuitable for forming pairs across liquidity levels. We speculate that the desired pairs are not moving close enough together to deliver a low enough sum of squares deviation.

These findings suggest that the simplistic GGR pairs trading algorithm is not, as hypothesized in the introduction, an optimal hedge against liquidity risk for hedge funds seeking to limit their exposure in existing strategies to liquidity shocks. To cater a pairs trading strategy to be a liquidity hedging strategy, future research could focus on a modified pairs trading algorithm that does not focus on pairs selected on a distance measure alone, but rather a combination of metrics. One could force the long leg to be consistently long a higher *ILLIQ* decile relative to the short leg. Do and Faff suggest considering the number of zero-crossings, which is defined as the number of times the normalized time series cross during the formation period. This metric could aid in identifying cross *ILLIQ* decile pairs that may possess the best of both worlds, a combination of stocks moving together that cross frequently while simultaneously having a long illiquid leg and a short liquid leg. It would be interesting to see whether this modification would provide a positive alpha during sub-periods with shocks to liquidity.

Given that we observe a negative correlation between momentum and pairs trading, it would suggest that a pairing of the two strategies could yield interesting results. This method should stand in comparison to Asness et al. (2013). The authors combined momentum and value in a 50/50 portfolio that outperforms the individual strategies. The benefit of this approach, as explained by the authors is that momentum and value benefit from the negative correlation between the two strategies. There is a negative correlation due to momentum being positively correlated to liquidity risk while value is negatively correlated to liquidity risk. We find that the pairs trader loads positively on momentum stocks when the VIX spikes, and that the pairing would therefore not function as a liquidity hedge, but merely attempt to benefit from the general negative correlation.

Rebalancing vs. no rebalancing

Our findings from the factor regressions show a significant market loading, also evidenced in figure 10, where we can see that the strategy suffers from a time-varying net exposure to the market that is different from zero throughout the period 1963-2019.

One option to consider to alleviate this market exposure is to apply daily or periodic rebalancing. However, we would still be facing a trade-off between market exposure and transaction costs, as we found that the major costs affecting the profitability of pairs trading are the trading commissions, so adjusting the positions would most likely only worsen the situation.

GGR (2006), as well as Do and Faff (2010), comment on the top pairs potentially moving too closely together to not even compensate the pairs trader for the transaction costs alone, as deviations are too small to justify the costs of entering the trade. This could explain why the top pairs do not earn high returns after transaction costs. We sought to alleviate some of this effect by filtering our data such that companies with multiple share classes only appeared with one share class in our sample. However, this still does not help to produce significant positive returns after transaction costs. One could try implementing the strategy by trading pairs that are not high in the SSD rank, so that the price spreads to enter a position are big enough to justify the commissions, but this analysis would still be subject to major data-mining to find an optimal rank of pairs to trade.

Even though commissions, market impact and short selling fees are the costs that have the greatest impact on the profitability of pairs trading, the analysis is prone to the lack of application of any other costs. For example, as the strategy is assumed to be self-financing, we do not consider any costs or interests received on margins or collaterals. Moreover, we do not apply any dividend tax to our stocks, which as a long position holder, the trader is liable to pay a dividend tax, whereas when holding a short position he must pay the dividend to the lender and does not receive a tax deduction from dividends on short positions.

Choice of liquidity factor

Amihud (2014) introduced the *IML* factor, which is based on his 2002 illiquidity measure. Given that Amihud lays the foundation for the liquidity CAPM analysis in Acharya and Pedersen's (2005) analysis, we chose this to be our liquidity measure and included the *IML* factor in all our factor regressions. But there were other noteworthy options. Pastor and Stambaugh (2003) criticize Amihud's (2002) ILLIQ measure, as it is volume based, and volume is not necessarily related to illiquidity. As mentioned in our literature review, Pastor and Stambaugh refer to the 1987 stock market crash. Liquidity was at a record low, but trading volume set its record high. While we hope to isolate sub-periods where it is particularly profitable for the pairs trader, Pastor and Stambaugh would argue that the IML factor does not necessarily capture this liquidity premium. Therefore, future research could follow the advice of Pastor and Stambaugh and not use trading activity measures to proxy for time variation in liquidity.

Long and Short Excess performance

As introduced in section 4.4, we attempt to calculate an excess return for both the long and the short leg. Calculating excess returns on a pairs trading portfolio could naively be calculated as the monthly return generated by the pairs traders long portfolio and short portfolio in excess of an equally-weighted market return. However, this method does fall short to the limitations of how long the pairs trader may be active in the market. If the pairs trader is out of the market for even just a single day, a mismatch between the return of the pairs trader and the market occurs. We attempted to alleviate this limitation by adjusting the daily equal-weighted returns and calculating daily excess returns instead of monthly. In this section, we will present our resulting long and short excess mean returns during sub-periods with market liquidity shocks. Shocks to funding liquidity and measures of Sharpe ratios can be found in the Appendix.

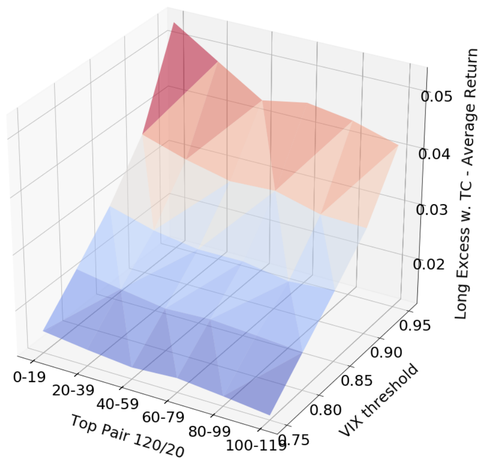
Figure 20 exhibits mean returns when the VIX either surges or drops. We observe that both the stocks in our long portfolio and short portfolio outperform the market during VIX surges by a large margin. The opposite holds for market drops. Given that we are long and short the same notional amount in the market as we are in our long leg and our short leg, we can only speculate about why we observe this pattern, and it would require more research to understand the drivers than the time frame of this study allows for. In essence, it would seem that our long and short stocks exhibit low market betas.

