Common Short Selling and Excess Comovement*

Marco Valerio Geraci†, **Jean-Yves Gnabo**‡, and **David Veredas**⋆

†*Research Department, National Bank of Belgium*

‡*CeReFiM & NaXys, Université de Namur*

⋆*Vlerick Business School & Ghent University*

February 15, 2022

Abstract

We show that common short sold capital can explain future four-factor excess return correlation one month ahead, controlling for many pair characteristics, including similarities in size, book-to-market, and momentum. We explore the possible explanations that could give rise to this result. Contrary to the predictions of price pressure, we find that the relationship weakens significantly with stock illiquidity. Instead, consistent with the informed trading hypothesis, the relationship is stronger when short positions originate from hedge funds, from active investors, and from short sellers with high past performance. Stocks connected by common short sellers are associated with non-transitory negative cumulative abnormal returns. Finally, we show that our results can be used to obtain diversifications benefits.

Keywords: short selling, comovement, hedge funds

JEL Classification: G12, G14

*The views expressed in this paper are those of the authors and do not necessarily reflect the view of the National Bank of Belgium or of the Eurosystem. The authors gratefully acknowledge the support of the Cambridge-INET Institute and the Communauté française de Belgique (Projet d'Actions de Recherche Concertées grant 13/17-055). We thank two anonymous referees, as well as Laurent Barras (discussant), Christiane Baumeister, Sophie Béreau, Oscar Bernal, Catherine Dehon, Pelin Ilbas, Oliver Linton, Tarik Roukny, Davy Paindaveine, Pedro Saffi, and Kristien Smedts (discussant), and conference and seminar participants at the University of Cambridge, the 2018 Belgian Financial Research Forum, the 2019 International Forecasting Symposium, the 2018 and 2019 International Association of Applied Econometrics Conference, the 2019 European Economic Association Meeting in Manchester, the 2018 Paris December Finance Meeting, and the 2019 Econometric Society Winter Meeting for helpful comments and suggestions.
Corresponding author email: marcovalerio.geraci@nbb.be.

1 Introduction

A large part of the cross-sectional variation of stock return correlation can be explained by few fundamental factors. For example, equity returns of companies from closely related sectors and countries tend be more correlated than the equity returns of companies from more distant sectors and countries. Another fundamental factor is firm size—Huberman et al. (1988), among others, have shown that the equity returns of similarly-sized firms are more closely correlated than the equity returns of firms of different size.

Apart from fundamental factors, research has shown that supply and demand dynamics of equity markets have a major role in determining the cross-sectional variation of the correlation of equity returns. Pindyck and Rotemberg (1993) have shown that the excess comovement of equity returns can be in part explained by institutional ownership. More recently, Antón and Polk (2014) and Bartram et al. (2015) have found that mutual fund flows and active reallocations of funds can trigger comovement of equity returns above and beyond classic fundamentals.

Whereas previous studies have focused on the relationship between long positions and comovement of asset returns, in this study, we explore the role of short selling positions for comovement. Specifically, we exploit new data on short positions disclosed to the Financial Conduct Authority (FCA) of the United Kingdom (UK) to construct a measure of common short selling.[1] The measure, which is intuitive and easy to compute, captures the strategies of short sellers taking net negative positions against multiple stocks. This can be calculated only if we know the identity of short sellers, which the disclosure data provide. We use this new measure to predict (in-sample) the future excess comovement of equity returns.

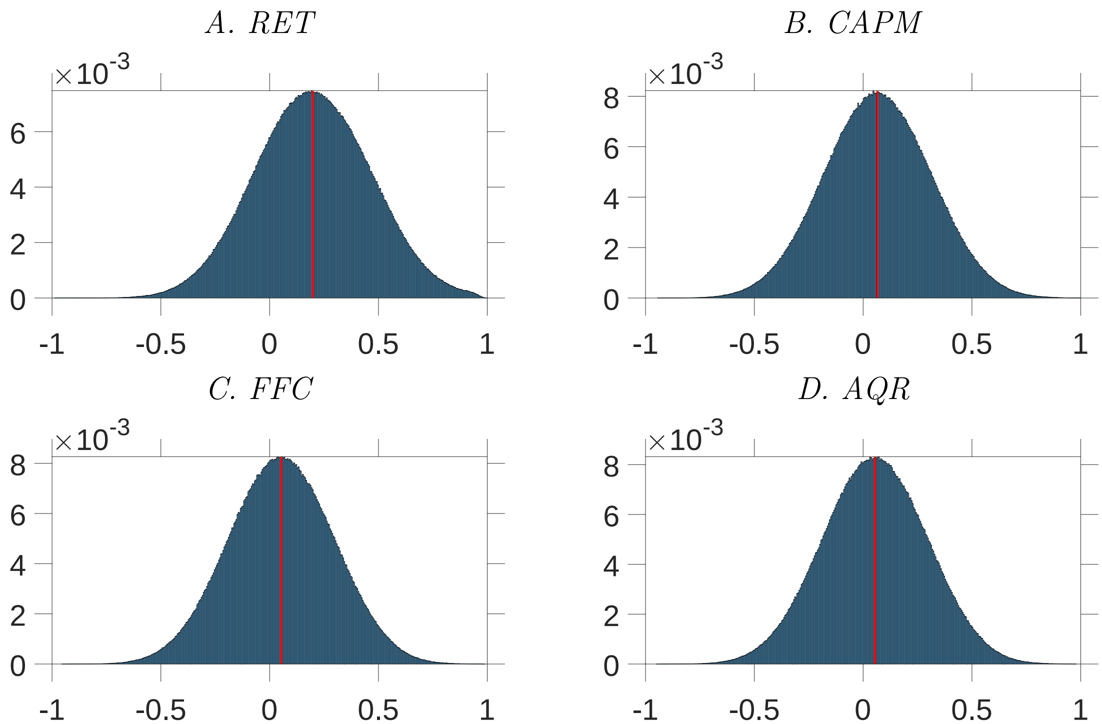
By definition, excess comovement is the part of correlation that isn't explained by common factors, such as the Fama and French (1993) value and size factors and the Carhart (1997) momentum factor. For the stocks in our sample, Figure 1 presents the distribution of the within-month pairwise correlation of daily raw returns, as well as the distribution of excess correlation, computed from the daily residual returns from different factor models.[2]

[1] Our data on short selling disclosures cover large net short positions, above the European regulatory threshold of 0.5% of company capital. We focus on stocks listed on the London Stock Exchange (LSE).

[2] Our sample covers 356 stocks traded on the LSE that have at least one public short selling disclosure reported by the UK's FCA.

The average correlation of raw returns is 20%, whereas the four-factor excess correlation i.e., the correlation of daily residuals from a Fama-French-Carhart factor model, averages 5.4%. Excess correlation remains sizeable and is extremely variable across the cross-section of stock pairs.

Figure 1: Distribution of pairwise return correlations, Jan. 2013—Dec.2019



This figure shows the distribution of the monthly realised correlation of daily returns for stock-pairs from an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. Chart A depicts the distribution of monthly realised correlation of raw daily returns. Charts B through D depict the distribution of monthly realised correlation of residual returns from alternative factor models. Specifically, residual returns are computed using: (B) the local CAPM; (C) the local Fama and French (1993) model augmented with the local Carhart (1997) momentum factor (FFC); (D) the local AQR model, which is the FFC model augumented with local betting-against-beta (Frazzini and Pedersen, 2014) and quality-minus-junk (Asness et al., 2018) factors. Local factors are associated to the country (or region) of stocks. Factor data are from AQR's website, whereas stock price data, used to compute returns, are from Refinitiv EIKON.

Because correlation is a key input to the portfolio allocation problem, understanding and trying to anticipate correlations and excess correlations is important for asset managers. Engle and Colacito (2006), among others, have shown that mispecification of correlation can lead to sizeably lower portfolio returns. Relatedly, predicting correlations is important for risk management and hedging purposes, as well as for pricing derivatives such as correlation swaps and index options.

We find that the amount of common short sold capital can predict four-factor residual return correlation one month ahead, controlling for common ownership and common analyst coverage, as well as for similarities in size, book-to-market, momentum, and several other common characteristics. In our most flexible specification, a standard deviation increase in common short sold capital is associated with a future rise of 2.9% of the average excess comovement for a given stock pair.

The results of our paper are also of interest to regulators. On several occasions, particularly during the aftermath of the financial crisis, short selling has been banned for fears that, in a declining market, it might exacerbate downward price spirals (Finnerty, 2005). Our unique data set allows us to verify two hypotheses explaining the uncovered relationship, including the possibility that short sellers induce comovement through price pressure. In the second part of the paper, we analyse this hypothesis in more depth and draw implications for financial stability policy.

The effect of illiquidity helps us shed light on the role of price pressure in the relationship between common short selling and excess comovement. According to the theoretical studies of Brunnermeier and Pedersen (2005) and Cont and Wagalath (2013), price pressure by short sellers can induce shifts in correlation and this effect should be stronger the higher the stock illiquidity. If the positive relationship that we uncover were stronger for the most illiquid stock pairs, it would bring further evidence in support of the price pressure mechanism.

Quite to the contrary, we find that the association between common short selling and future excess correlation weakens significantly for the most illiquid stock pairs. This result should be somewhat reassuring for regulators, as it does not support, at least directly, the prediction that liquidity conditions could lead to contagion by short selling. Still, we cannot completely exclude that the price pressure story occurs in ways that our framework cannot pick up. We discuss these alternative interpretations, including the possibility that short sellers adopt stealth trading techniques, in more details in Section 5.2.

Next, we verify the role of informed trading for the relationship between common short selling and excess correlation. Studies have shown that short sellers are sophisticated market agents, who trade on the basis of superior information and are able to predict future stock price movements (Boehmer et al., 2008, Diether et al., 2009b, Boehmer et al., 2018). By shorting several stocks, short sellers expect future price declines. As declines occur, they should coincide with higher correlation between the shorted stocks.

We exploit the valuable advantage of our data that allows us to identify short sellers and classify them according to several traits. We find that the effect of common short selling is more predictive of future excess correlation when it originates from informed agents, such as hedge funds, active investors, with high turnover and high concentration, and short sellers with high past performance. This indicates that informed trading has a role in explaining the relationship between common short selling and excess correlation.

Finally, we analyse portfolios of stocks that are connected through common short sellers. First, we find that connected portfolios are associated with negative four-factor abnormal cumulative returns that do not revert in the short run and persist for several months after the portfolio construction. Because price declines are non-transitory, this result lends further support to the information hypothesis explaining the link between common short selling and excess comovement. Second, we compare the realised volatility, over 253 trading days, of similar connected and non-connected portfolios, matched by size-deciles. We find that, on average, compared to their matched connected counterpart, portfolios of stocks that do not share any common short sellers have 13.6% lower volatilities. This shows that the uncovered relationship between common short selling and excess comovement can reveal diversification opportunities for portfolio managers.

We contribute to a growing body of literature that makes use of short selling disclosure data (Boehmer et al., 2018, Jones et al., 2016, Jank et al., 2021). Previous studies have used this data to analyse the behaviour of short sellers and the relation between short positions and underlying stock returns. In contrast, we use this data to study the relation between short selling and comovement.

The disclosure data are partially censored, such that only large short positions are observable. However, compared to previous short selling data, disclosure data come with at least two advantages. First, rather than proxying for short selling, such as short selling indicators constructed from securities lending data, the disclosure data cover actual net short positions submitted by short sellers to the regulator. Second, the data allow us to identify short sellers taking the short positions, which is crucial for constructing our common short selling measure and for our analysis of informed trading. Alternative data, such as, for instance, short interest data, capture the aggregate (market-wide) level of short selling of stocks. This would not allow to retrieve information on common short positions, which we show are useful to explain the commonalities of stock returns.

The rest of the paper is organized as follows. In Section 2, we describe the short selling disclosure data and the sample construction. In Section 3, we outline our regression setting. In Section 4, we present the results showing the predictive power of common short selling for forecasting future excess correlation. In Section 5, we investigate the role of illiquidity and informed trading for our results. In Section 6, we analyse price reversals and diversification opportunities for connected portfolios. We draw our conclusions in Section 7.

2 Data and Sample

2.1 UK Short Selling Disclosure Data

According to EU regulation N. 236/2012, ratified in November 2012 by the European Parliament and the European Council, every financial subject detaining a net short position above 0.2% of shares outstanding of a company is required to disclose their position to the competent market authority—the FCA, in the UK. Furthermore, any short position that passes the threshold of 0.5%, and every change by 0.1% after that, has to be disclosed publicly. Public disclosures include the name and ISIN of the shorted share, the name of the short seller, and the quantity short sold, in terms of percentage of shares outstanding. Compared to other short selling data, such as short interest data, short disclosures are actual net short positions obligatorily submitted to the regulator and, therefore, are subject to attentive scrutiny. In calculating their net short selling position, short sellers are required to include synthetic short positions obtained through options.

