Short Selling and Excess Return Correlation*

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

†*Cambridge-INET Institute, University of Cambridge*

‡*CeReFiM, Université de Namur*

⋆*Vlerick Business School and Ghent University*

First version: 29 October 2017

This version: December 3, 2017

Download the latest version: https://www.dropbox.com/s/74ctb5eozxeufqc/short_corr.pdf?dl=0

Abstract

We show that the number of common short sellers shorting two stocks can predict their four-factor excess return correlation one month ahead, controlling for many pair characteristics, including similarities in size, book-to-market, and momentum. We verify that this result holds out-of-sample and show that it can be used to establish a trading strategy that yields positive cumulative returns over 12 months. We explore the possible mechanisms that could give rise to this relationship. We find that neither the price-impact mechanism nor the liquidity-provision mechanism can explain the uncovered relationship. Rather, it seems that the relationship is due to informed short selling, which we identify using several indicators of value obtained from financial statement analyses.

Keywords: short selling, correlation, informed trading

JEL Classification: G14

*Marco Valerio Geraci and Jean-Yves Gnabo acknowledge the support of the Communauté française de Belgique through the Projet d'Actions de Recherche Concertées 13/17-055 grant. Corresponding author email: mvg23@cam.ac.uk.

I Introduction

In their theoretical setting, Cont and Wagalath (2013) show that, by heavily short selling two or more assets, short sellers can induce positive comovement in their prices, even if the assets are fundamentally uncorrelated. In this paper, we empirically investigate this claim by analysing the association between short selling of stock pairs and their future realized correlation.

We use public short selling disclosures made to the Financial Conduct Authority (FCA) of the United Kingdom to connect stocks based on their common short sellers. We find that the number of common short sellers can predict abnormal stock return correlation one month ahead, controlling for similarities in size, book-to-market, momentum, and several other common characteristics. In our most flexible specification, a standard deviation increase in the number of common short sellers is associated with a future rise of 2.13% of the average four-factor excess return correlation for a given stock pair. We show that when predicting correlation our variable capturing the number of common short sellers can lead to significant out-of-sample forecast gains. Moreover, we show that the forecasted correlation can be used to establish a trading strategy that can yield positive cumulative abnormal returns of over 9% after 12 months, gross of transaction costs.

Several possible mechanisms can explain the positive association between short selling and future correlation because, fundamentally, several reasons may lead a short seller to initiate a short position (see e.g., Dechow, Hutton, Meulbroek, and Sloan (2001)). We investigate three mechanisms which we believe are most important and determine whether they have support in our data.

Short selling could induce higher correlation by applying negative price pressure on several stocks. This price impact effect should be stronger for illiquid stocks (Brunnermeier and Pedersen, 2005, Brunnermeier and Oehmke, 2014). We use this prediction to verify whether there is evidence of the price impact effect in our data. Specifically, we test whether the uncovered association between the number of common short sellers and future correlation is stronger for less liquid stocks. Our empirical results do not find support for this mechanism. At least at the frequency and periodicity of our study, the positive association between short selling and future excess correlation does not appear to corroborate the price impact of short selling.

Further, this result can help us clarify the role of a second possible previously studied mechanism—the liquidity provision mechanism. Short sellers may be acting as voluntary liquidity providers in times of high excess demand for low-liquidity stocks (Boehmer and Wu, 2013). When excess demand subsequently reverts, correlation should increase as prices of those same short sold stocks fall. In line with Diether, Lee, and Werner (2009), we cannot confirm that the liquidity provision mechanism is operating in our data.

As an alternative, our results could be a by-product of informative trading strategies of short sellers. Ultimately, by initiating a short position, a short seller predicts that stock prices will decline in the future to gain a positive return from the trade. Previous studies have shown that short sellers are sophisticated market agents, who trade on the basis of superior information. Christophe, Ferri, and Angel (2004) find that short sellers can predict future disappointing earnings announcements, and Lynch, Nikolic, Yan, and Yu (2013) show that short sellers can predict negative market-wide returns. In addition, Curtis and Fargher (2014) show that short sellers focus on overpriced stocks according to several fundamental value-to-price ratios.