As previously mentioned, Petkova and Zhang (2005) find that value stocks tend to have higher market betas when markets go down. In section 5.5 we explain why we theoretically should stock up on value stocks during sub-periods with a spike in VIX. However, we did not find any evidence of significant loading on the value factor in our factor regression. Therefore, we cannot definitively dispute our findings as an inconsistency with Petkova and Zhang's findings.

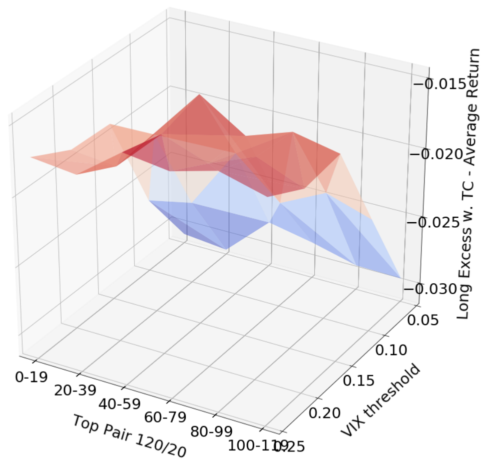
In section 5.4.1 we found that Utilities is the main sector we trade, both on the long and the short leg. This sector is commonly characterized as having stable and low-beta stocks, so this could serve as a possible explanation for our portfolios exhibiting low market exposure. Another possible explanation goes back to our equally-weighted market return approximation. If we still overestimate the equally-weighted market returns, we will naturally subtract a larger market return during market downturns, which could then appear as outperformance by both our long and short portfolios.

In summary, while intriguing, these results fall into the category of future research needed.

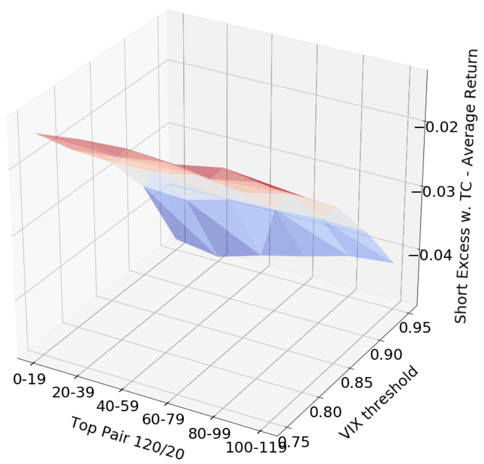
Figure 20: ΔVIX - Long and Short Excess Portfolios - Mean w. TC



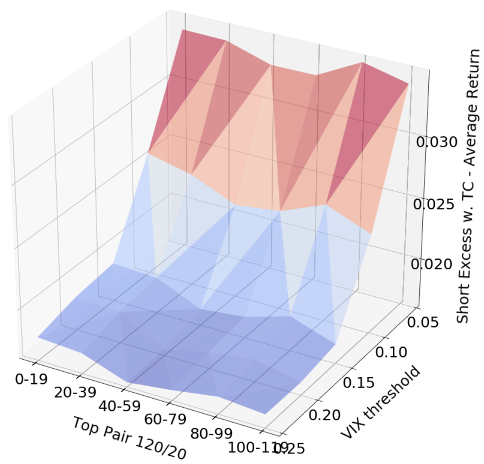
(a) VIX surge - Long Excess - Mean



(b) VIX drop - Long Excess - Mean



(c) VIX surge - Short Excess - Mean



(d) VIX drop - Short Excess - Mean

7 Conclusion

Our initial hypothesis was based on empirical literature that finds persistence in liquidity. Persistence in liquidity implies that returns are predictable. High illiquidity today should predict high expected illiquidity in the following period, which translates into investors requiring larger returns. Therefore, as expected returns should increase with illiquidity costs, it would imply that higher illiquidity costs should result in the investor requiring a larger compensation. A higher expected return leads to a lower price. This was hypothesized to be the pairs trader's trading signal. As the pairs trader provides liquidity during times with shocks to funding and market liquidity, he/she should have a high expected return. From our results it is evident that the liquidity premium obtained by the pairs trader is very limited. We asked four initial research questions, we will answer now on the basis of our analysis.