We collected all publicly disclosed short selling positions that were published in the UK on the FCA's website between November 2012 and December 2019. We focus on short selling disclosures published by the UK's FCA for three main reasons. First, as will become clear in the next sections, the study involves a large effort to match data across different sources. To keep this exercise manageable, we decided to focus the scope of the paper on a single country of those comprised by the EU short selling disclosure requirement. Second, the UK has a highly-recognised and active stock market, guaranteeing reliable stock and company coverage from the data sources used in our study. Third, as evidenced by Jones et al. (2016), a large part of EU short selling disclosures occur in the UK.[3]

In the remainder of this section, we will briefly describe the raw FCA public short selling disclosure data, whereas in the next section we will construct our final matched data sample.

The disclosures involve 657 unique stocks and 454 different short sellers. Most of the stocks are of UK companies of all sectors. Table 1 shows the summary details for the collected disclosure data.

Panel A of Table 1 summarizes the information given by public disclosures of short positions to the FCA. During the sample period, 60,099 disclosures were made, which included 8,016 position originations (i.e., the first disclosure of a net short position above 0.5% of the shares outstanding), 44,503 updates (i.e., any increments or decrements of 0.1% of the shares outstanding after the 0.5% threshold), and 7,580 position terminations (i.e., disclosures under the 0.5% and representing the closing of the short position).

The number of disclosures increased steadily over the years of our sample, with the exception of 2019.[4] This may suggest that, over time, short sellers became more active and/or that they became more accustomed to the new disclosure regulation.

Panel B of Table 1 presents additional descriptive statistics regarding the disclosure data. The upper part of the panel shows that, on average, short sellers take position on about five different stocks per year. The standard deviation is quite large, with some short sellers taking position on as many as 116 different stocks over one year. The median holding period length of a disclosed short position is of 28 trading days, whereas the average holding period is of 81.7 trading days. This average reflects some short positions being held for long periods.

[3]We checked that this is still valid today. For the sample period, UK disclosure data involve 649 unique stocks (reduced to 374 after matching and cleaning). For the same period, French disclosure data cover 173 unique stocks, German disclosure data cover 331 stocks, and Italian disclosure data cover 168 stocks.

[4]Note that we only observe data for the last two months of 2012.

Table 1: Descriptive statistics of public short selling disclosure made to the FCA, Nov. 2013—Dec. 2019

Panel A: Number of Disclosed Positions, Originations, Stocks, and Short Sellers

Year	Disclosures	Originations	Stocks	Short Sellers
2012	793	323	165	106
2013	4489	617	261	159
2014	5151	717	262	162
2015	7167	1008	279	185
2016	9301	1232	317	214
2017	10751	1384	321	224
2018	12557	1587	355	229
2019	9890	1148	357	203

Panel B: Summary Statistics for Stocks and Short Sellers

Variable	Year	Mean	Med.	S.D.	Min	Max
# of stocks per short seller	2012	2.9	1	5.6	1	53
	2013	4	2	8.3	1	80
	2014	4.7	2	8.6	1	75
	2015	5.1	2	10	1	89
	2016	5.3	2	11.7	1	116
	2017	5.7	2	13	1	102
	2018	6.5	2	14.4	1	113
	2019	6.3	2	13.5	1	103
# of short sellers per stock	2012	1.8	1	1.7	1	12
	2013	2.5	1	2.4	1	14
	2014	2.9	2	2.7	1	15
	2015	3.4	2	3.4	1	18
	2016	3.6	2	3.7	1	23
	2017	4	2	4.3	1	29
	2018	4.2	2	4.6	1	26
	2019	3.6	2	3.5	1	18

The table reports short selling disclosure public data collected from the website of the UK's FCA. Panel A shows the number of disclosed position and the number of disclosures that were originations of a short position. Panel B shows the summary statistics regarding the number stocks and short sellers involved in the disclosure data.

2.2 Stock Sample Construction

From the initial set of 657 stocks that had at least one disclosed short selling position, we construct the sample for our study. First, to make sure that the sample is composed of comparable and sufficiently liquid stocks, we only consider those stocks that are primary shares of companies traded on the London Stock Exchange (LSE).[5]

Table 2: Stock sample construction

Cleaning step	N. Stocks
1. Stocks with at least one FCA short sale disclosure	657
2. Remove non-LSE stocks	470
3. Remove non-common shares	469
4. Remove stocks with more than 50% missing price data	374
5. Match with ownership data	356

Next, for these stocks and for the time period covered, we searched for historical price data and company information from Refinitiv EIKON. To compute the control variables for our regression analysis, we also require ownership data from Refinitiv EIKON and analyst earnings estimates from the Institutional Brokers Estimate System (IBES).

Finally, we require stocks to have a price for 50% of the trading days from Oct. 2007 to Dec. 2019. This allows to have a steady number of observations throughout the sample period.[6] Table 2 summarises the steps for the sample construction.

After these restrictions the final matched sample involves 356 stocks and 43,393 disclosed short selling positions. Figure 2 summarizes the matched sample according to the Thompson Reuters Business Classification (TRBC) economic sector code. The sector with the most stocks was Industrials with 80 stocks, whereas the Cyclical Consumer Goods & Services sector had the most disclosures and short sellers. As outlined in the next section, sector information is used to control for similarities across stocks in our model.

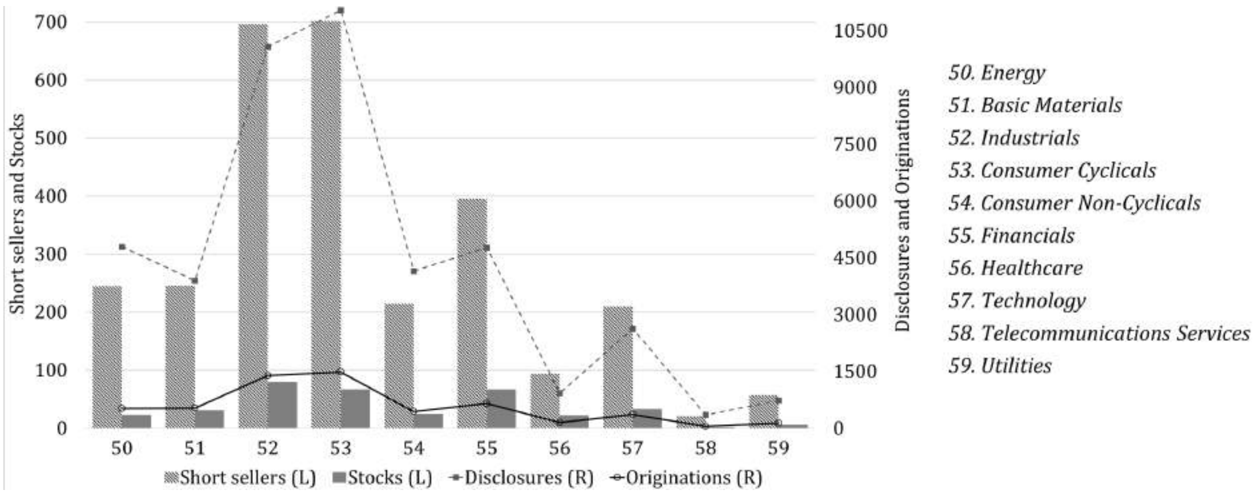
We present additional descriptive statistics of the stock sample in Table A.2 of the Internet Appendix.[7] The average (median) market capitalisation of companies in our sample is about \$5.5 billion (\$1.2 billion). For the purpose of comparison, the average market capitalisation of companies listed on the LSE, at the end of 2018, was of \$3.2 billion. We also

[5]The LSE operates with a price monitoring rule, which is very similarly to a circuit-breaker. Specifically, the rule works with a static and a dynamic threshold. Both thresholds are set at the stock-specific level, depending on liquidity of shares. When the price threshold is passed, automatic execution of trades is suspended for five minutes and a call auction takes place, to determine the best price.

[6]As explained in Section 4.2, results are robust to alternative constructions of the stock sample.

[7]Available at: https://www.dropbox.com/s/m08yplsk7wmaxnz/appendix_anon2.pdf?dl=0

Figure 2: Sample information by stock NACE Rev. 2 classification.



The chart shows the matched sample of 356 stocks by company classifier, following the Economic sector TRBC codes. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. To be read against the left axis, the bars depict, for any given sector, the number of stocks in the sample (full) and the number of short sellers taking position against those stocks (hatched). To be read against the right axis, the lines depict the total number of disclosures (dashed) and the number of disclosures that were originations (straight).

note that the stocks in our sample have low analyst coverage—on average, every year, only 0.89 analysts issue a 1-year ahead earnings forecast.[8] For a sample of stocks from the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ stock Market (NASDAQ), Hameed et al. (2015) report 4.60 analysts on average. Lastly, according to Refinitiv EIKON, 21.1% of the final sample of stocks have traded options.

3 Methodology

3.1 The Model

We follow the approach proposed by Antón and Polk (2014), who studied the impact of mutual fund holdings on the correlation of abnormal returns. Here, we are interested in the effect of common short selling.

We define $SSCAP_{ij,t}$ as the net value of stocks i and j, above the disclosure threshold of 0.5% shares outstanding, shorted by S common short sellers at quarter-end t, scaled by the

[8]Prior to July 2015, no stock in our dataset is covered by any analysts. If we consider only data after June 2015, the average number of analysts covering the stocks in our sample is 3.05 per year.

stock pair's market capitalisation. Specifically,

$$SSCAP_{ij,t} = \frac{\sum_{s=1}^{S}(W_{i,t}^{s} + W_{j,t}^{s})}{MV_{i,t} + MV_{j,t}}, \tag{1}$$

where $W_{i,t}^{s}$ is the value of the net short position held by common short seller s against stock i at quarter-end t and $MV_{i,t}$ is market value of stock i at quarter-end t. The value of the short position, $W_{i,t}^{s}$, is computed using the publicly disclosed short position weight, multiplied by the value of market capital of firm i on the reported day of the position. If the short seller reported more than one disclosure during the quarter, then we used the most recent disclosure.

Table 3 summarises the distribution of $SSCAP$. The average common (net) short sold capital of stock pairs is 0.03%, but can reach up to 9% of common capital. $SSCAP$ is sparse because, over any given quarter, short sellers tend to take few common positions across several stocks. However, as we will show in the next section, it contains explanatory power for future excess correlation.

Table 3: The Cross-Sectional Distribution of SSCAP, Jan. 2013—Dec. 2019

			Percentiles						
Year	Mean	Std	0	25	50	75	95	99	100
All	0.0003	0.002	0	0	0	0	0	0.0095	0.0901
2013	0.0001	0.001	0	0	0	0	0	0.0049	0.0475
2014	0.0001	0.001	0	0	0	0	0	0.0044	0.0455
2015	0.0002	0.0015	0	0	0	0	0	0.0072	0.0666
2016	0.0003	0.002	0	0	0	0	0	0.0095	0.0774
2017	0.0004	0.0025	0	0	0	0	0	0.0123	0.0812
2018	0.0005	0.0027	0	0	0	0	0.0049	0.0132	0.0901
2019	0.0005	0.0025	0	0	0	0	0.0045	0.0122	0.0679

The table reports the cross-sectional distribution of (scaled) common short sold capital. $SSCAP_{ij,t}$ is the net capital of stocks i and j short sold by common short sellers at quarter-end t, scaled by market capitalisation of the stock pair. The distributions of $SSCAP$ is shown for the whole sample and for individual years of sample coverage.

As can be noticed from Table 3, $SSCAP$ increases over time. To make the cross-sections comparable and ease interpretability of the regression coefficients, at each quarter, we normalise $SSCAP$ to have zero mean and unit standard deviation. We denote the normalised variable $SSCAP^*$.

We use $SSCAP^*_{ij,t}$ to forecast the future within-month realised correlation of each stock pair's daily four-factor residual returns, $\rho^{FFC}_{ij,t}$. Specifically, we estimate the following regression model,

$$\rho^{FFC}_{ij,t+1} = a + b_s \times SSCAP^*_{ij,t} + \sum_{k=1}^{n} b_k \times CONTROL_{ij,k} + \epsilon_{ij,t+1}. \tag{2}$$

To compute monthly realised correlation of residual returns for a given stock pair-month, we require daily residual return observations for at least 50% of the trading days of that month. The four factors used to compute residual returns are the local market excess return, the local size and value factors (Fama and French, 1993), and the local momentum factor (Carhart, 1997).[9] In Section 4.2, we show that results are robust to alternative factor models used to compute residual returns.