If common short sellers trade according to superior information that stock prices should decline in the future, this will coincide with higher future correlation across the shorted stocks. Thus, if this mechanism is occurring, the positive relationship between common short selling and future excess correlation can be explained by informative trading. To verify this mechanism, we isolate the number of common short sellers that can be associated with informative trading using several measures of value obtained from financial statement analyses. We then show that the informed number of common short sellers is a strong predictor of future returns, hence corroborating the informative trading mechanism.

To further validate the informative short selling mechanism, we carry out additional analysis to verify if there is evidence of informative trading in our data, whilst controlling for several of the determinants of short selling. We find evidence of both informative trading and non-informative momentum trading occurring. Overall, our results indicate that the informative trading mechanism is a likely explanation for the uncovered relationship between short selling and future correlation.

The literature surrounding short selling has focused primarily on providing evidence that short selling has useful information for predicting future asset returns (Boehmer, Huszar, and Jordan, 2010). To our best knowledge, our study is the first to verify the predictive power of short selling with respect to the comovement of asset returns. This is important as correlation is a key risk metric in finance.

It is well known that correlation tends to move over time, increasing markedly during market events and periods of crisis. A wide range of models have been proposed that treat correlation as an exogenous process unlinked to market developments.[1] Recently, theoretical models have interpreted correlation as an endogenous function of asset demand and supply dynamics, such as speculative attacks, herding, and predatory short selling (Cont and Wagalath, 2013, Yang and Satchell, 2007).

[1]Multivariate generalized autoregressive heteroskedasticity models, for example, treat the covariance of asset returns as a function of past covariance and past realization of returns. Stochastic volatility models, on the other hand, assume that the covariance structure of asset returns is determined by a stochastic process which is independent of returns.

Our paper contributes to the literature on correlation modelling by providing empirical evidence of the association between short selling and future realized correlation. Brunnermeier and Pedersen (2005) and Cont and Wagalath (2013) had posited the possibility that short sellers could trigger sudden shifts in correlation, which have been the focus of the contagion literature (Forbes and Rigobon, 2002, Billio and Pelizzon, 2003, Corsetti and Sbracia, 2005). Whereas we show that it is possible to predict these correlation shifts using short selling information, we cannot confirm the price-impact mechanism that they have put forward in their theoretical studies.

Lastly, most studies on short selling use data on the aggregate amount of short selling activity—measured as short interest, short ratio, or short volume. In this study, we look at actual net short positions using public short selling disclosures made to the Financial Conduct Authority (FCA) of the United Kingdom. Thanks to this data, we are able to connect stocks according to the common short sellers and study the impact of common short selling on their future correlation. This would not have been possible with traditional datasets on short interest or securities lending proxying for short selling. Considering the study by Jones, Reed, and Waller (2016), who analyse the market impact of the European Union (EU) disclosure requirement, our paper is the second to make use of this data and the first to use it for understanding stock price correlations. We focus our study on the UK because it has the largest and most liquid stock exchange in Europe.[2]

The rest of the paper is organized as follows. In Section II, we describe the short selling disclosure data and in Section III we outline the methodology used. In Section IV, we present the results showing the predictive power of the number of common short sellers for forecasting future excess correlation. This section includes an out-of-sample exercise and a trading strategy exploiting the uncovered relationship. In Section V, we investigate three possible mechanisms that can explain our results. In Section VI, we analyse the determinants of the number of common short sellers. In Section VII, we draw our conclusions.

II Data

A Sample

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 relevant market authority—the FCA, in the UK. Furthermore, any short position that passes the

[2]Jones et al. (2016) study the effects of the disclosure regulation for 12 European countries (Austria, Belgium, Finland, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, and the UK) and find that UK disclosures represent over half of their sample disclosures.

threshold of 0.5%, and every change by 0.1% after that, has to be disclosed publicly on the FCA's website. 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. 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 available on the FCA's website, between November 2012 and April 2017. Compared to short interest data, short selling disclosures are actual net short positions obligatorily submitted to the regulator and, therefore, are subject to attentive scrutiny.