First, we asked: is the pairs trader a liquidity provider, and therefore subject to liquidity risk? And second: if the answer to the first question is positive, does the level of liquidity provision change over time? In our analysis of loadings on size deciles and the Amihud *ILLIQ* measure, we observe that the pairs trader on average is net long smaller and more illiquid stocks, which would indicate that the pairs trader is a liquidity provider, and therefore subject to liquidity risk. However, the average illiquidity decile traded on the long leg and the short leg is not very different and does perhaps not justify a large liquidity provision premium. Over time, two periods particularly stand out: the dot-com bubble and the 2008 financial crisis. The pairs trader goes from trading an average *ILLIQ* decile of between 3 and 4 (highly liquid stocks) to trading an average between 7 and 8 (highly illiquid stocks). These findings support our conclusion, that flight to liquidity generates signals for the pairs trader in more illiquid stocks. Conversely, we find that the difference between illiquidity loading on the long and short leg does not diverge as much as anticipated. We speculate that the reason for these findings arises during our formation of pairs. If stock A is much more illiquid than stock B, the chance of those two moving together to the extent that they end up in top 20 in the distance approach is overly optimistic. This does not dispute the hypothesis, but it challenges the rather simplistic pairing algorithm's ability to actively select pairs with a large difference in liquidity level, yet with the same underlying drivers of returns. During our robustness test of the choice of which top pairs portfolio to trade, we found that the top portfolio performed slightly worse than the subsequent portfolios. These findings could be a weak indication of lower ranked pairs being able to reflect a liquidity risk premium trade better than a top pair. In summary, the current pairs formation algorithm allows for minimal exploitation of liquidity provision in times of liquidity shocks. On average it would appear we do provide liquidity, but the average difference between the long and short leg illiquidity deciles loading is minimal, and compensation for raking this liquidity risk must therefore be limited.

Third, we asked: is the pairs trader compensated higher in sub-periods with high barriers to arbitrage? As Do and Faff, we find that the pairs trader makes an insignificant amount after accounting for transaction costs. We focus our analysis on sub-periods signified by large shocks to market liquidity and funding liquidity captured by the VIX index and TED spread, respectively. For the full period, we generate a loss each month as a pairs trader, net of transaction costs. By isolating our trading to months where we observe liquidity shocks, results are improved to earning between 20 and 47 bps per month. The average return of 47 bps are yielded when the lagged changes in TED spread is plunging. One could argue that a big drop in the TED spread should indicate that a further improvement of funding liquidity should cause our pairs to converge as the barriers to arbitrage are lowered. In summary, we find that, net of transaction costs, the pairs trader does receive a slight degree of compensation in terms of higher raw returns, but none of which is significant in an economic or statistical sense. These findings are consistent with our hypothesis that the pairs trader does obtain a higher return during months characterized by high barriers to arbitrage. Yet the results are not statistically and hardly economically significant enough to yield a clear conclusion that this is compensation for providing liquidity.

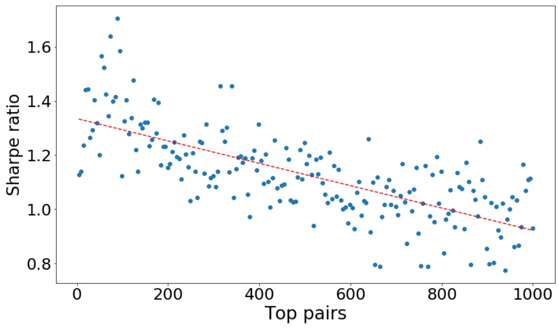
Finally, we asked whether a liquidity CAPM model indicates whether the pairs trader is compensated for taking liquidity risk. From our Liquidity CAPM analysis, we could from the summary statistics deduce that an illiquid sector, denoted as a sector with a high value of c, tends to display higher commonality in liquidity with the market, captured by β_2, relative to a liquid sector. Moreover, we find a slight pattern that the liquidity sensitivity to market returns, β_4 tends to be more negative for illiquid sectors. However, we do not observe much evidence of a relation between the liquidity level and return sensitivity to market liquidity, β_3. The signs of our liquidity beta estimates are consistent with those found by Acharya and Pedersen (2005), and they serve as evidence of flight to liquidity. However, from the cross sectional regressions considering the full sample period, no evidence was found that the pairs trader is compensated for taking liquidity risk

In summary, we conclude that the simple GGR algorithm for forming pairs does not allow the pairs trader to take sufficient liquidity risk to generate the hypothesized higher return. Returns are seemingly all captured by transaction costs, and the pairs trader is only mildly improving results by trading the strategy during months characterized by liquidity shocks.

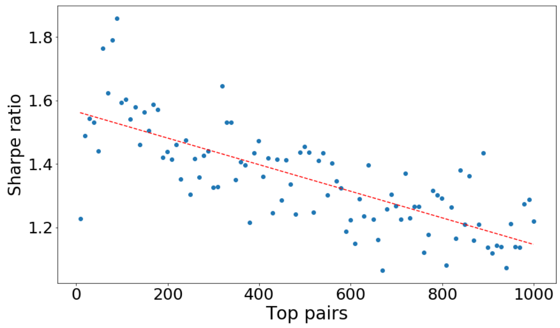
8 Appendix

Figure 21: Sharpe Ratios of Top-1000 pairs divided in sub-portfolios - Committed Capital

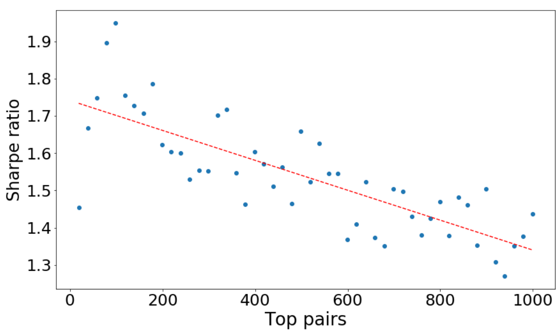
This figure shows the Sharpe ratios of the committed capital returns for different combinations of the number of pairs in each portfolio, up to the pair ranked 1000 in the *SSD* metric, with pair 1 being the pair with the lowest *SSD* during the formation period. Sub-figure A, shows the Sharpe ratios of portfolios comprised of 5 pairs each, with the left-most blue dot representing the metric for trading only the first 5 pairs with the lowest SSD, and the right-most blue dot indicating the Sharpe ratio of the portfolio which only trades pairs ranked between the 996th to 1000th in the SSD metric. The same explanation applies to the other sub-figures, which show the results of forming portfolios of 10, 20 and 50 pairs. The dashed red line is the fitted line that minimizes the residuals of all the observations.