Our dependent variable of Equation 2, ρ^{FFC}, captures the part of correlation that isn't explained by the four well-known factors. Table 4 shows that, for our sample of stocks, the average four-factor residual return correlation is 5.35% . This represents over one quarter of the average correlation of raw returns. Also, together with Figure 1, Table 4 shows that both raw return cross-sectional correlations and residual return cross-sectional correlations are extremely variable across stock pairs. In some periods, stock pairs have an excess correlation up to 98.6%.

Given that the unexplained part of correlation remains substantial, we include in Equation 2 a large set of controls, which we present in detail in the next section. All variables on the right-hand side of Equation 2 are updated quarterly.

If common short sold capital is associated with higher future excess correlation, then b_s will be positive and significant. To limit the effect of serial correlation, we estimate b_s using the Fama and MacBeth (1973) regressions i.e., we run Equation 2 cross-sectionally for every t and compute the temporal average of b_s. Generally, we find that autocorrelation in our estimates is low and limited to the first lag. We account for autocorrelation up to three lags

[9]To compute residual returns, we use local factors i.e., factors associated to the country of a given stock. For six stocks, local factors were not available. For these exceptions, we use regional factors. Excess returns are computed with respect to the daily U.S. T-bill rate. Daily factor data are from AQR's website. Further details on the factor models are provided in Section A.1 of the Internet Appendix.

Table 4: **Summary statistics of stock price return correlation, Jan. 2013—Dec. 2019**

Variable	$\bar{N}$	Mean	St. Dev.	P10	P50	P90
ρ^{RET}	62,502	0.200	0.262	0.019	0.198	0.380
ρ^{CAPM}	62,502	0.066	0.243	-0.099	0.066	0.232
ρ^{FFC}	62,502	0.054	0.238	-0.109	0.053	0.217
ρ^{AQR}	62,502	0.054	0.237	-0.108	0.054	0.217
ρ^{ICAPM}	62,502	0.068	0.242	-0.097	0.068	0.234
ρ^{IFFC}	62,502	0.054	0.238	-0.109	0.054	0.217
ρ^{IAQR}	62,502	0.054	0.236	-0.108	0.054	0.216

The table presents summary statistics for the pairwise monthly realised correlation of residual returns and raw returns for an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. To compute monthly realised correlation of returns and residual returns for a given stock pair-month, we require daily observations for at least 50% of the trading days of that month. ρ^{RET} is the monthly pairwise realised correlation of daily raw returns. ρ^{CAPM}, ρ^{FFC}, and ρ^{AQR} are the monthly pairwise correlation of the daily residuals from, respectively, the local CAPM model, the local Fama and French (1993) and Carhart (1997) four-factor model, and the local six-factor AQR model, which is the FFC model augmented with local betting-against-beta (Frazzini and Pedersen, 2014) and quality-minus-junk (Asness et al., 2018) factors. Local factors are associated to the country (or region) of stocks. ρ^{ICAPM}, ρ^{IFFC}, and ρ^{IAQR} are the monthly pairwise correlation of the residuals from, respectively, the international CAPM model, the international Fama-French-Carhart model, and the international AQR model. International factor models are the local CAPM, FFC, and AQR models augmented with the corresponding global factors. Stock price data are from EIKON Refinitiv, whereas daily factors and the Treasury bill rate are from AQR's website. Note that the column headed $\bar{N}$ relates to the average number of observations across cross-sections. All other columns relate to pooled sample statistics.

(one quarter) with Newey and West (1987) robust standard errors.

3.2 Controls

In Equation 2, we include a large set of controls that explain stock return correlations beyond the four factors used to compute excess returns.

First, we control for common ownership of stock pairs. Let $HCAP_{ij,t}$ be the value of i and j held by common owners, scaled by the market capitalization of the two stocks. $HCAP$ controls for excess correlation created by common owners purchasing and selling stocks. By including $HCAP$ in our specification, we aim to separate the excess correlation due to short selling activity from long strategies of investors. Because our measure of common short selling is constructed using net short positions, $HCAP$ alleviates concerns that movements in $SSCAP$ might be due to changes in long positions.

Next, we control for industry effects using the Thompson Reuters Business Classification

(TRBC). TRBC offers the widest coverage for the stocks in our sample. It consists of four levels of classification (Economic Sector, Business Sector, Industry Group, and Industry). We created the variable $NUMTRBC_{ij,t}$, which captures the number of consecutive equal level codes, starting from the most generic, in the TRBC of stocks i and j. Alternative definitions of the industry control, based on different classification codes (e.g., SIC), yield similar results.

We also compute a series of additional size, style, and pair characteristic controls.

In terms of size, we control for the size of the two companies i and j using their market capitalisation. Chen et al. (2017) show that stocks of similar size tend to be more highly correlated. Hence, we captured similarity in size using $SAMESIZE_{ij,t}$, which we define as the negative absolute difference in the cross-sectional percentile ranking of the market capitalization of i and j at quarter-end t. As size is a proxy for the number of shares available to short sell (Dechow et al., 2001), it can also control for short selling costs. Thus, we included $GAVSIZE_{ij,t}$, which is the geometric average of the cross-sectional percentile ranking of the market capitalization of i and j at quarter-end t.

In terms of style, we control for similarities in the book-to-market ratio and the momentum of the two stocks. We define $SAMEBM_{ij,t}$ and $SAMEMOM_{ij,t}$ as the negative absolute difference in the cross-sectional percentile ranking of, respectively, the book-to-market ratio, and the momentum of the two stocks.[10]

Book-to-market ratios are positively associated with future returns (Rosenberg et al., 1985, Fama and French, 1992). Moreover, Curtis and Fargher (2014) show that short sellers tend to concentrate on stocks with high book-to-market. Hence, we include the geometric average of the cross-sectional percentile rank of the book-to-market of the two stocks, $GAVBM_{ij,t}$. Furthermore, because short sellers might ride on declining prices, which are, by definition, correlated, we include the geometric average of the percentile rank of the momentum of the two stocks, $GAVMOM_{ij,t}$.

[10]We define momentum as the cumulative stock return over the last year, excluding the most recent month.

Table 5: Summary statistics of stock pair variables, Jan. 2013—Dec. 2019

Variable	N	Mean	St. Dev.	P10	P50	P90
$SSCAP$	63,190	0.000	0.002	0.000	0.000	0.000
$SSVOL$	63,190	0.159	2.047	0.000	0.000	0.000
$SSFLOAT$	63,190	0.001	0.004	0.000	0.000	0.000
NSS	63,190	0.040	0.234	0.000	0.000	0.000
$HCAP$	63,190	0.102	0.107	0.005	0.074	0.167
A	63,190	0.023	0.248	0.000	0.000	0.000
$SAMESIZE$	60,997	-0.335	0.236	-0.503	-0.295	-0.136
$SAMEBM$	60,236	-0.335	0.236	-0.503	-0.295	-0.135
$SAMEMOM$	63,190	-0.335	0.236	-0.501	-0.296	-0.135
$GAVSIZE$	60,997	0.444	0.230	0.259	0.432	0.618
$GAVBM$	60,236	0.444	0.230	0.259	0.432	0.618
$GAVMOM$	63,190	0.444	0.230	0.259	0.432	0.618
$DIFFLEV$	52,891	3.102	8.717	0.320	0.852	2.244
$DIFFPRICE$	62,789	1.579	1.332	0.593	1.261	2.197
$NUMTRBC$	63,190	0.256	0.736	0.000	0.000	0.000
$RETCORR$	62,712	0.260	0.187	0.143	0.263	0.383
$ROECORR$	60,307	0.057	0.569	-0.445	0.085	0.567
$VOLCORR$	62,789	0.088	0.297	-0.121	0.080	0.299
$GEODIST$	62,835	677.3	677.3	53.48	208.4	398.5
$DCOUNTRY$	63,190	0.789	0.408	1.000	1.000	1.000
$DCITY$	63,190	0.129	0.336	0.000	0.000	0.000
$DINDEX$	63,190	0.569	0.495	0.000	1.000	1.000

The table presents summary statistics for stock pair variables of an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. For a given stock pair, $SSCAP$, $SSVOL$, and $SSFLOAT$ are the net capital of the stock pair shorted by common short sellers and scaled, respectively, by the stock pair's market capital, trading volume, and equity float. NSS is the number of common short sellers. These measures are constructed using short selling disclosure data from the UK's FCA, so capture net short positions larger than the regulatory threshold of 0.5% shares outstanding. $HCAP$ is the (scaled) capital held by common owners. A is the number of common analysts issuing an earnings forecast of a stock pair over the past year. $SAMESIZE$, $SAMEBM$, and $SAMEMOM$ are the negative of the absolute difference in the cross-sectional percentile ranking of, respectively, size, book-to-market, and momentum, for the stock pair. $NUMTRBC$ is the number of consecutively equal digits in the TRBC code for the stock pair. $GAVSIZE$, $GAVBM$, and $GAVMOM$ are the the geometric average of the cross-sectional percentile ranking of, respectively, size, book-to-market, and momentum, for the stock pair. For a stock pair, $RETCORR$, $ROECORR$, and $VOLCORR$ measure the correlation of, respectively, the past 2-year monthly return, the past 5-year return on equity, and the past 2-year monthly abnormal trading volume. $DIFFLEV$ and $DIFFPRICE$ are, respectively, the absolute difference in leverage and price of a stock pair. $DCOUNTRY$ and $DCITY$ are dummy variables capturing whether both stocks in a stock pair have their headquarters in the same country and city. $DINDEX$ is a dummy variable capturing whether both stocks of a stock pair are members of the same index. All variables are updated quarterly. Ownership, company, and market data are from Refinitiv EIKON. Earning's forecast data are from IBES.

We control for a series of stock pair characteristics. To address concerns for potential reverse causality in our regression model, we control for the past 2-year monthly price return correlation of stock pairs, which we denote $RETCORR_{ij,t}$. As companies with similar profits are expected to have correlated stock returns (Chen et al., 2017), we also control for the past 5-year correlation of the return on equity for every pair, $ROECORR_{ij,t}$. We also included a control variable capturing similarity in abnormal trading volumes of stock pairs, $VOLCORR_{ij,t}$, which measures the monthly correlation in abnormal trading volumes over the past two years.[11] We control for the absolute difference in the price level of the two stocks, which we denote $DIFFPRICE_{ij,t}$, as well as the absolute difference in their leverage, $DIFFLEV_{ij,t}$.

Further, we create variables to control for geographical location (Pirinsky and Wang, 2006). First, $GEODIST$ measures the geographical distance (in Kilometers) between the headquarters of two companies. Second, dummy variables $DCOUNTRY$ and $DCITY$ measure, respectively, whether two companies have headquarters in the same country and city.

Past studies have reported that index membership affects correlation (see, Barberis et al. (2005) and Greenwood (2007)). We construct a dummy variable $DINDEX_{ij,t}$, which is equal to one if, at quarter-end t, both stocks i and j are members of the same index. We construct the variable by checking the constituents (over the period of study) of a list of 840 indices.[12] We report summary statistics of the stock pair variables in Table 5.

We update all controls quarterly. For our regression analysis of Equation 2, we standardise all independent variables, with the exception of the dummy variables, so that they have zero mean and unit standard deviation. This is to ease interpretation of regression coefficients.

4 Results

4.1 Main Regression Estimates

Table 6 presents the results of the Fama and MacBeth regression specified in Equation 2, using $SSCAP^*$ to predict the realised correlation of four-factor residual returns.

[11] We define abnormal trading volume as the residual of a regression of volume on an annual trend and monthly dummies.

[12] The list is determined by checking index membership, during the period of study, of the 356 stocks in our sample. Index constituents data are from Refinitiv EIKON.

The first column of Table 6 reports the baseline specification with just $SSCAP^*$ and a constant. The coefficient on $SSCAP^*$ is positive and significant, with a coefficient equal to 0.00343. Given that $SSCAP^*$ is standardized to have zero mean and unit standard deviation, the regression constant reflects the average residual correlation when $SSCAP$ is at its mean. Thus, the coefficient can be interpreted with respect to the average residual correlation. A standard deviation increase in common short sold capital is associated with an increase of the predicted residual return correlation by about 6.4% of its average.

The second column of Table 6 shows results controlling for common ownership, similarity in sector, size, book-to-market, and momentum. The coefficient on $HCAP^*$ is positive and highly significant. This result is consistent with common owners inducing higher correlation by trading of stocks held in common (Bartram et al., 2015).

Recall that the dependent variable is the correlation of the residuals of a four-factor asset pricing model, which includes the size, book-to-market, and the momentum factor. Despite this, similarity in size book-to-market, and momentum still have a strong positive and significant association with future excess correlation.

Consistent with early studies of Pindyck and Rotemberg (1993), the similarity in sector of the two companies is a key determinant of correlation. The coefficient on $NUMTRBC^*$ is statistically significant with a coefficient of 0.01383 and a t-statistic of 37.69.