The disclosures involve 525 unique stocks and 360 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, 32,110 disclosures were made, which included 4,500 position originations (i.e., the first disclosure of a net short position above 0.5% of the shares outstanding), 23,547 updates (i.e., any increments or decrements of 0.1% of the shares outstanding after the 0.5% threshold), and 4,063 position terminations (i.e., disclosures under the 0.5% and representing the closing of the short position).

Given that we only observed data for the last two months of 2012 and for the first half of 2017, Panel A of Table 1 shows that the number of disclosures increased steadily over the years of our sample. 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 in our data take position on about four different stocks per year. The standard deviation is quite large, with some short sellers taking position on as many as 117 different stocks over one year. We found that the median holding period length of a disclosed short position was of 25 trading days.

The lower part of Panel B in Table 1 shows that, on average, the stocks in our sample had around 3 short sellers per year taking position on them. Although the standard deviation of short sellers per stock was not as large as that of stocks per short seller, we observed that some stocks had as many as 23 different short sellers taking position against them.

For all the stocks that had at least one disclosed short selling position, we searched for historical price data using Thompson Reuters Eikon and for company information using Amadeus Bureau Van Dijk. We managed to match the data for 323 stocks. For these stocks and for the time period covered, we also collected analyst earnings estimates from the Institutional Brokers Estimate System (I/B/E/S). Analyst data will be used to build the controls for pairwise realised correlation.

Table 1: Descriptive statistics regarding disclosure data.

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

Year	Disclosures	Originations	Stocks	Short Sellers
2012	796	324	166	107
2013	4483	615	261	158
2014	5149	715	262	162
2015	7163	1001	279	185
2016	9322	1215	317	215
2017	5197	630	260	169

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	5.5	1	53
	2013	4.1	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	117
	2017	4.9	2	10.4	1	76
# 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	3.2	2	3.1	1	18

(Continued)

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. Panel C shows the matched data sample by company classifier, following the NACE Rev. 2 classification system of the European Commission. The information includes the number of stocks that are part of the sample, the number of short selling disclosures made to the FCA, the number of disclosures that were short selling originations, and the number of short sellers making the disclosures. Fifteen stocks were unclassified according to the data provider.

Our final matched sample involves 27,280 disclosed short selling positions. Panel C of Table 1 summarizes the matched sample according to the EU NACE Revision 2 classification system. The sector with the most stocks was the manufacturing sector with 54 stocks, whereas the wholesale and retail trade sector had the most disclosures and short sellers. Sector information is used to control for similarities across stocks in our model, outlined in the next section.

Table 1–*Continued*

Panel C: Sample information by stock NACE Rev. 2 classification.

	Stocks	Disc.	Orig.	Short Sellers
B. Mining and quarrying	40	4477	592	281
C. Manufacturing	54	3479	478	278
D. Electricity, gas, steam and air conditioning	3	305	57	20
E. Water supply; sewerage, waste management	3	96	21	13
F. Construction	21	2022	281	153
G. Wholesale and retail trade	36	5198	673	366
H. Transportation and storage	14	1009	144	87
I. Accommodation and food service activities	6	807	106	77
J. Information and communication	35	2354	359	193
K. Financial and insurance activities	30	1222	173	115
L. Real estate activities	13	803	102	73
M. Professional, scientific and technical activities	26	2412	306	181
N. Administrative and support service activities	17	954	164	93
O. Public administration and defence	3	56	18	7
Q. Human health and social work activities	3	1022	86	42
R. Arts, entertainment and recreation	2	780	109	41
S. Other service activities	2	55	10	9

III Methodology

A 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 construct our main covariate, NSS, from the short selling disclosure data described in Section II.A. Specifically, we let $NSS_{ij,t}$ be the number of common short sellers disclosing a short position in both stocks i and j during the quarter ending at t.