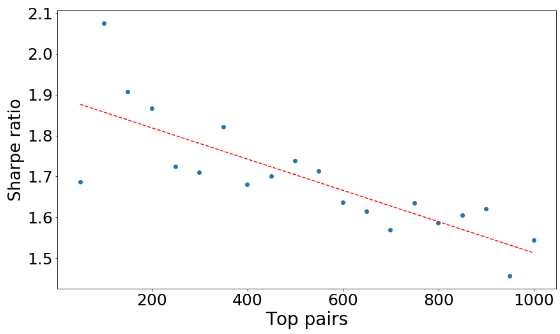
(a) Steps of 5 pairs



(b) Steps of 10 pairs



(c) Steps of 20 pairs



(d) Steps of 50 pairs

Figure 22: Market illiquidity innovations

This figure illustrates the historic standardized market illiquidity innovations and stands as a comparison to Acharya and Pedersen's (2005) Figure 1. The series is standardized using the mean and standard deviation using the full sample up until 2019.

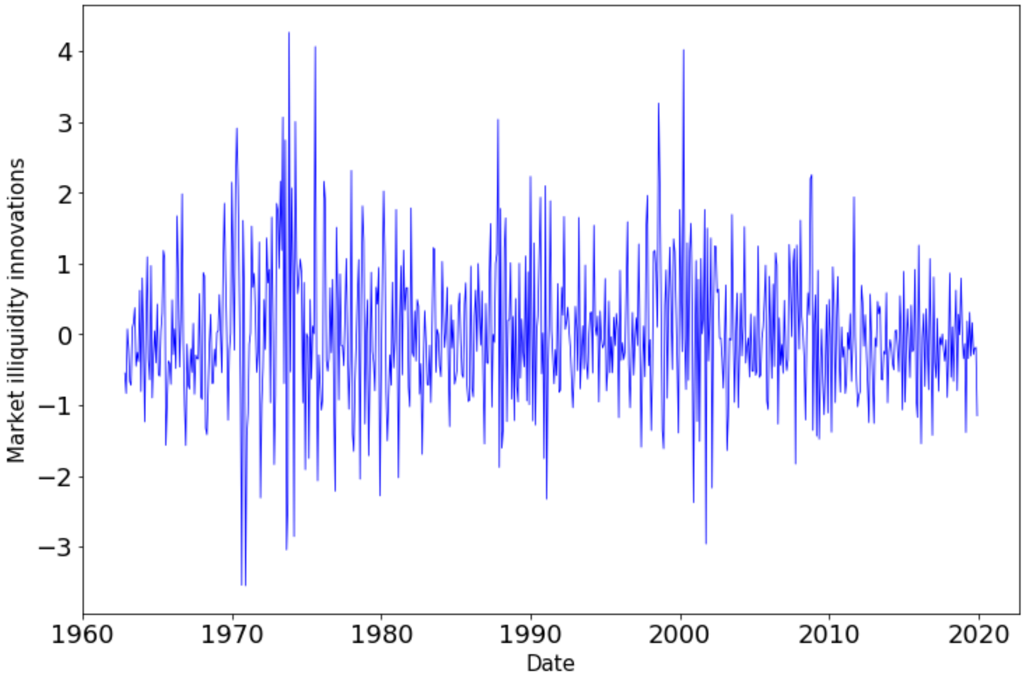


Table 21: Sector Portfolio Size and Formation

Sector	Number of Companies	HSICIG
Utilities	310	490-499
Financials	1250	600-649 and 670-679
Transport	273	400-479
Industrials	1620	330-359 and 370-399
Technology	503	360-369
Wholesale	352	500-519
Retail	639	520-599
Real Estate	165	650-669
Mining	626	100-149
Construction	148	150-179
Consumer Goods	588	0-99 and 200-239
Materials	1131	240-329
Consumer Services	1239	700-999
Communications	205	480-489

Table 22: Liquidity-adjusted CAPM - Δ VIX 0.1 - Net TC

This table reports 8 cross sectional regression models explaining 14 sector portfolio excess returns by a liquidity-adjusted CAPM model. The data used in the regression reflect months where monthly changes in the VIX is below the 0.1 percentile. Each model consider special cases of the relation

$$E(r_{sec,t}) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p + \lambda \beta_{net}^p$$

The table presents parameter estimates and t-Statistics reported in parentheses below. The t-statistics are calculated using a Fama Macbeth framework, that takes pre-estimation of the betas into account. R^2 and R^2_{adj} are provided in the last two columns.