In the second specification, we also include the size control $GAVSIZE^*$, the geometric average of the percentile ranking of the stock pair size. After adding these controls, the size of the coefficient on $SSCAP^*$ decreases, but remains significant at 5% confidence level with a p-value of 2.3%.

In the third specification of Table 6, we add additional controls for pair characteristics. The terms capturing similarities in past correlation, past profits, and past abnormal trading volume are all positive and significant. The coefficient on $DIFFLEV^*$ is positive and significant at 1%, meaning that stocks that have similar leverage have lower correlation of excess returns. The coefficients on $DIFFPRICE^*$ is insignificant. The coefficients on $GEODIST$ is negative and significant at the 5% confidence level, indicating that stocks of companies that are geographically closer are more strongly correlated. With these additional controls the coefficient of $SSCAP^*$ is significant at the 5% level (p-value of 1.1%).

Table 6: Fama-MacBeth regressions of monthly realised pairwise correlation of daily four-factor residual returns on *SSCAP* and stock pair controls, Jan. 2013—Dec. 2019

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.05344	0.05355	0.05420	0.05085
	(13.80)	(13.90)	(13.83)	(12.87)
$SSCAP^{*}$	0.00343	0.00099	0.00110	0.00145
	(7.04)	(2.33)	(2.61)	(3.57)
$HCAP^{*}$		0.00311	0.00242	0.00261
		(5.72)	(4.20)	(4.62)
$SAMESIZE^{*}$		0.00899	0.00739	0.00728
		(10.57)	(7.82)	(8.01)
$SAMEBM^{*}$		0.00219	0.00154	0.00142
		(7.53)	(5.23)	(4.14)
$SAMEMOM^{*}$		0.00995	0.00766	0.00620
		(12.80)	(10.44)	(9.11)
$NUMTRBC^{*}$		0.01383	0.01157	0.01152
		(37.69)	(34.02)	(34.20)
$GAVSIZE^{*}$		0.01482	0.01193	0.01122
		(12.84)	(10.25)	(9.46)
$GAVBM^{*}$				0.00003
				(0.04)
$GAVMOM^{*}$				0.00490
				(4.21)
$RETCORR^{*}$			0.01243	0.01279
			(16.65)	(19.11)
$ROECORR^{*}$			0.00210	0.00193
			(5.94)	(5.75)
$VOLCORR^{*}$			0.00610	0.00571
			(6.60)	(7.15)
$DIFFLEV^{*}$			0.00200	0.00200
			(3.45)	(3.51)
$DIFFPRICE^{*}$			0.00083	0.00125
			(1.65)	(2.32)
$GEODIST^{*}$			-0.00161	-0.00096
			(-2.55)	(-1.65)
$DCOUNTRY$				0.00319
				(1.74)
$DCITY$				-0.00297
				(-2.66)
$DINDEX$				0.00185
				(1.32)

Continued on next page

Table 6—*Continued*

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
R^2	0.05728 (7.83)	0.07510 (9.00)	0.07931 (9.23)	0.08212 (9.31)
N. Obs.	62,502 (220.48)	58,407 (132.74)	49,871 (106.78)	49,871 (106.78)
Size controls	No	Yes	Yes	Yes
Firm attributes	No	No	Yes	Yes
Style controls	No	No	No	Yes
Dummy variables	No	No	No	Yes

The table reports Fama and MacBeth (1973) regression coefficients, computed by running cross-sectional regressions each month between Jan. 2013 and Dec. 2019 (84 months) and then averaging regression coefficients over the sample period. Observations relate to stock pairs from an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. The dependent variable is the realised pairwise correlation in month $t+1$ of the daily residual returns from a four-factor model. The four factors are: the market excess return, the size and value factors (Fama and French, 1993), and the momentum factor (Carhart, 1997). Factors are local i.e., they are associated to the country (or region) of stocks. Stock price data are from Refinitiv EIKON, whereas daily factors are from AQR's website. The independent variables include $SSCAP$, which is the net (scaled) capital of the stock pair shorted by common short sellers, and controls at time t. $SSCAP$ is constructed using public disclosure data from the UK's FCA, so it captures common net short positions larger than the regulatory threshold of 0.5% share capital. $HCAP$ is the (scaled) capital held by common owners. $SAMESIZE$, $SAMEBM$, and $SAMEMOM$ are the negative of the absolute difference in the cross-sectional percentile ranking of, respectively, size, book-to-market, and momentum, for the stock pair. $NUMTRBC$ is the number of consecutively equal digits in the TRBC code for the stock pair. $GAVSIZE$, $GAVBM$, and $GAVMOM$ are the the geometric average of the cross-sectional percentile ranking of, respectively, size, book-to-market, and momentum, for the stock pair. For a stock pair, $RETCORR$, $ROECORR$, and $VOLCORR$ measure the correlation of, respectively, the past 2-year monthly return, the past 5-year return on equity, and the past 2-year monthly abnormal trading volume. $DIFFLEV$ and $DIFFPRICE$ are, respectively, the absolute difference in leverage and price of a stock pair. $DCOUNTRY$ and $DCITY$ are dummy variables capturing whether both stocks in a stock pair have their headquarters in, respectively, the same country and city. $DINDEX$ is a dummy variable capturing whether both stocks of a stock pair are members of the same index. Stock-pair controls are constructed using ownership and company data from Refinitiv EIKON. All independent variables are updated quarterly and, with the exception of dummy variables, are cross-sectionally normalised (to have zero mean and unit standard deviation), which we denote by $*$. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 3 lags (one quarter).

In the fourth column of Table 6, we add book-to-market and momentum measures for the stock pairs, $GAVEBM^*$ and $GAVMOM^*$. The coefficient estimate on the former control variable is insignificant, whereas the coefficient on the latter is highly significant. We also add three dummy variables. The coefficient on the dummy variables $DCOUNTRY$ is positive and significant at the 10% confidence level, indicating that stocks of companies based in the same country are more strongly correlated. Quite to the contrary, the coefficient on $DCITY$ is negative and significant at the 1% confidence level. The effect on $DINDEX$ is positive, as expected, but insignificant. In this specification, which is the most complete and flexible, the coefficient on $SSCAP^*$ is 0.00145, which underlines that an increase in one standard deviation in common short sold capital is associated with an increase of the predicted correlation of excess returns of about 2.9% of the average abnormal correlation. The effect is strongly significant, with an associated t-statistic of 3.57 and p-value of 0.1%.

The coefficients reported in Table 6 show that the effect of a standard deviation increase in common short selling is smaller than the effect of a standard deviation increase in common ownership. It's likely that, to some extent, this difference reflects the fact that it is much easier to "go long" than to "go short" and that long positions are generally much larger than short positions.

Consider interpreting results in terms of stock pair capital, using the average cross-sectional standard deviation of $SSCAP$, given in Table 3. In the most complete specification, a 1% increase in common short sold capital is associated with an increase of excess correlation of 0.00728, equivalent to 14.3% of the average excess correlation. Taking into account the standard deviation of $HCAP$, given in Table 5, a 1% increase in common ownership is associated with an increase of excess correlation of 0.00024, that is 0.5% of its average. Hence, our results show that a common short position of, for example, just 0.5% of stock pair capital predicts a larger increase in excess correlation than a long position of 10% of common capital.

In untabulated results, we find that fitted values that are due to $SSCAP^*$ range from an average minimum of 0.0379 to an average maximum of 0.1048, around an average excess correlation of 0.0508.[13] As a mean for comparison, the fitted values due to $HCAP^*$ range from 0.0367 to 0.0709, showing that common (net) short positions have similar explanatory power to common long positions.

[13] To calculate the range of these fitted values, we first orthogonalise $SSCAP^*$ with respect to all the controls used in the fourth specification. We then forecast the realised correlation of four-factor residual returns using the orthogonalised $SSCAP^*$ and save the minimum and maximum forecasted value for each cross-section. Finally, we average these values across time.

Overall, despite the high variability of the correlation of excess returns, which have an average standard deviation of 0.24, in the fourth specification, the regression has an an average R^2 of 8.21% and the association between $SSCAP^*$ and the future correlation of residual returns is significant at 1% confidence level.

Notice that $SSCAP^*$ explains excess returns of a four factor model, even after accounting for many controls and characteristics. As explained in more detail in the next subsection, the results of Table 6 are robust to using raw returns or alternative factor models.

4.2 Robustness

To check the robustness of our results, we carry out a series of alternative specifications of the regressions given in Table 6.

First, we run the regressions in Table 6 using robust regressors, by rank-transforming (and standardising) all right-hand variables, except for the dummy controls, of Equation 2. We denote rank-transformed (and normalised) common short selling as $SSCAP^\dagger$. Panel A of Table 7 shows that, although the significance of the estimate on $SSCAP$ decreases, it remain significant across all specifications, with with 1% confidence level in the most complete regression.

Second, we run the regressions in Table 6 using alternative definitions of the main covariate, common short selling. For these regressions we use $SSVOL_{ij,t}$ and $SSFLOAT_{ij,t}$, which are the (net) common short sold capital of stock pairs i and j at quarter-end t scaled by, respectively, the dollar trading volume and the free float of the stock pair. This is to take into account the liquidity of stock pairs, which might limit the capital exposure of short sellers.[14] Additionally, we look at the effect of the number of common short sellers, NSS. Again, as for $SSCAP$, these common short selling variables are constructed using the disclosure data, so account for large (net) short positions, above the disclosure threshold.

Panel B of Table 7 presents the results for the third and fourth specifications of Table 6, using the alternative measures of common short selling. As with the regressions of Section 4.1, all variables are standardised to have zero mean and unit standard deviation. Panel B of Table 7 shows that changing the scale of $SSCAP$ does not alter our main results.

[14] We thank an anonymous referee for suggesting these measures.

Moreover, the last two column of Panel B show that the number of common short sellers is also predictive of future comovement.

Third, we verify robustness of results with different specifications of the dependent variable i.e., with different measures of the comovement of residual returns. Panel C of Table 7 presents our two most complete regression specifications using Kendall's and Spearman's measures of rank-correlation as the left-hand side variables. The coefficient on $SSCAP$ remains stable and significant.

We also re-run regressions using a continuous transformation of Pearson's correlation coefficient as dependent variable. Here, we follow Pindyck and Rotemberg (1993) and transform the realised monthly correlation to $c^{FFC} = \tan(\pi\rho^{FFC}/2)$. This transformation pushes high (absolute) correlation values towards minus/plus infinity. This continuous measure of excess correlation leaves results unchanged. The first two columns of Panel C of Table 7 report the two most complete regression specifications using c^{FFC} as regressand.

Fourth, we verify whether the result holds for the correlation of raw returns (ρ^{RET}) and the correlation of alternative residual returns. Specifically, we use the residual returns from the local classic Capital Asset Pricing Model (CAPM) and the correlation of residual returns from a six-factor asset pricing model (referred to as AQR), which is the local FFC model augmented with the local betting-against-beta (BAB) factor of Frazzini and Pedersen (2014) and the local quality-minus-junk (QMJ) factor of Asness et al. (2018).[15] We denote the correlation of these residual returns as, respectively, ρ^{CAPM} and ρ^{FFC}.

In addition to the local CAPM, FFC, and AQR models, we also test the international versions of these models, which include both local factors (specific to the country or region of the given stock) and global factors. For example, the international FFC includes eight factors: the local and global market, size, value, and momentum factors. We denote the correlation of the residuals from the international CAPM, the international FFC, and international AQR factor models as, respectively, ρ^{ICAPM}, ρ^{IFFC}, and ρ^{IAQR}. Descriptive statistics on (residual) correlations are provided in Table 5, whereas further details on the factor models are outlined in the Internet Appendix.

[15] Jank and Smajlbegovic (2017) find evidence that short sellers trade on the BAB and QMJ factors. By including these factors in our robustness tests, we control for the possibility that part of the relationship between common short selling and excess comovement is due to BAB and QMJ.