As in Antón and Polk (2014), we work with the cross-sectional variation in NSS. To make the cross-sections comparable, we rank-transform the variable at each quarter and then normalize (to obtain zero mean and unit standard deviation) to ease interpretability. The normalized rank-transformed variable is denoted NSS^*.

We use $NSS^*_{ij,t}$ to forecast the future within-month realized correlation ($\rho_{ij,t+1}$) of each stock pair's daily four-factor excess returns.[3] We control for a vast series of pair characteristics, which are outlined below. All variables on the right-hand side of Equation 1 are updated quarterly, meaning that variables relating to month t contain data ending at the end of the last quarter preceding t.

$$\rho_{ij,t+1} = a + b_s \times NSS^*_{ij,t} + \sum_{k=1}^{n} b_k \times CONTROL_{ij,k} + \epsilon_{ij,t+1} \tag{1}$$

If the number of common short sellers shorting stocks i and j is associated with higher future correlation in the excess returns of stocks i and j, then b_s will be positive and significant.

In order to limit the effect of serial correlation, we estimate b_s using the Fama and MacBeth (1973) regressions i.e., we run Equation 1 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. Therefore, we perform inference on b_s using the Newey and West (1987) robust standard errors, accounting for autocorrelation up to one lag (one month).

B Controls

We included a large set of controls in Equation 1. We followed Antón and Polk (2014) and Chen, Chen, Chen, and Li (2017) to identify the controls that could explain stock return correlations beyond the Fama and French (1993) and Carhart (1997) factors.

Antón and Polk (2014) show that common analyst coverage is highly predictive of future stock return correlation. This is because analysts tend to cover stocks that are similar across several dimensions and stocks that are similar tend to have a higher realized correlation. Furthermore, the similarity captured by common analyst coverage might be different from the similarity captured by the other controls included in our regression. Therefore, we controlled for common analyst coverage of stock pairs using the variable $A_{ij,t}$, which is equal to the number of analysts issuing an earnings forecast of both stocks i and j over the past year. Including A in our regressions helps us control for similarity factors that cause stocks to be correlated and that might contaminate the effect of common short selling.

We attempted to control for industry effects using the first 5 digits NACE (Rev. 2) code,

[3]We used the Fama and French (1993) and Carhart (1997) daily return factors for European markets available from Ken French's website.

which classifies the activities of firms in the EU. From these digits, we created the variable $NUMNACE_{ij,t}$, which captures the number of consecutive equal digits, starting from the first, in the NACE codes of stocks i and j. We also computed a series of additional size, style, and pair characteristic controls.

In terms of size, we controlled for the size of the two companies i and j using their market capitalization. Chen et al. (2017) show that stocks of similar size tend to be more highly correlated. Hence, we captured differences in size using $SAMESIZE_{ij,t}$, which we defined as the negative absolute difference in the cross-sectional percentile ranking of the market capitalization of i and j at the end of the quarter preceding period 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 $SIZE1_{ij,t}$ and $SIZE2_{ij,t}$, which we define as, respectively, the larger and smaller percentile market capitalization of the pair. In our most flexible specifications, we also controlled for their interaction ($SIZE1_{ij,t} \times SIZE2_{ij,t}$) and a series of nonlinear combinations of these variables.

In terms of style, we controlled for similarities in the book-to-market ratio and the momentum of the two stocks. We defined $SAMEBM_{ij,t}$ 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.[4]

It is well known that book-to-market ratios are positively associated with future returns (Stattman, 1980, Rosenberg, Reid, and Lanstein, 1985, Fama and French, 1992). As will be discussed in Section V.B, book-to-market is also a possible driver of informative short selling. Hence, we included the percentile rank of the book-to-market ratio of the two stocks, $BM1_{ij,t}$ and $BM2_{ij,t}$, to keep the effect of common short selling separate from the effect of book-to-market. Furthermore, we included the percentile rank of the momentum of the two stocks, $MOM1_{ij,t}$, and $MOM2_{ij,t}$. We did this because short sellers might ride on declining prices, which are, by definition, more correlated.