Model	Constant	E(c)	b_1	b_2	b_3	b_4	b_{net}	R^2	R^2_{adj}
Model 1	0.00596	0.4814					-0.00152	0.098	0.023
	(4.14)						(-1.14)		
Model 2	0.0065	-0.58252					-0.00041	0.509	0.42
	(2.66)	(-2.45)					(-0.23)		
Model 3	0.00352		-0.2113					0.144	0.073
	(1.11)		(-1.42)						
Model 4	0.01091	0.4814	-0.26635				-0.00133	0.494	0.403
	(5.38)		(-2.94)				(-1.28)		
Model 5	0.01271	-0.71103	-0.27681				0.00039	0.742	0.665
	(4.58)	(-3.83)	(-3.01)				(0.28)		
Model 6	0.00533		-0.19547				-0.00301	0.363	0.248
	(1.78)		(-1.46)				(-1.95)		
Model 7	0.01052	0.4814	-0.30889	-0.00126	-0.05918	0.00193		0.584	0.399
	(4.48)		(-3.22)	(-1.19)	(-1.21)	(0.21)			
Model 8	0.01165	-0.64949	-0.30899	0.00016	-0.05839	4e-05		0.785	0.651
	(3.86)	(-3.21)	(-3.16)	(0.12)	(-1.18)	(0.0)			

Table 23: Liquidity-adjusted CAPM - Δ VIX 0.9 - Net TC

This table reports 8 cross sectional regression models explaining 14 sector portfolio excess returns by a liquidity-adjusted CAPM model. The data used in the regression reflect months where monthly changes in the VIX is above the 0.9 percentile. Each model consider special cases of the relation

$$E(r_{sec,t}) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p + \lambda \beta_{net}^p$$

The table presents parameter estimates and t-Statistics reported in parentheses below. The t-statistics are calculated using a Fama Macbeth framework, that takes pre-estimation of the betas into account. R^2 and R^2_{adj} are provided in the last two columns.

Model	Constant	E(c)	b_1	b_2	b_3	b_4	b_{net}	R^2	R^2_{adj}
Model 1	-0.00825	0.4814					-0.00059	0.012	-0.071
	(-4.24)						(-0.38)		
Model 2	-0.00527	-0.82978					0.00117	0.502	0.412
	(-1.72)	(-3.18)					(0.67)		
Model 3	-0.01393		-0.00114					0.0	-0.083
	(-5.27)		(-0.01)						
Model 4	-0.00805	0.4814	-0.01776				-0.00042	0.017	-0.162
	(-3.67)		(-0.24)				(-0.23)		
Model 5	-0.00364	-0.92703	-0.06677				0.00213	0.537	0.398
	(-1.0)	(-3.23)	(-0.87)				(1.02)		
Model 6	-0.01327		0.02967				-0.00183	0.054	-0.118
	(-4.72)		(0.31)				(-0.79)		
Model 7	-0.01079	0.4814	0.28364	-0.0077	0.14888	-0.03424		0.157	-0.218
	(-2.68)		(0.68)	(-1.06)	(0.68)	(-1.09)			
Model 8	-0.00686	-0.89298	0.27313	-0.00473	0.16952	-0.03248		0.589	0.332
	(-1.37)	(-2.8)	(0.68)	(-0.65)	(0.8)	(-1.07)			

Table 24: Liquidity-adjusted CAPM - Δ TED 0.1 - Net TC

This table reports 8 cross sectional regression models explaining 14 sector portfolio excess returns by a liquidity-adjusted CAPM model. The data used in the regression reflect months where monthly changes in the TED-spread is below the 0.1 percentile. Each model consider special cases of the relation

$$E(r_{sec,t}) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p + \lambda \beta_{net}^p$$

The table presents parameter estimates and t-Statistics reported in parentheses below. The t-statistics are calculated using a Fama Macbeth framework, that takes pre-estimation of the betas into account. R^2 and R^2_{adj} are provided in the last two columns.

Model	Constant	E(c)	b_1	b_2	b_3	b_4	b_{net}	R^2	R^2_{adj}
Model 1	7e-05	0.4814					-0.0007	0.022	-0.059
	(0.05)						(-0.52)		
Model 2	0.00293	-0.7815					0.00164	0.703	0.649
	(1.43)	(-4.66)					(1.07)		
Model 3	-0.00445		-0.07029					0.09	0.014
	(-1.98)		(-1.09)						
Model 4	0.00075	0.4814	-0.04696				3e-05	0.1	-0.063
	(0.51)		(-0.98)				(0.02)		
Model 5	0.00347	-0.78008	-0.0389				0.00224	0.724	0.642
	(1.61)	(-4.6)	(-0.87)				(1.33)		
Model 6	-0.00422		-0.04086				-0.00194	0.141	-0.015
	(-1.83)		(-0.55)				(-0.81)		
Model 7	0.00065	0.4814	-0.20561	-0.00532	-0.1645	-0.01109		0.55	0.35
	(0.57)		(-2.16)	(-2.12)	(-2.23)	(-1.97)			
Model 8	0.00138	-0.57855	-0.19446	-0.00298	-0.15587	-0.00927		0.814	0.698
	(0.59)	(-2.87)	(-1.84)	(-0.74)	(-1.84)	(-1.33)			

Table 25: Liquidity-adjusted CAPM - Δ TED 0.9 - Net TC

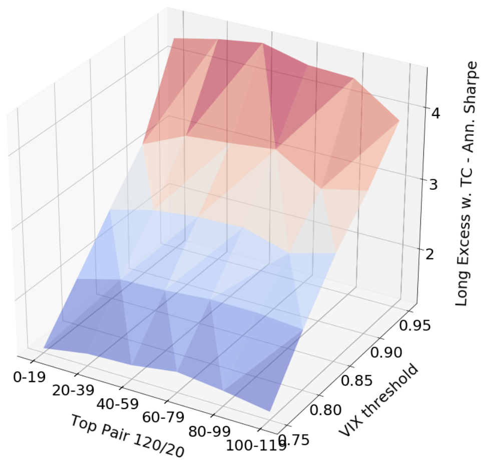
This table reports 8 cross sectional regression models explaining 14 sector portfolio excess returns by a liquidity-adjusted CAPM model. The data used in the regression reflect months where monthly changes in the TED-spread is above the 0.9 percentile. Each model consider special cases of the relation

$$E(r_{sec,t}) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta_1^p + \lambda_2 \beta_2^p + \lambda_3 \beta_3^p + \lambda_4 \beta_4^p + \lambda \beta_{net}^p$$