Table 7: Fama-MacBeth regressions of monthly realised correlation of daily four-factor residual returns on $SSCAP$ and controls, with alternative measures of correlation and common short selling, Jan. 2013—Dec. 2019

Panel A: Rank Transformed Regressors

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.05344	0.05357	0.05436	0.05128
	(13.8)	(13.9)	(13.91)	(13.68)
$SSCAP^{\dagger}$	0.0036	0.00079	0.00084	0.00126
	(6.93)	(1.8)	(2.05)	(3.22)
Other control reported in the Internet Appendix				
Size controls	No	Yes	Yes	Yes
Firm attributes	No	No	Yes	Yes
Style controls	No	No	No	Yes
Dummy variables	No	No	No	Yes

Panel B: Alternative measures of common short selling

	Dependent Variable: Correlation of 4F Residuals					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.05421	0.05084	0.05421	0.05084	0.0542	0.05089
	(13.84)	(12.86)	(13.83)	(12.88)	(13.83)	(12.88)
$SSVOL^{*}$	0.00052	0.00071				
	(1.83)	(2.59)				
$SSFLOAT^{*}$			0.00083	0.00118		
			(2.06)	(3.05)		
NSS^{*}					0.00122	0.00162
					(2.91)	(4.02)
Other control reported in the Internet Appendix						
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm attributes	Yes	Yes	Yes	Yes	Yes	Yes
Style controls	No	Yes	No	Yes	No	Yes
Dummy variables	No	Yes	No	Yes	No	Yes

Continued on next page

Table 7—*Continued.*

Panel C: Alternative correlation measures						
	Dependent Variable: Correlation of 4F Residuals					
	Continuous		Kendall		Spearman	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.05803	0.05446	0.04255	0.03981	0.06435	0.06031
	(13.71)	(12.69)	(14.34)	(13.81)	(14.3)	(13.91)
$SSCAP^{*}$	0.00117	0.00154	0.00086	0.00109	0.00126	0.00163
	(2.56)	(3.53)	(3.02)	(3.92)	(2.96)	(3.88)
Other control reported in the Internet Appendix						
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm attributes	Yes	Yes	Yes	Yes	Yes	Yes
Style controls	No	Yes	No	Yes	No	Yes
Dummy variables	No	Yes	No	Yes	No	Yes

The table reports Fama and MacBeth (1973) regression coefficients, computed by running cross-sectional regressions each month between Jan. 2013 and Dec. 2019 (84 months) and then averaging regression coefficients over the sample period. Observations relate to stock pairs from an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. The dependent variable is the realised pairwise correlation in month $t+1$ of the daily residual returns from a four-factor model. The four factors are: the market excess return, the size and value factors (Fama and French, 1993), and the momentum factor (Carhart, 1997). Factors are local i.e., they are associated to the country (or region) of stocks. Stock price data are from Refinitiv EIKON, whereas daily factors are from AQR's website. The independent variables include $SSCAP$, which is the net (scaled) capital of the stock pair shorted by common short sellers, and controls at time t. In Panel A, all independent variables are ranked-transformed and normalised, which we denote by †. Panel B reports results using alternative measures of common short selling. $SSVOL$ is the net capital of the stock pair shorted by common short sellers, scaled by the stock pair's trading volume. $SSFLOAT$ is the net capital of a stock pair shorted by common short sellers, scaled by the stock pair's equity float. NSS is the number of common short sellers. Short selling measures are constructed using public disclosure data from the UK's FCA, so measures capture common net short positions larger than the regulatory threshold of 0.5% share capital. Panel C reports results using alternative measures of the correlation of four-factor residual returns as dependent variables. In these specifications we use three measures: a) the continuous transformation of Pearson's pairwise correlation (ρ) proposed by Pindyck and Rotemberg (1993), $c = \tan(\pi\rho/2)$; b) Kendall's rank correlation; and c) Spearman's rank correlation. Estimates for the remaining controls may be found in the Internet Appendix. In Panel B and C, all independent variables (except the dummy variables) are normalised to have zero mean and unit standard deviation, which we denote by $*$. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 3 lags (one quarter).

Table A.7 in the Internet Appendix reports results from the most complete regression specification, using residual correlation from these alternative factor models. Overall, results are robust to the factor model choice. When we use the AQR models to compute residual return correlation, the coefficient estimate on $SSCAP^{*}$ decreases in size, but remains significant at 1% level of confidence. As can be noted from the first column of Table A.7, the effect of $SSCAP^*$ remains significant at 5% when using raw returns (RET) to compute excess correlation.

Fifth, for the period Jul. 2015 to Dec. 2019, we test the robustness of our results by including the (normalised) number of common analysts, A^*, as an additional control variable in our models. Analysts tend to specialise in stocks that are similar across many different dimensions, some of which might not be easily observable. In this sense, A^* helps us proxy for those unobservable factors that might be driving correlation and that are not captured in our regression specifications.

We define $A_{ij,t}$ equal to the number of analysts issuing an earnings forecast of both stocks i and j over the 12-months prior to t. Because analyst coverage for our stock sample is either missing or zero prior to mid-2015, our variable for common analyst coverage is also zero for that period. Therefore, for our robustness checks with A^*, we run the Fama-MacBeth regressions only for the period Jul. 2015—Dec. 2019.

We present results in Table A.8 in the Internet Appendix. Although, across all specifications, the coefficients on $SSCAP^*$ are larger than for the full sample shown in Table 6, significance decreases for the specifications that include A^*. Nonetheless, the estimates on $SSCAP^*$ remains significant at the 5% level of confidence (p-value of 4.1%) in the third specification, and at 1% in the fourth and most complete specification (p-value of 0.6%).

Lastly, to address the variability of the sample across the regression specifications, we conducted two additional robustness checks, which we report in the Internet Appendix.

First, we restrict the sample of stock-pairs to that of specifications 3 and 4 of Table 6. Regression results, reported in Table A.9 of the Internet Appendix, show that, for this subsample of stock-pairs, the coefficient estimates for specifications 1 and 2 remain basically unchanged.

Second, we re-run regressions for a subsample of 195 stocks, for which we have complete data for all the controls. This creates an almost perfectly balanced sample of $n \times (n-1)/2 = 18,915$ stock-pairs observations.[16] For this smaller, more balanced, subsample, there is

[16]The sample is not perfectly balanced because, for some stock pair-months, we might be missing sufficient daily residual return to compute realised correlation.

substantially less variability in the number of observations across regression specifications. Results are reported in Table A.10 of the Internet Appendix and show that the coefficient on $SSCAP^*$ remains strongly statistically significant.

5 Two Explanations

5.1 Hypotheses Development

We put forward two competing explanations for the uncovered relationship between common short positions and excess comovement.

The first one is that, by taking large short positions on two or more stocks, short sellers exert price pressure, which materialises as comovement.

This price pressure mechanism has been used to explain the relationship between common ownership and excess comovement (Bartram et al., 2015, Antón and Polk, 2014). It has also been illustrated in theoretical studies on short selling. Brunnermeier and Pedersen (2005) and Cont and Wagalath (2013) suggest that short sellers can drive down prices of several stocks inducing shifts in correlation and contagion. Both models predict that the effect of short selling on correlation inversely depends on market depth of the stocks short sold—the more illiquid the stocks, the greater should be the impact of common short sellers on correlation.

We draw on these studies to develop our first hypothesis, which we test in Section 5.2.

Hypothesis 1: According to the price pressure mechanism, the relationship between $SSCAP$ and excess comovement should be stronger for more illiquid stock pairs.

Note that the effect does not necessarily have to work in one direction. Positive price pressure might equally explain the relationship between common short selling and excess comovement. In the event of a short squeeze, for example, short sellers would have to buy stocks in order to cover their positions, driving up prices and correlation. To the extent that a higher level of $SSCAP$ is associated with a higher probability of future short squeeze, the relationship between $SSCAP$ and future excess correlation could also be due to positive price pressure.

Therefore, we do not restrict our analysis to either negative or positive price pressure. Rather, we test the prediction that the relation between $SSCAP$ and excess comovement should be stronger for more illiquid stock pairs. Our goal is not to prove a causal relation, which is outside the scope of this paper, but to verify whether the predictions of the price pressure mechanism are consistent with our results for $SSCAP$.

There are a wide range of alternatives to Hypothesis 1, which are discussed further in Section 5.2. These involve, among other, the possibility that short sellers avoid illiquid stocks or adopt stealth trading strategies.

Our second hypothesis is that the uncovered relationship is due to short sellers trading on superior information. There is considerable evidence in the literature that short sellers are informed traders. For example, Boehmer et al. (2008) and Diether et al. (2009b) show that short sellers can correctly predict future returns. Furthermore, studies have found that short sellers tend to focus on overpriced stocks, thus trading in a non-predatory way against market sentiment (Dechow et al., 2001, Curtis and Fargher, 2014).

If common short sellers are informed traders and expect stock prices to decline in the future, as declines materialise, their positions should be associated with higher future correlation. That is, $SSCAP$ predicts future price declines and, as these declines occur, we observe higher correlations. In this sense, the relationship between common short selling and excess comovement is non-causal.

From a temporal perspective, prior evidence of the long-horizon predictability of short selling supports this reasoning. For example, Boehmer et al. (2010) find that heavily shorted stocks display negative monthly abnormal returns for over six months. Jones et al. (2016), show that initial short disclosures are associated with cumulative abnormal returns of -5.23% after 90 days. If, as studies have shown, the predictability of informed short positions persists for several months, this effect should last for the time lag adopted in our regressions.

A counter argument is that informed short positions predict average future stock returns, whereas correlation measures deviations around the average. However, if the overall tendency is for prices of stock pairs to decline, this should result in higher correlation compared to stock pairs for which we do not expect this general tendency. A similar argument has led to the development of unbiased correlation estimators based on opening and closing prices (see, for example, Rogers and Zhou (2008)).

This leads to our second hypothesis explaining the relationship between common short selling and excess comovement.

Hypothesis 2: According to the informative trading mechanism, the relationship between $SSCAP$ and excess comovement should be stronger when common short positions are taken by informed traders.

We test Hypothesis 2 in Section 5.3 by exploiting two additional aspects of our framework: 1) short sellers' investor profiles and 2) their past short selling performance.

5.2 Illiquidity and Price Pressure

To verify Hypothesis 1, we construct a dummy variable that captures the illiquidity of a stock pair. Specifically, $DAMIHUD_{ij,t}$ is equal to one if, during quarter t, the Amihud measure of both stocks i and j is above the cross-sectional median.[17] We add the dummy variable and its interaction with $SSCAP^*$ to the two most flexible models of our main regressions. Results, reported in the first two columns of Table 8, show that illiquidity is associated with lower future excess comovement, when controlling for size, style, and common characteristics.

Contrary to what is predicted by the price pressure mechanism, the interaction term with $SSCAP^*$ is negative and significant at 10% in the first specification and insignificant in the second. The association between common short selling and excess comovement is weaker for more illiquid stocks. In fact, the total effect of common short selling for illiquid stock pairs, is statistically irrelevant.

This result is confirmed with alternative liquidity measures. We define the dummy $DTURN_{ij,t}$ as equal to one if, during quarter t, both stocks i and j have average daily turnover below the cross-sectional median.[18] The third and fourth columns of Table 8 show that the coefficient on the interaction between $SSCAP^*$ and $DTRUN$ is negative and significant at the 5% confidence level, indicating that the effect of common short selling is less

[17] The Amihud (2002) measure of stock i is defined as $(\sum_{d=1}^{D_t} |r_{i,t_d}|)/(D_t \sum_{d=1}^{D_t} V_{i,t_d})$, where D_t is the number of trading days in quarter t, $\mid r_{i,t_d} \mid$ is the absolute value of the daily return of stock i on day t_d, and V_{i,t_d} is the daily dollar volume of shares traded.

[18] Turnover is the volume of shares traded, as a percentage of shares outstanding.

strong for the most illiquid stocks. Again, the total effect of common short selling for illiquid stocks is statistically insignificant, meaning that $SSCAP^*$ has an effect only for liquid stock pairs.

We verify results with a third dummy variable of illiquidity, based on the free float of the stock pair. $DFLOAT_{ij,t}$ is equal to one if, during quarter t, both stocks i and j have their free float, as percentage of their shares outstanding, below the cross-sectional median.

The fifth and sixth columns of Table 8 present results using $DFLOAT$. Similar to the results for $DAMIHUD$, Column 5 and 6 indicate that higher illiquidity is associated with lower levels of excess comovement. Table 8 also shows that the coefficient on interaction term between $DFLOAT$ and $SSCAP^*$ is negative and highly signifcant in both specifications.

Generally, across all regressions, our results indicate that the association between $SSCAP$ and excess correlation holds when at least one of the two stocks is liquid. When the stock pair is illiquid i.e., when both stocks are below the cross-sectional median of the liquidity indicator, then the effect of $SSCAP$ weakens significantly or vanishes. This result holds for alternative thresholds used to construct the illiquidity dummies.

These results are confirmed with several additional robustness checks.

First, we use continuous liquidity measures of the Amihud indicator, turnover, and free float. Results are presented in Table A.12 in the Internet Appendix, and show that that the positive association between common short selling and excess return comovent continues to hold for liquid stocks, but that this effect weakens as stock pair illiquidity increases.

Second, we use alternative factor models to compute the correlation of residual returns. Table A.13 of the Internet Appendix shows that using residual returns from the AQR six-factor model leaves results unvaried. We obtain comparable outcomes when we use the correlation of raw returns and of residual returns from alternative factor models.

Finally, alternative measures of common short selling, such as those described in Section 4.2, yield similar results.