Finally, we controlled for a series of stock pair characteristics. To address concerns for potential reverse causality in our regression model, we controlled for the past 2-year monthly correlation of stock pairs, which we denote $RETCORR_{ij,t}$. We also controlled for the past 5-year correlation of the return on equity for every pair and we called this $ROECORR_{ij,t}$. We did this because companies with similar profits are expected to have correlated stock returns (Chen et al., 2017). 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.[5]

[4]The momentum was computed as the cumulative stock return over the last year, excluding the most recent month.

[5]We computed abnormal trading volumes as the residual of the regression of volume on an annual trend and monthly dummies.

We controlled for the absolute difference in the price level of the two stocks, which we denoted $DIFFPRICE_{ij,t}$, as well as the absolute difference in their leverage, $DIFFLEV_{ij,t}$, and in their sales growth, $DIFFGROW_{ij,t}$.

Finally, we created a dummy variable capturing common membership to the FTSE100, denoted $DINDEX$, which should account for any index effects on correlation. Since most of our stocks are of UK companies, we controlled for geographical location using the dummy variable $DCITY$, which measures whether two companies had their headquarters in the same city.

All controls are updated quarterly and, except for the dummy variables, rank-transformed and normalized to make the coefficients more easily interpretable.

IV Results

A Excess stock correlation with common short selling

Table 2 shows the results of the Fama and MacBeth regressions using NSS^* to predict the realized correlation of abnormal returns, as specified in Equation 1.

The first column of Table 2 describes the baseline specification using just NSS^* with a constant. The coefficient on NSS^* is positive and significant, with a coefficient equal to 0.00544. Given that NSS^* is standardized to have zero mean and unit standard deviation, the constant term, which is 0.07001, reflects the average abnormal correlation between stock returns. The coefficient on NSS^* can thus be interpreted with respect to the average abnormal correlation. A standard deviation increase in the number of common short sellers is associated with an increase of the predicted excess return correlation of about 7.8% of the average excess return correlation.

In the second column of Table 2, we show results obtained whilst controlling for analyst coverage, similarity in sector, size, book-to-market, and momentum. The coefficient on analyst coverage is positive and highly significant. Antón and Polk (2014) interpret this result as consistent with analysts covering similar stocks. The most important determinant of correlation appears to be the $NUMNACE^*$, the similarity in sector of the two companies. The coefficient on $NUMNACE^*$ is statistically significant with a coefficient of 0.00874 and a t-statistic of 15.36. Similarity in size and book-to-market do not seem to be as important in this specification, whereas similarity in momentum, captured by $SAMEMOM^*$, is significant with a coefficient of 0.01212 (t-statistic of 8.27). Recall that we have already controlled for exposure to the size, book-to-market, and momentum factor by using the four-factor abnormal returns of Fama and French (1993) and Carhart (1997) to compute realized correlation.

Table 2: Common short selling originations and excess correlation

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.07001	0.08627	0.08747	0.07893
	(15.6)	(7.2)	(7.02)	(4.92)
NSS^*	0.00544	0.0016	0.0013	0.00168
	(5.97)	(1.98)	(1.65)	(2.61)
A^*		0.0035	0.00284	0.00311
		(6.78)	(5.85)	(6.65)
$SAMESIZE^*$		0.01946	0.02155	0.02887
		(1.3)	(1.14)	(1.39)
$SAMEBM^*$		0.00086	-0.00013	-0.10995
		(2.02)	(-0.27)	(-2.22)
$SAMEMOM^*$		0.01212	0.01126	-0.01714
		(8.27)	(7.4)	(-0.89)
$NUMNACE^*$		0.00874	0.00745	0.0075
		(15.36)	(16.53)	(16.2)
$SIZE1^*$		0.02072	0.02305	0.03198
		(1.21)	(1.02)	(1.29)
$SIZE2^*$		0.02221	0.01728	0.00887
		(1.25)	(0.8)	(0.37)
$SIZE1^* \times SIZE2^*$		-0.01053	0.01716	-0.12181
		(-1.81)	(10.59)	(-2.1)
Other controls reported in the Internet Appendix				
$R2$	0.0787	0.13	0.14368	0.15407
No. Obs.	46,093	27,023	20,589	20,589
Nonlinear size controls	No	Yes	Yes	Yes
Pair characteristic controls	No	No	Yes	Yes
Nonlinear style controls	No	No	No	Yes