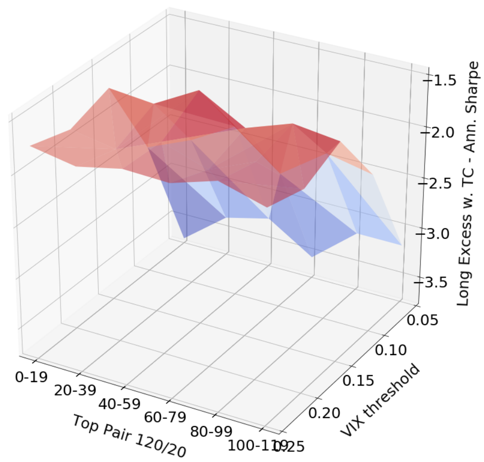
The table presents parameter estimates and t-Statistics reported in parentheses below. The t-statistics are calculated using a Fama Macbeth framework, that takes pre-estimation of the betas into account. R^2 and R^2_{adj} are provided in the last two columns.

Model	Constant	E(c)	b_1	b_2	b_3	b_4	b_{net}	R^2	R^2_{adj}
Model 1	0.00346	0.4814					-0.00575	0.178	0.11
	(1.89)						(-1.61)		
Model 2	0.00326	-0.48808					-0.00445	0.487	0.394
	(1.23)	(-1.98)					(-1.01)		
Model 3	-0.00419		0.07346					0.013	-0.07
	(-1.92)		(0.39)						
Model 4	0.00381	0.4814	-0.04802				-0.00588	0.187	0.039
	(1.75)		(-0.34)				(-1.58)		
Model 5	0.00378	-0.51157	-0.04647				-0.00433	0.492	0.339
	(1.16)	(-1.91)	(-0.3)				(-0.94)		
Model 6	-0.00071		0.03894				-0.00924	0.307	0.181
	(-0.29)		(0.24)				(-2.16)		
Model 7	0.00448	0.4814	-0.05722	-0.0066	-0.08836	0.00431		0.327	0.027
	(1.89)		(-0.36)	(-1.74)	(-1.22)	(0.27)			
Model 8	0.00203	-0.18178	-0.02237	-0.00853	-0.11398	0.00995		0.585	0.326
	(0.57)	(-0.47)	(-0.14)	(-1.47)	(-1.27)	(0.51)			

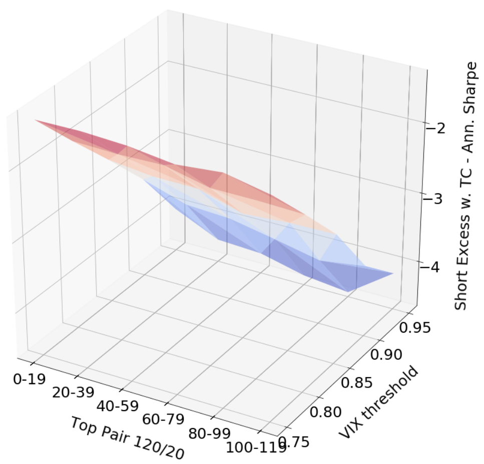
Figure 23: ΔVIX - Long and Short Excess Portfolios - Ann. Sharpe w. TC



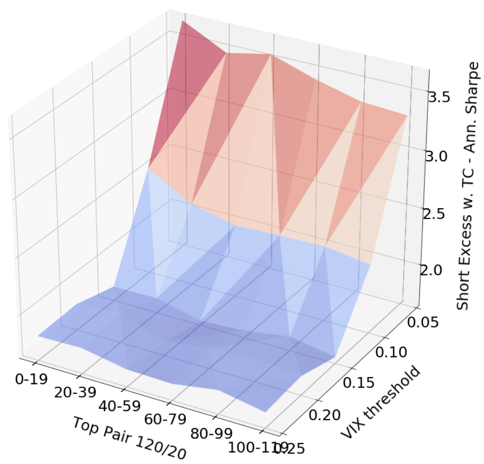
(a) VIX surge - Long Excess - Ann. Sharpe