Because the effect of $SSCAP$ does not increase with the illiquidity of stocks, our results contradict the price pressure mechanism outlined in Hypothesis 1.

Table 8: Fama-MacBeth cross-sectional regressions of monthly realised pairwise correlation of daily four-factor residual returns on the interaction of $SSCAP$ with illiquidity dummies, and stock pair control variables, Jan. 2013—Dec. 2019

	Dependent Variable: Correlation of 6F Residuals					
	(1)	(2)	(3)	(4)	(5)	(6)
$Constant$	0.05637 (13.94)	0.05302 (13.13)	0.05265 (13.68)	0.04980 (12.67)	0.05518 (13.51)	0.05212 (12.45)
$SSCAP^{*}$	0.00111 (2.37)	0.00145 (3.20)	0.00133 (3.10)	0.00163 (3.94)	0.00149 (3.30)	0.00184 (4.28)
$DAMIHUD$	-0.00884 (-4.78)	-0.00920 (-4.71)				
$SSCAP^{*} \times DAMIHUD$	-0.00122 (-1.95)	-0.00102 (-1.61)				
$DTURN$			0.00620 (4.78)	0.00482 (3.71)		
$SSCAP^{*} \times DTURN$			-0.00173 (-2.37)	-0.00175 (-2.37)		
$DFLOAT$					-0.00390 (-3.85)	-0.00364 (-3.45)
$SSCAP^{*} \times DFLOAT$					-0.00156 (-3.53)	-0.00150 (-3.31)
Other control reported in the Internet Appendix						
Tot. $SSCAP^{*}$ effect for illiquid pairs	-0.00011 (-0.19)	0.00043 (0.76)	-0.00041 (-0.57)	-0.00011 (-0.15)	-0.00007 (-0.15)	0.00034 (0.73)

Continued on next page

Table 8—*Continued*

	Dependent Variable: Correlation of 6F Residuals					
	(1)	(2)	(3)	(4)	(5)	(6)
$R2$	0.07971	0.08253	0.07984	0.08259	0.07967	0.08246
	(9.89)	(9.98)	(9.90)	(9.99)	(9.89)	(9.98)
N. Obs.	49,871	49,871	49,871	49,871	49,860	49,860
	(106.78)	(106.78)	(106.78)	(106.78)	(105.83)	(105.83)
Size controls	Yes	Yes	Yes	Yes	Yes	Yes
Pair characteristic controls	Yes	Yes	Yes	Yes	Yes	Yes
Style controls	No	Yes	No	Yes	No	Yes
Dummy controls	No	Yes	No	Yes	No	Yes

The table reports Fama and MacBeth (1973) regression coefficients, computed by running cross-sectional regressions each month between Jan. 2013 and Dec. 2019 (84 months) and then averaging regression coefficients over the sample period. Observations relate to stock pairs from an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. The dependent variable is the realised pairwise correlation in month $t+1$ of the daily residual returns from a four-factor model. The four factors are: the market excess return, the size and value factors (Fama and French, 1993), and the momentum factor (Carhart, 1997). Factors are local i.e., they are associated to the country (or region) of stocks. Stock price data are from Refinitiv EIKON, whereas daily factors are from AQR's website. The independent variables include $SSCAP$, which is the net (scaled) capital of the stock pair shorted by common short sellers, and controls at time t, as well as the interaction of $SSCAP$ with three illiquidity dummies: $DAMIHUD$, $DTRUN$, $DFLOAT$. $SSCAP$ is constructed using public disclosure data from the UK's FCA, so it captures common net short positions larger than the regulatory threshold of 0.5% share capital. $DAMIHUD$ is equal to one if both stocks in a stock pair have an Amihud (2002) measure above the cross-sectional median, and zero otherwise. $DTRUN$ is equal to one if both stocks have an average daily turnover below the cross-sectional median. $DFLOAT$ is equal to one if both stocks have free float (as percentage of shares outstanding) below the cross-sectional median. Stock data used to compute the illiquidity measures are from Refinitiv EIKON. Estimates for the remaining controls may be found in the Internet Appendix. All independent variables are updated quarterly and, with the exception of dummy variables, are cross-sectionally normalised (to have zero mean and unit standard deviation), which we denote by $*$. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 3 lags (one quarter).

This is consistent with the empirical short selling literature. Shkilko et al. (2012), for example, find that long sales (buys) drive prices more than short sales (covers). This interpretation would also reconcile our results with those of Antón and Polk (2014) and Bartram et al. (2015) on common ownership. It would mean that the price pressure mechanism underlines the relationship between common ownership and comovement, but not that between common short selling and comovement.

On the other hand, the result that the effect of $SSCAP$ vanishes for the most illiquid stock pairs is more puzzling.

One possible explanation for this result is that short sellers act as liquidity providers for illiquid stocks, alleviating upward price pressure and, hence, reducing excess comovement. Evidence of liquidity provision by short sellers has been found by Boehmer and Wu (2013) and Diether et al. (2009a), among others.

Another possibility is that short sellers avoid illiquid stock pairs because of high transaction costs or to reduce risks of being caught in a short squeeze (Dechow et al., 2001). Boehmer et al. (2010) find that lightly shorted stocks are less liquid, whereas factor models have less explanatory power for these stocks. Although our regression specifications control for the difficulty to short through firm size of stock pairs, this remains an alternative explanation for our liquidity results.

Lastly, the result could be due to short sellers trying to trade illiquid stocks less aggressively than liquid stocks to either conceal information or to minimise price impact. While this type of stealth behaviour is known to occur among long traders, evidence of it occurring among short sellers is less conclusive. With US shorting flow data, for example, Boehmer et al. (2008) find that informed short sellers will favour large trades over many small or medium-sized trades. With European disclosure data, Jank et al. (2021) note that short sellers bunch positions under the disclosure threshold. However, they find that bunching is motivated by disclosure avoidance and unrelated to short sellers' stock liquidity concerns.

To shed light on these alternative hypotheses, we would require short positions under the 0.5% disclosure thresholds. This would allow us to examine whether the relationship between $SSCAP$ and correlation strengthens or weakens for illiquid stocks under the threshold and, accordingly, to exclude some of the different possible interpretations discussed.

Nonetheless, these results call for further examination of the effect of $SSCAP$ to explain the observed relationship between common short selling and excess comovement. Hence, in the next section, we investigate to Hypothesis 2.

5.3 Informed Trading

Hypothesis 2 posits that the effect of $SSCAP$ should be stronger for informed short sellers. With this in mind, we investigate $SSCAP$ originating from different types of short seller.

First, from Refinitiv EIKON, we obtain the investor profiles for 323 out of the 454 short sellers disclosing short positions against our sample of stocks. We collect the following information: investment orientation, investor type, portfolio turnover (%), and number of instruments held.[19]

According to Aragon and Martin (2012) and Agarwal et al. (2013), hedge funds are highly informed agents. Thus, we begin by distinguishing short sellers between hedge funds and non-hedge funds.[20] We classify short sellers in our sample as hedge funds if, according to their Refenitiv EIKON investor profile, their investor type is either "Hedge Fund" or "Investor Advisor/Hedge Fund". According to this classification system, our sample of short sellers comprises 234 hedge fund and 89 non-hedge funds.[21]

The first two columns of Table 9 show the effect of $SSCAP$ for short sellers that are classified as hedge funds and non-hedge funds. The regression coefficient on $SSCAP$ for hedge funds is larger than that on non-hedge funds and, in the most complete specification, this difference is statistically significant at the 10% significance level (p-value at 8.9%).

We obtain similar results when we distinguish short sellers based on their investment orientation. Refinitiv EIKON classifies investors' investment orientation as either active or passive. Active investors are more prone to stock-picking and using proprietary trading strategies, whereas passive investors involve less buying and selling and more long-haul investments. Our sample of short sellers includes 276 active investors and 47 passive investors.

[19]Investor profiles are not available historically, so they refer to the date of collection (March 2020).

[20]We do not explicitly make the distinction between hedge funds and hedge fund managers.

[21]Using SEC ADV forms paired with the methodology of Brunnermeier and Nagel (2004) and Griffin and Xu (2009), we obtain a similar classification, with 75% concordance.

Table 9: Fama-MacBeth cross-sectional regressions of monthly realised pairwise correlation of daily six-factor residual returns on *SSCAP* by short seller type and stock pair control variables, Jan. 2013—Dec. 2019

	Dependent Variable: Correlation of 6F Residuals							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hedge Fund	0.00089	0.00120						
	(2.20)	(3.08)						
Non-Hedge Fund	0.00034	0.00043						
	(1.67)	(2.08)						
Active			0.00091	0.00123				
			(2.15)	(2.99)				
Passive			0.00034	0.00040				
			(2.00)	(2.41)				
High Turnover					0.00028	0.00038		
					(1.83)	(2.56)		
Low Turnover					0.00062	0.00061		
					(1.37)	(1.38)		
High Concentration					0.00024	0.00026		
					(1.90)	(2.00)		
Low Concentration					0.00047	0.00080		
					(0.85)	(1.52)		
High Performance							0.00120	0.00153
							(3.20)	(4.15)
Low Performance							0.00002	0.00013
							(0.10)	(0.56)
Other controls reported in the Internet Appendix								
Group difference	0.00052	0.00074	0.00057	0.00083			0.00118	0.00140
	(1.18)	(1.72)	(1.21)	(1.75)			(3.46)	(4.01)

Continued on next page

Table 9—*Continued*

	Dependent Variable: Correlation of 6F Residuals							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R2$	0.07935	0.08216	0.07934	0.08215	0.07915	0.08197	0.07935	0.08216
	(9.87)	(9.96)	(9.87)	(9.96)	(9.83)	(9.92)	(9.87)	(9.96)
No. Obs.	49,871	49,871	49,871	49,871	49,981	49,981	49,871	49,871
	(106.78)	(106.78)	(106.78)	(106.78)	(103.37)	(103.37)	(106.78)	(106.78)
Size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair characteristic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style controls	No	Yes	No	Yes	No	Yes	No	Yes
Dummy controls	No	Yes	No	Yes	No	Yes	No	Yes

The table reports Fama and MacBeth (1973) regression coefficients, computed by running cross-sectional regressions each month between Jan. 2013 and Dec. 2019 (84 months) and then averaging regression coefficients over the sample period. Observations relate to stock pairs from an unbalanced sample of 356 stocks. The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. The dependent variable is the realised pairwise correlation in month $t+1$ of the daily residual returns from a four-factor model. The four factors are: the market excess return, the size and value factors (Fama and French, 1993), and the momentum factor (Carhart, 1997). Factors are local i.e., they are associated to the country (or region) of stocks. Stock price data are from EIKON Refinitiv, whereas daily factors and the Treasury bill rate are from AQR's website. The independent variables include $SSCAP$ by short seller type, which is the net (scaled) capital of the stock pair shorted by common short sellers of a given type, and controls at time t. Each row reports the coefficient on $SSCAP$ for different types of short sellers. $SSCAP$ by short seller type is constructed using public disclosure data from the UK's FCA, so it captures common net short positions larger than the regulatory threshold of 0.5% share capital. To determine short seller type, we identify short sellers from the regulatory disclosure data and match them with investor profile information from Refinitiv EIKON. Estimates for the remaining controls may be found in the Internet Appendix. All independent variables are updated quarterly and, with the exception of dummy variables, are cross-sectionally normalised (to have zero mean and unit standard deviation). t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 3 lags (one quarter).

The third and fourth columns of Table 9 show that common short positions of active investors are more strongly associated with future excess correlation than those of passive investors. Again, the difference is statistically significant at the 10% significance level in the most complete specification. As active investing requires processing extensive information rapidly, these results are in line with the informed trading explanation.

Our sample of short sellers is skewed towards hedge funds and active investors. This could be problematic for our analysis as it could lead to an unbalanced comparison between $SSCAP$ for hedge funds (active investors) against $SSCAP$ for non-hedge funds (passive investors). To mitigate this issue, we use an alternative definition of activeness. Specifically, we separate short sellers according to their portfolio turnover and concentration, keeping the size of groups balanced.[22]

We define a short seller as a high (low) turnover entity if it ranks above (below) the cross-sectional median in terms of portfolio turnover. Similarly, we define high (low) concentration those entities that rank below (above) the cross-sectional median in terms of number of stocks in their investment portfolio.[23]

The fifth and sixth columns of Table 9 show that the relationship between common short selling and future excess return correlation is stronger for short sellers that are high turnover and high concentration investors. These type of agents often re-balance their portfolios and invest in few specific stocks.

These results point towards common short positions of informed short sellers having a stronger association with future excess comovement. However, this rests on the implicit assumption that hedge funds, active investors, or investors with high turnover and high concentration are more informed than non-hedge funds, passive investors, or investors with low turnover and low concentration.