Table 2 reports the Fama and MacBeth (1973) cross-sectional regression of the monthly realized correlation of abnormal returns on the number of common short sellers and several stock pair control variables. The dependent variable is the realized correlation of a stock pair 4-factor Fama and French (1993) and Carhart (1997) abnormal returns in month $t+1$. The independent variables are updated quarterly and include NSS^*, which is the number of common short sellers short selling the stock pair. A^* is the number of common analyst covering the stock pair. $SAMESIZE$, $SAMEBM$, 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. $NUMNACE$ is the number of consecutively equal digits in the NACE code for the stock pair. All independent variables (except the dummy variables) are rank-transformed and normalised (to have zero mean and unit standard deviation). Estimates for the remaining controls can be found in the Internet Appendix. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

In the second specification we also included several nonlinear size controls. We included the third-order polynomial of $SAMESIZE^*$ and the second-order polynomial of $SIZE1^*$, $SIZE2^*$, and their interaction. We show in Table 2 only the coefficients for the first-order terms of $SAMESIZE^*$. The remaining coefficient estimates can be found in the Internet Appendix.[6] After adding these additional controls, the size of the coefficient on NSS^* decreases, but remains significant at 10% significance level with a t-statistic of 1.98.

In the third specification, displayed in the third column of Table 2, we add additional controls for pair characteristics. The coefficient estimates for these variables are reported in the Internet Appendix. The terms capturing similarities in past correlation, past profits, and past abnormal trading volume are all positive and significant. The coefficient on $DIFFGROW^*$ is positive and significant, meaning that stocks that have similar sales growth rates have higher correlation of excess returns. The coefficients on $DIFFLEV^*$ and $DIFFPRICE^*$ are also positive but insignificant, except for the fourth specification in which the coefficients on $DIFFLEV^*$ is significant. The coefficients on the dummy variables $DINDEX$ and $DCITY$ are both insignificant. With these additional controls the coefficient of NSS^* stands slightly above at the 10% significance level (t-statistic of 1.65, p-value of 10.4%).

In the fourth column of Table 2, we added the third-order polynomials of $SAMEBM^*$ and $SAMEMOM^*$, as well as the second-order polynomials of book-to-market and momentum for each stock in the pair, and their interaction. Again, coefficient estimates for these controls are given in the Internet Appendix. In this specification, which is the most complete and flexible, the coefficient on NSS^* gains in significance (t-statistic of 2.61 and p-value of 1.2%). The coefficient equals 0.00168, which underlines that an increase in one standard deviation in the number of common short sellers is associated with an increase of the forecasted correlation of excess returns of about 2.13% of the average abnormal correlation.

In untabulated results, we find that fitted values that are due to NSS^* range from an average minimum of 0.0705 to an average maximum of 0.0938, around an average abnormal correlation of 0.0789.[7]

Overall, despite the high variability of the correlation of excess returns, which have an average standard deviation of 0.25 and an average R^2 of 15% in the fourth specification, the association between NSS^* and the future correlation of excess returns is significant.

[6] Available at: https://www.dropbox.com/s/vmvyq56fke34wl8/appendix.pdf?dl=0.

[7] To calculate the range of these fitted values, we first orthogonalise NSS^* with respect to all the controls used in the fourth specification. We then forecast the realized correlation of 4-factor excess returns using the orthogonalised NSS^* and save the minimum and maximum forecasted value for each cross-section. Finally, we average these values across time.