(b) VIX drop - Long Excess - Ann. Sharpe

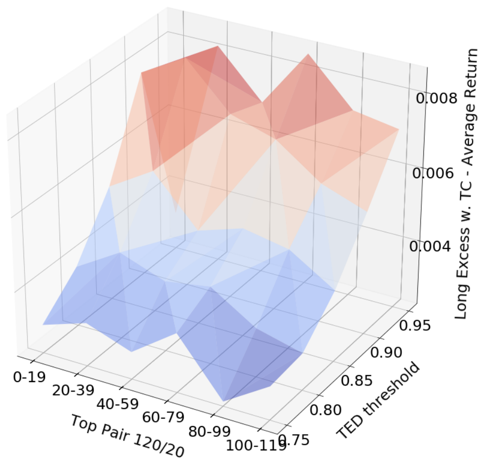


(c) VIX surge - Short Excess - Ann. Sharpe

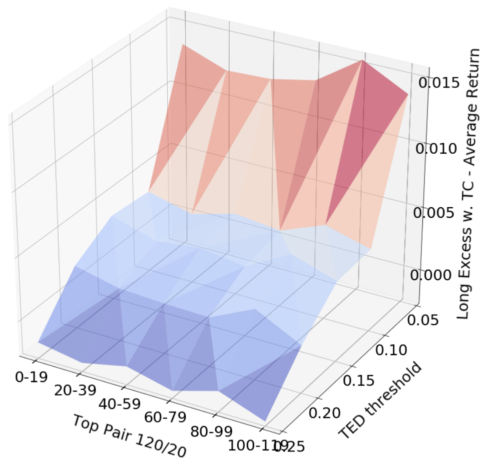


(d) VIX drop - Short Excess - Ann. Sharpe

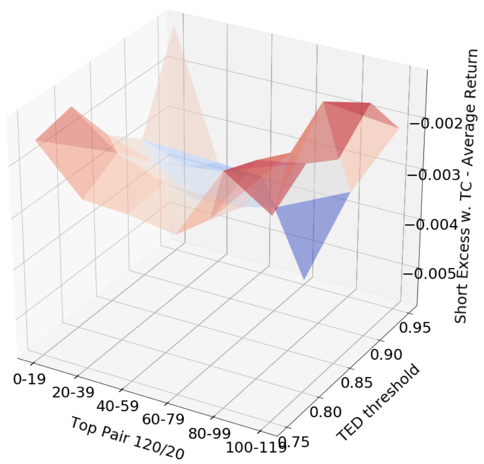
Figure 24: ΔTED - Long and Short Excess Portfolios - Mean w. TC



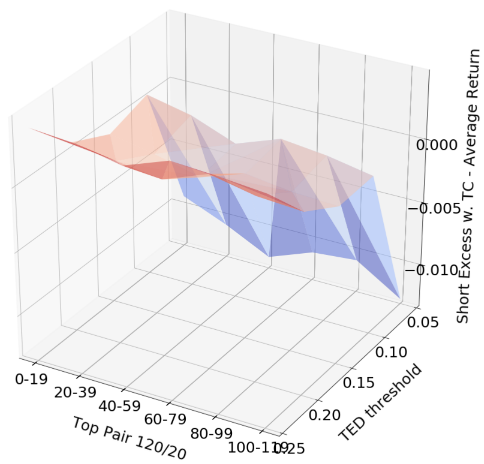
(a) TED surge - Long Excess - Mean



(b) TED drop - Long Excess - Mean

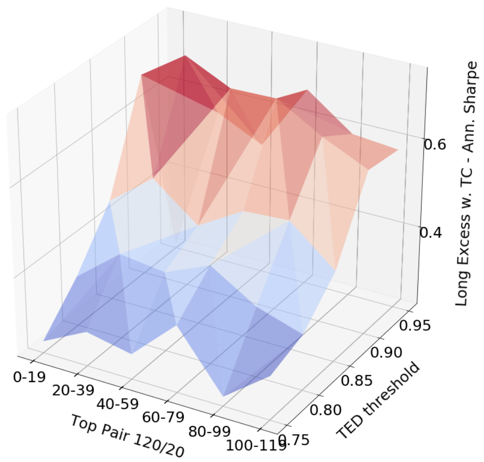


(c) TED surge - Short Excess - Mean

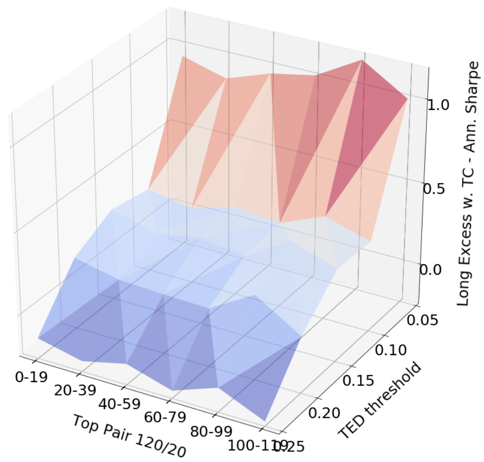


(d) TED drop - Short Excess - Mean

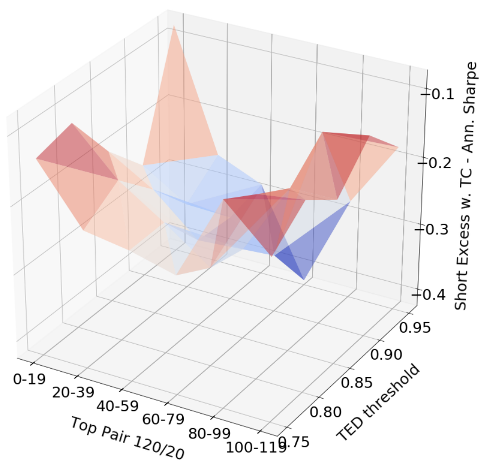
Figure 25: ΔTED - Long and Short Excess Portfolios - Ann. Sharpe w. TC



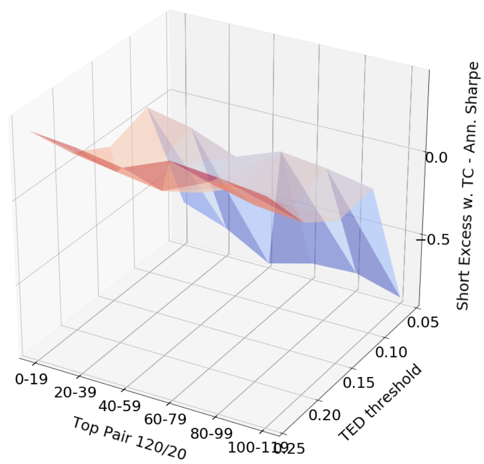
(a) TED surge - Long Excess - Ann. Sharpe



(b) TED drop - Long Excess - Ann. Sharpe



(c) TED surge - Short Excess - Ann. Sharpe



(d) TED drop - Short Excess - Ann. Sharpe

9 References

Acharya, V. V. and L. H. Pedersen

2005. Asset pricing with liquidity risk. *Journal of Financial Economics*.

Akbas, F., E. Boehmer, E. Genc, and R. Petkova

2012. The Time-Varying Liquidity Risk of Value and Growth Stocks. *SSRN Electronic Journal*.

Amihud, Y.