An alternative explanation consistent with our results could be that $SSCAP$ signals agreement between short sellers and hence greater overall disagreement between short sellers and long investors. Higher disagreement can lead to more volatility and hence greater

[22]The issue remains a concern if hedge funds or active investors with high turnover are more active in the shorting market than non-hedge funds or passive investors with low turnover. In Table A.3 of the Internet Appendix, we verify that, except for high/low concentration groups, the number and size of short positions does not vary significantly between the groups of short sellers.

[23]We obtain similar results if we use terciles, rather than medians, to group short sellers.

covariation in stock returns. This disagreement mechanism might be strongest when short sellers are "opinionated" agents, such as active investors or ones with high turnover.

To address this concern, we repeated the analysis with a different measure of informativeness, that should be less susceptible to the disagreement mechanism—short sellers' past performance. The informativeness of short positions is often measured by their ability to predict future returns (see, for example, Christophe et al., 2004). In first instance, opinionatedness should not be related to performance.

With this in mind, at each period, we computed short sellers' performance over the previous 12 months based on the disclosure data.[24] We labelled short sellers as having high (low) performance if, over the preceding 12 months, their short portfolio returns were above (below) the cross-sectional median.[25] Then, using these two groups, we constructed $SSCAP$ for high performance and low performance short sellers.

The seventh and eighth columns show that common short positions of short sellers with high past performance are better able to predict future excess comovement. The difference between high and low performance short sellers is highly significant, at the 1% confidence level in both specifications. Furthermore, the size of coefficient on $SSCAP$ for high performance short sellers is larger than that found for hedge funds and for active investors.

We verified the robustness of results to alternative specifications. Table A.14 in the Internet Appendix reports regressions using, as dependent variable, the correlation of six-factor residual returns from the AQR model. Results are close to those presented in Table 9. Alternative factor models, such as those presented in Section 4.2, obtain similar outcomes.

6 Further Evidence

6.1 Price Reversals

The informative content of common short positions can be also assessed by analysing the evolution of returns after the positions are taken. Particularly, we would expect non-informative

[24] To construct the short portfolio based on disclosure data, we followed the conservative equally-weighted approach of Greppmair et al. (2020).

[25] We obtain similar results if we use terciles, rather than medians, to group short sellers.

short positions to be associated with price reversals (Boehmer and Wu, 2013). This is because, generally, price reversals are transient and not information based. Thus, informed short sellers should, in theory, not position themselves against these events.

To adopt such an analysis in our context of common short positions, we define the connected portfolio i as the portfolio of stocks that are connected, through one or more common short seller, to stock i. The return on the equally-weighted connected portfolio of stock i is computed as the average of all the connected stocks of i. That is,

$$r_{iC,t}^{EQ} = \frac{\sum_{j=1}^{J} \mathbb{I}(SSCAP_{ij,t-1}) r_{j,t}}{\sum_{j=1}^{J} \mathbb{I}(SSCAP_{ij,t-1})}, \tag{3}$$

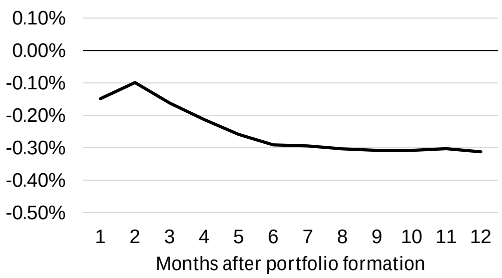
where $\mathbb{I}(\cdot)$ is the indicator function, which is equal to one if it's argument is positive and zero otherwise, and J is the number of stocks connected through common short positions to i. With the goal of verifying whether portfolios of connected stocks are associated with price reversals, we analyse the cumulative buy-and-hold abnormal returns of the connected portfolios over the 12 months after portfolio creation—$t+1, t+2, ..., t+12$.

Specifically, we retrieve the abnormal returns of the connected portfolio by regressing excess portfolio returns on the global excess-market return and on global FFC factors. To avoid weighing our results on the more recent part of the sample, during which more stocks are present, we focus our analysis on 195 stocks for which a balanced sample of observations is available.

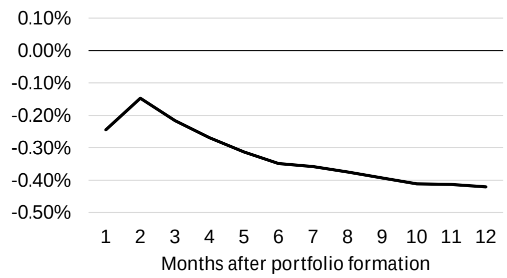
Figure 3 presents the average abnormal cumulative returns for the connected portfolios. Chart A shows that equally-weighted portfolios of connected stocks earn, on average, negative abnormal returns that persist for over one year. These cumulative returns reach -0.30% after six months, and then appear to stabilise. Rather than revert, cumulative abnormal returns persist. This lends evidence to the informative trading story of Hypothesis 2. Stocks connected through common short positions are associated with permanent, rather than transient, price shifts.

Figure 3: Monthly cumulative abnormal returns of connected portfolios, Jan. 2013—Dec. 2019

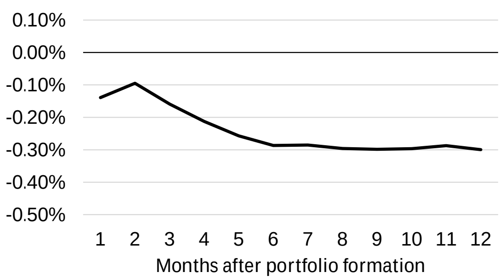
A. Equally-weighted portfolios



B. Value-weighted portfolios



C. $SSCAP$-weighted portfolios



The charts present the average monthly abnormal cumulative returns of connected portfolios using data on a balanced sample of 195 stocks from January 2013 to December 2019 (84 months). The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. We define the connected portfolio i as the portfolio of stocks that are connected, through one or more common short seller, to stock i. Buy-and-hold cumulative abnormal returns of these portfolios are computed over the twelve months after portfolio formation. Abnormal returns are retrieved by regressing portfolio returns in excess of the U.S. T-bill rate on the global market excess return, the global size and value factors (Fama and French, 1993), and the global momentum factor (Carhart, 1997). Price data are from Refinitiv EIKON, whereas factor data are from AQR. Chart A depicts the monthly abnormal cumulative return of connected portfolios computed using an equal weighting of stocks within each portfolio. Chart B and Chart C depict the monthly abnormal cumulative return of connected portfolios computed using, respectively, value weighting and weights based on common short sold capital. Common short sold capital is constructed using public disclosure data from the UK's FCA, so it captures common net short positions larger than the regulatory threshold of 0.5% share capital.

These results continue to hold when we use alternative portfolio weights for the connected stocks. Chart B of Figure 3 shows the average monthly four-factor abnormal returns of connected portfolios constructed using value weights, whereas Chart C depicts the same measure for portfolios constructed using $SSCAP$ weights. In the latter case, the return on connected portfolio of stock i is computed as

$$r_{iC,t}^{SSCAP} = \frac{\sum_{j=1}^{J} SSCAP_{ij,t-1} r_{j,t}}{\sum_{j=1}^{J} SSCAP_{ij,t-1}}. \tag{4}$$

In both cases, cumulative abnormal returns are negative and persistent.

In Figure A.2 of the Internet Appendix, we present robustness of this result to alternative definitions of abnormal returns. If we use simple cumulative excess returns above the U.S. T-bill rate, we observe reversal after about nine months. If we look at cumulative returns in excess of the market using the CAPM model, we obtain a much flatter picture, indicating even slower reversion. Paired with Figure 3, these results indicate that, when benchmarked against the market and FFC factors, connected portfolio returns are associated with persistently negative cumulative returns that are slow to revert. This result is further confirmed when abnormal returns are defined using the six-factor AQR model.

6.2 Portfolio Diversification

Finally, we explore the possibility of using the our results on common short selling and comovement to improve the diversification of portfolios.

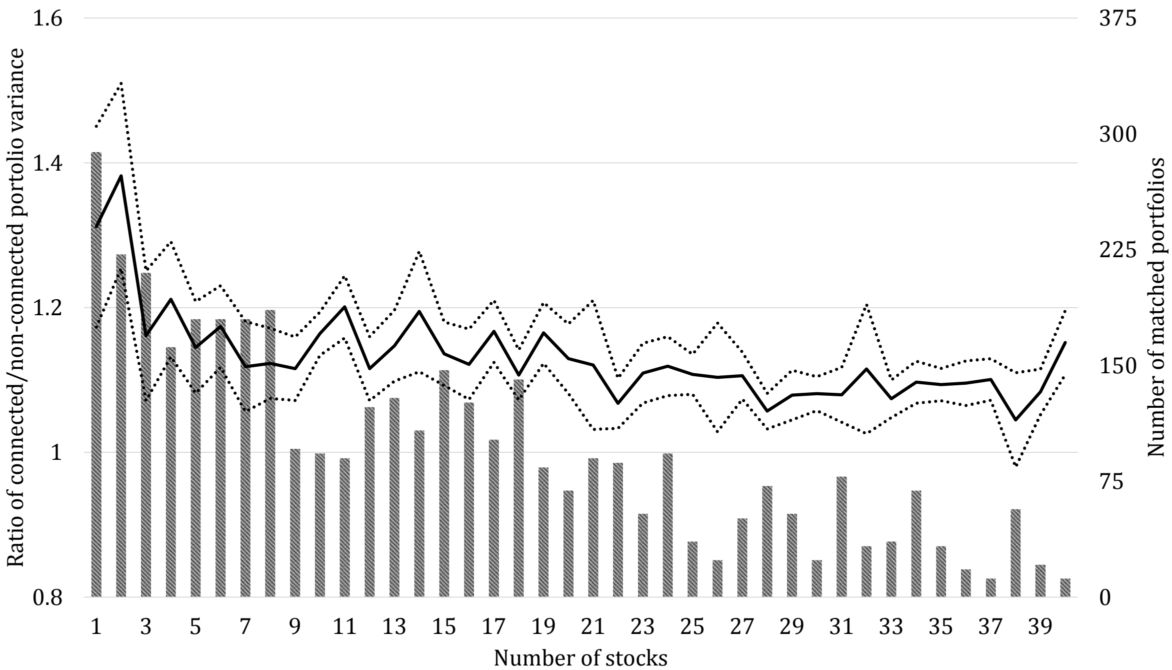
At the beginning of each month $t + 1$, we construct the connected portfolio, for every stock i, using the non-zero entries of $SSCAP_{ij,t}$. We then study the volatility of daily excess returns of equally-weighted connected portfolios for the successive 253 days (without re-balancing). As a mean of comparison, every month, we also construct non-connected portfolios of matching stocks.

Specifically, our non-connected portfolios contain stocks that, according to the disclosure data, at period t, had no common short sellers. As a matching criteria, we impose that the stocks in the non-connected portfolios belong to the same size deciles as the stocks in the corresponding connected portfolio.[26] We then compute the equally-weighted daily returns, in excess of the U.S. T-bill rate, of these connected and non-connected portfolios portfolios,

[26] Attempting to match over more characteristics, does not return a sufficient number of matches. However, we note that size may act as a proxy for other characteristics.

as well as their realised yearly volatility.

Figure 4: Ratio of yearly realised excess return volatility of equally-weighted connected and non-connected portfolios, Jan. 2013—Dec. 2019



The chart plots the average ratio of yearly realised volitities of the returns of matched connected portfolios and non-connected portfolios (bold solid, left axis), as well as the uncertainty bands (light dashed, left axis) representing 2.5 standard deviations from the average. At the beginning of every month $t + 1$, we define connected portfolio i as the portfolio of stocks that, at quarter-end t, are connected, through one or more common short seller, to stock i. At the same, for each connected portfolio, we defined a matching non-connected portfolio of stocks that belong to the same size deciles as those in the connected portfolio and do not share any common short seller at quarter-end t. We computed the daily equally-weighted excess return of these portfolios and compared their realised volatility over the successive 253 trading days. The figure also plots the number of matched portfolios over which the average volatility ratio is calculated (bars, right axis). Portfolios are constructed from a balanced sample of 195 stocks from January 2013 to December 2019 (84 months). The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. Excess returns are computed using price data from Refinitiv EIKON and the U.S. Treasury bill rate. We determine common short selling connections using public disclosure data from the UK's FCA, which captures common net short positions larger than the regulatory threshold of 0.5% share capital. Uncertainty bands are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 12 lags (one year).

Figure 4 depicts the average of the ratio of the connected and non-connected portfolio excess return volatilities. As the size of the portfolio increases i.e., as the number of (matched) stocks in the portfolios grows, diversification of connected portfolios strengthens and the average ratio tends towards one. However, when both portfolios include 40 stocks, the ratio of the volatilities still hovers around 1.1. This means that choosing stocks without common short sellers obtains substantial diversification benefits.

Table 10: Summary statistics of yearly realised volatilities of matching connected and non-connected portfolios, Jan. 2013—Dec. 2019

	Portfolio Yearly Realised Volatilities		
	Connected Portfolios	Non-Connected Portfolios	% of Rejections
Mean	0.0138	0.0120	29.77
	(65.81)	(64.08)	(24.34)
Median	0.0118	0.0105	

The table reports the mean and median volatilities for 3,870 matched connected and non-connected portfolios, as well as the rejection frequency of Levene's (1960) absolute deviation test on equality of variances. At the beginning of every month $t+1$, we define connected portfolio i as the portfolio of stocks that, at quarter-end t, are connected, through one or more common short seller, to stock i. At the same, for each connected portfolio, we defined a matching non-connected portfolio of stocks that belong to the same size deciles as those in the connected portfolio and do not share any common short seller at quarter-end t. We computed the daily equally-weighted returns of these portfolios in excess of the U.S. T-bill rate and computed the realised volatility of excess returns over the 253 trading days after portfolio formation. Portfolios are constructed from a balanced sample of 195 stocks from January 2013 to December 2019 (84 months). The sample includes stocks that have at least one public short selling disclosure reported by the UK's FCA, subject to sufficient stock data being available. Excess returns are computed with respect of the U.S. T-bill rate. Daily stock price data are from Refinitiv EIKON. We determine common short selling connections using short selling disclosure data from the UK's FCA, which captures common net short positions larger than the regulatory threshold of 0.5% share capital. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 12 lags (one year).

The chart also depicts the number of portfolios used to compute the average ratio. Because it becomes more difficult to fulfill the matching criteria for large portfolios, the number of portfolios declines as portfolio size increases. Inevitably, this means that results are less precise for portfolios of large sizes. However, for small portfolio sizes, the result hints that there are diversification benefits to grouping stocks with no common short sellers.

Table 10 reports the average and median volatilities across 3,870 connected and non-connected portfolios. The non-connected portfolio has, on average, a 13.6% smaller volatility than connected portfolios. This difference is statistically significant at 1% confidence level.

We verified the difference in riskiness between connected and non-connected portfolios by conducting tests on the equality of variances. We used Levene's (1960) absolute deviation test, which is more robust to departures from normality than the classic F-test. The alternative hypothesis to the test is that the connected portfolio variance is not equal to that of the matching non-connected portfolio.

Table 10 shows that, for almost 30% of the matched portfolios, the test rejected the null hypothesis of equality of variances at 5% confidence level.[27]

These results are robust to various alterations to our analysis. First, we repeated the analysis with alternative matching criteria, based on TRBC Economic Sector. Second, we varied the weighting of stocks in the connected and non-connected portfolios. Specifically, we adopted value-weights and weights based on common short sold capital, $SSCAP$. Third, we varied the length of stock holding period, increasing it to two years and decreasing it to six-months. Lastly, we repeated the analysis using residual returns from different factor models (CAPM, FFC, and AQR).

These robustness results are presented in Figure A.3 and Table A.16 of the Internet Appendix. Results are in line with those presented in this section. In particular, the rejection frequencies for the tests on the equality of variances across connected and non-connected portfolios vary between 27% and 35%.

7 Conclusion

Overall, our results provide a helpful tool for investors and portfolio managers to estimate future correlation. If two stocks are connected by common short sellers, they have stronger comovement than two stocks that are not connected by common short sellers. We show that this result can be used to obtain portfolio diversification benefits. Our results should be also reassuring for financial markets regulators, as the weight of the evidence tends towards the informed trading hypothesis more than the price pressure hypothesis. However, our analysis is not immune to some key limitations that we discuss together with their relevance for future extensions of this study.

First, despite the wealth of our controls and our robustness checks, we cannot completely exclude that our results are not related to fundamental comovement. It is still possible that short sellers trade according to some unobserved stock characteristics that we do not capture in our framework. We simply cannot rule out that this effect might be more pronounced

[27]Using the classical F-test, with the alternative that the connected portfolio volatility is greater than that of the matching non-connected portfolio, the rejection frequency increases to 52.2%.

for hedge funds, active investors (or ones with high turnover and high concentration), and highly performing short sellers.

To exclude this possibility completely, we would ideally have an exogenous event that induces a change in $SSCAP$, without directly affecting comovement. However, due to our brief sample period, and the nature of our data, we lack this sort of setting.

In any case, this does not change the significance of our results and the predictive power of $SSCAP$ for future correlation. In fact, it would mean that $SSCAP$ proxies for unobserved fundamental factor not captured by other measures widely used in the literature. Future work could then focus on determining the drivers of common short position and analysing whether these are related to discount rates or cashflow news. In this sense, the evidence we gathered on hedge funds and informed trading would prove a useful starting point.

Second, the fact that our data is restricted large short position disclosures covering just one market leads to some additional considerations.

Although sample restrictions undoubtedly limit the global validity of the results, we have presented several robustness specifications that strengthen their local validity for an important market, such as the UK. Recently Boehmer et al. (2021) have uncovered cross-country differences in the predictive power of short selling data for stock price returns. Given these findings, future extensions could verify the results of our study in different settings.

Furthermore, despite we capture a good portion of the short selling picture, due to the European disclosure rule, we are missing a large number of smaller short positions below the 0.5% threshold. In it's 2018 Report on Trends, Risks and Vulnerabilities, ESMA notes that EU positions above the 0.5% threshold represent less than one third of all positions above the 0.2% threshold (at which institutions privately report to national authorities). The literature has found that large short positions are generally informed, and that the informativeness increases with position size (Avramov et al., 2006, Boehmer et al., 2008, Easley and O'Hara, 1987). This is in line with our result that common short selling relates to excess comovement through the informed trading. Thus, a further path for future work would be to verify our results with smaller short positions, using data exclusively disclosed to national regulators.

References

V. Agarwal, W. Jiang, Y. Tang, and B. Yang. Uncovering hedge fund skill from the portfolio holdings they hide. *The Journal of Finance*, 68(2):739–783, mar 2013. doi:10.1111/jofi.12012.

Y. Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, jan 2002. doi:10.1016/s1386-4181(01)00024-6.

M. Antón and C. Polk. Connected stocks. *The Journal of Finance*, 69(3):1099–1127, may 2014. doi:10.1111/jofi.12149.

G. O. Aragon and J. S. Martin. A unique view of hedge fund derivatives usage: Safeguard or speculation? *Journal of Financial Economics*, 105(2):436–456, aug 2012. doi:10.1016/j.jfineco.2012.02.004.

C. S. Asness, A. Frazzini, and L. H. Pedersen. Quality minus junk. *Review of Accounting Studies*, 24:34–112, 2018.

D. Avramov, T. Chordia, and A. Goyal. The impact of trades on daily volatility. *Review of Financial Studies*, 19(4):1241–1277, 2006.

N. Barberis, A. Shleifer, and J. Wurgler. Comovement. *Journal of Financial Economics*, 75 (2):283–317, feb 2005. doi:10.1016/j.jfineco.2004.04.003.

S. M. Bartram, J. M. Griffin, T.-H. Lim, and D. T. Ng. How important are foreign ownership linkages for international stock returns? *The Review of Financial Studies*, 28(11):3036–3072, 2015.

E. Boehmer and J. Wu. Short Selling and the Price Discovery Process. *Review of Financial Studies*, 26(2):287–322, February 2013. ISSN 1465-7368. doi:10.1093/rfs/hhs097. URL `http://dx.doi.org/10.1093/rfs/hhs097`.

E. Boehmer, C. M. Jones, and X. Zhang. Which Shorts Are Informed? *Journal of Finance*, 63(2):491–527, Apr 2008. doi:10.1111/j.1540-6261.2008.01324.x.

E. Boehmer, Z. R. Huszár, and B. D. Jordan. The Good News in Short Selling. *Journal of Financial Economics*, 96(1):80–97, 2010. doi:10.1016/j.jfineco.2009.12.002.

E. Boehmer, T. X. Duong, and Z. R. Huszár. Short covering trades. *Journal of Financial and Quantitative Analysis*, 53:723–348, 2018.

E. Boehmer, Z. R. Huszaár, Y. Wang, X. Zhang, and X. Zhang. Can shorts predict returns? A global perspective. SSRN Working Paper 2668357, Feb. 2021.

M. K. Brunnermeier and S. Nagel. Hedge funds and the technology bubble. *The Journal of Finance*, 59(5):2013–2040, oct 2004. doi:10.1111/j.1540-6261.2004.00690.x.

M. K. Brunnermeier and L. H. Pedersen. Predatory trading. *Journal of Finance*, 60(4): 1825–1863, 2005.

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

H. Chen, S. Chen, Z. Chen, and F. Li. Empirical investigation of an equity pairs trading strategy. *Management Science*, 2017. doi:https://doi.org/10.1287/mnsc.2017.2825.

S. E. Christophe, M. G. Ferri, and J. J. Angel. Short-selling prior to earnings announcements. *The Journal of Finance*, 59(4):1845–1875, 2004.

R. Cont and L. Wagalath. Running for the exit: Distressed selling and endogenous correlation in financial markets. *Mathematical Finance*, 4:718–741, 2013.

A. Curtis and N. L. Fargher. Does short selling amplify price declines or align stocks with their fundamental values? *Management Science*, 60(9):2324–2340, 2014.

P. M. Dechow, A. P. Hutton, L. Meulbroek, and R. G. Sloan. Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, 61:77–106, 2001.

K. B. Diether, K. Lee, and I. M. Werner. It's SHO time! Short–sale price tests and market quality. *The Journal of Finance*, 64:37–73, 2009a.

K. B. Diether, K.-H. Lee, and I. M. Werner. Short-sale strategies and return predictability. *Review of Financial Studies*, 22(2):575–607, 2009b.

D. Easley and M. O'Hara. Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19:69–90, 1987.

R. Engle and R. Colacito. Testing and valuing dynmaic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2):238–253, 2006.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47:427–465, 1992.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

E. F. Fama and J. D. MacBeth. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636, 1973.

J. D. Finnerty. Short selling, death spiral convertibles, and the profitability of stock manipulation. SEC working paper, 2005. URL `https://www.sec.gov/comments/s7-08-08/s70808-318.pdf`.

A. Frazzini and L. H. Pedersen. Betting against beta. *Journal of Financial Economics*, 111: 1–25, 2014.

R. Greenwood. Excess comovement of stock returns: Evidence from cross-sectional variation in nikkei 225 weights. *Review of Financial Studies*, 21(3):1153–1186, oct 2007. doi:10.1093/rfs/hhm052.

S. Greppmair, S. Jank, P. A. C. Saffi, and J. Sturgess. Securities lending and information acquisition. Working paper, July 2020. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3604977.

J. M. Griffin and J. Xu. How smart are the smart guys? A unique view from hedge fund stock holdings. *Review of Financial Studies*, 22(7):2531–2570, may 2009. doi:10.1093/rfs/hhp026.

A. Hameed, R. Morck, J. Shen, and B. Yeung. Information, analysts, and stock return comovement. *Review of Financial Studies*, 28(11):3153–3187, 2015.

G. Huberman, S. Kandel, and G. A. Karolyi. Size and industry-related covariations. IIBR Working Paper No. 960/88, Feb. 1988.

S. Jank and E. Smajlbegovic. Dissecting short-sale performance: Evidence from large position disclosures. 2017.

S. Jank, C. Roling, and E. Smajlbegovic. Flying under the radar: The effects of short-sale disclosure rules on investor behavior and stock prices. *Journal of Financial Economics*, 139:209–233, 2021.

C. M. Jones, A. V. Reed, and W. Waller. Revealing shorts: An examination of large short position disclosures. *The Review of Financial Studies*, 29(12):3278–3320, 2016.

H. Levene. Robust tests for equality of variances. In I. Olkin and H. Hotelling, editors, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 278–292. Stanford University Press, 1960.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

R. S. Pindyck and J. J. Rotemberg. The comovement of stock prices. *The Quarterly Journal of Economics*, 108(4):1073–1104, 1993.

C. Pirinsky and Q. Wang. Does corporate headquarters location matter for stock returns? *The Journal of Finance*, 61(4):1991–2015, aug 2006. doi:10.1111/j.1540-6261.2006.00895.x.

L. C. G. Rogers and F. Zhou. Estimating correlation from high, low, opening and closing prices. *Annals of Applied Probability*, 18(2):813—823, 2008.

B. Rosenberg, K. Reid, and R. Lanstein. Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11:9–17, 1985.

A. Shkilko, B. Van Ness, and R. Van Ness. Short selling and intraday price pressures. *Financial Management*, 41(2):345–370, apr 2012. doi:10.1111/j.1755-053x.2012.01189.x.