2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Amihud, Y.

2014. The Pricing of the Illiquidity Factor's Systematic Risk. *SSRN Electronic Journal*.

Amihud, Y. and H. Mendelson

1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–249.

Anderson, A.-M. and E. A. Dyl

2005. Market Structure and Trading Volume. *Journal of Financial Research*, 28(1):115–131.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen

2013. Value and Momentum Everywhere.

Avellaneda, M. and J. H. Lee

2010. Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7):761–782.

Avramov, D., S. Cheng, and A. Hameed

2013. Time-Varying Momentum Payoffs and Illiquidity. *SSRN Electronic Journal*.

Basu, S.

1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics*, 24(1):3–37.

Brennan, M. J. and A. Subrahmanyam

1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, 41(3):441–464.

Brown, J., D. K. Crocker, and S. R. Foerster

2007. Trading Volume Liquidity and Investment Styles. *SSRN Electronic Journal*.

Brunnermeier, M. and L. H. Pedersen

2007. Market Liquidity and Funding Liquidity. Technical report, National Bureau of Economic Research, Cambridge, MA.

Caldeira, J. and G. V. Moura

2013. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. *SSRN Electronic Journal*.

Canina, L., R. Michaely, R. Thaler, and K. Womack

1998. Caveat Compounder: A Warning about Using the Daily CRSP Equal-Weighted Index to Compute Long-Run Excess Returns. *The Journal of Finance*, 53(1):403–416.

Carhart, M. M.

1997. On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82.

Chordia, T., R. Roll, and A. Subrahmanyam

2000. Commonality in liquidity. *Journal of Financial Economics*, 56(1):3–28.

Chordia, T., R. Roll, and A. Subrahmanyam

2001. Market Liquidity and Trading Activity. *The Journal of Finance*, 56(2):501–530.

Chordia, T., R. Roll, and A. Subrahmanyam

2002. Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, 65(1):111–130.

Chung, K. H. and C. Chuwonganant

2014. Uncertainty, market structure, and liquidity. *Journal of Financial Economics*, 113(3):476–499.

Cummins, M. and A. Bucca

2012. Quantitative spread trading on crude oil and refined products markets. *Quantitative Finance*, 12(12):1857–1875.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam

1998. Investor Psychology and Security Market Under- and Overreactions. *Journal of Finance*, 53(6):1839–1886.

Datar, V. T., N. Y. Naik, and R. Radcliffe

1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2):203–219.

Do, B. and R. Faff

2010. Does simple pairs trading still work? *Financial Analysts Journal.*

Do, B. and R. Faff

2012. Are pairs trading profits robust to trading costs? *Journal of Financial Research.*

Engelberg, J., P. Gao, and R. Jagannathan

2009. An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity.

Fama, E. F. and K. R. French

1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and K. R. French

1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and J. D. MacBeth

1973. Risk, Return, and Equilibrium: Empirical Tests.

Gatev, E., W. Goetzmann, and K. G. Rouwenhorst

1999. Pairs Trading: Performance of a Relative Value Arbitrage Rule. Technical report, National Bureau of Economic Research.

Gatev, E., W. Goetzmann, and K. G. Rouwenhorst

2006. Pairs trading: Performance of a relative-value arbitrage rule.

Harris, M. and A. Raviv

1993. Differences of Opinion Make a Horse Race.

Haugen, R. A. and N. L. Baker

1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439.

Huck, N.

2009. Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, 196(2):819–825.

Huck, N.

2010. Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3):1702–1716.

Ibbotson, R. G., Z. Chen, and W. Y. Hu

2012. Liquidity as an Investment Style. *SSRN Electronic Journal*.

Jegadeesh, N. and S. Titman

1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65.

Jones, C. M.

2002. A Century of Stock Market Liquidity and Trading Costs.

Lintner, J.

1965. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1):13.

Liu, W.

2006. A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, 82(3):631–671.

Lo, A. and J. Wang

2000. Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. Technical report, National Bureau of Economic Research, Cambridge, MA.

Malkiel, B. G.

2003. The efficient market hypothesis and its. In *Journal of Economic Perspectives*, volume 17, Pp. 59–82.

Markowitz, H.

1952. Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Mossin, J.

1966. Equilibrium in a Capital Asset Market. *Econometrica*, 34(4):768.

Nagel, S.

2012. Evaporating Liquidity. Technical report, The Review of Financial Studies, Cambridge, MA.

Newey, W. and K. West

2014. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Applied Econometrics*, 33(1):125–132.

Pastor, L. and R. F. Stambaugh

2003. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Petkova, R., L. Zhang, W. E. Simon, A. Ang, M. Barclay, G. Bauer, K. Daniel, W. Ferson, A. Karolyi, L. Kogan, J. Lewellen, C. Mackinlay, B. Schwert, and J. Warner

2005. Is value riskier than growth? *Journal of Financial Economics*, 78:187–202.

Sharpe, W. F.

1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3):425–442.

Shumway, T.

1997. The Delisting Bias in CRSP Data. *The Journal of Finance*.

Subrahmanyam, A.

2010. The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research? *European Financial Management*, 16(1):27–42.