Figure 1: One-step ahead forecast Mean Squared Error for pairwise realised excess correlation using orthogonalised NSS^*

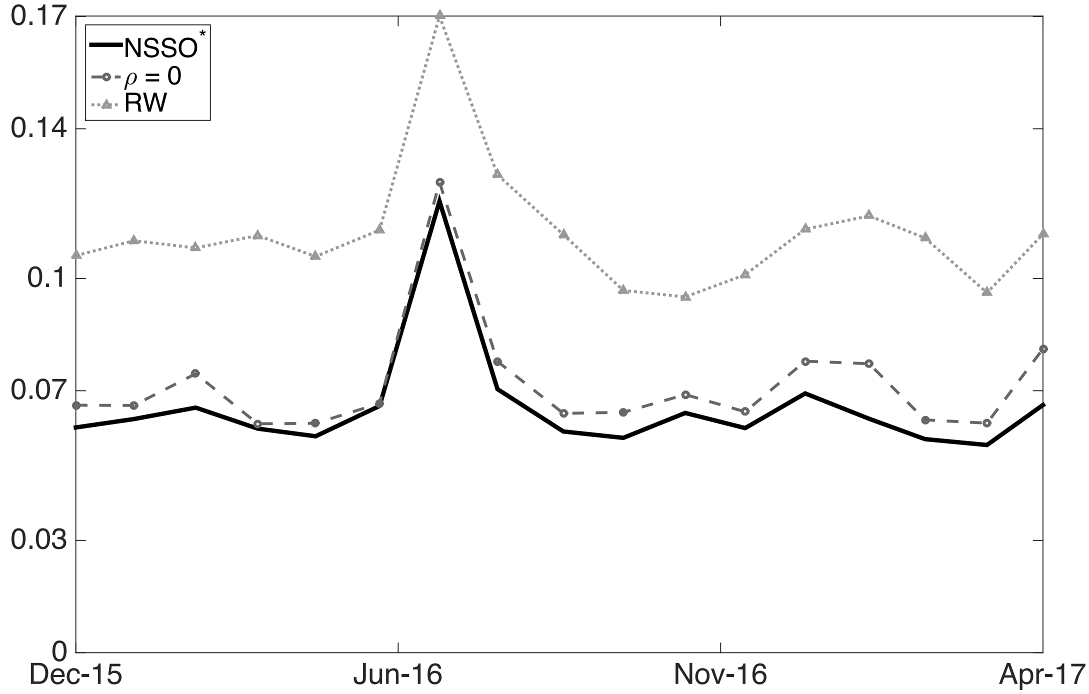


Figure 1 shows the one-month ahead mean squared forecast error of realised excess correlation using NSS^* orthogonalised with respect to the controls of specification 4 inTable 3.2. We used Fama and MacBeth (1973) regressions to estimate the regression coefficients.

B Out-of-sample forecasting performance

To understand if the in-sample forecast gains coming from NSS^* hold out of sample, we perform a simple out-of-sample forecasting exercise using the fourth specification given in Table 2. For this we initially used data up to the end of 2015, and then forecasted the one-month ahead excess correlation using an expanding window and re-estimating the coefficients at each step.

To distinguish the forecast gains coming from NSS^* and those coming from the other controls in specification 4, we orthogonalised NSS^* with respect to all the controls, and then used the residual value to forecast one-month ahead realised correlation. Figure 1 shows the results in terms of the mean squared forecast error.

Figure 1 shows that the orthogonalised variable, which we call $NSSO^*$, performs superiorly to the random walk forecasting rule. The average mean squared forecast error for the model with $NSSO^*$ is 0.0897, whereas it is 0.1376 using the random walk to forecast next month correlation. The figure also shows a large jump in the forecast mean square error at the end of June 2016, due to the EU membership referendum (Brexit) held in the UK.

An alternative to the random walk forecasting rule could be to agnostically set the forecast for future excess correlation to zero. Such a simple rule would yield a mean squared forecast error of 0.0976. The forecast gains of using $NSSO^*$ with respect to following the agnostic rule may seem slim. However, as previously mentioned, the correlation of excess returns is highly volatile, which makes forecasting the correlation of excess returns a difficult task. If we attempt to forecast simple return correlation, rather than 4-factor excess return correlation, the gains of the model are larger, with a forecast mean squared error of 0.1063 for $NSSO^*$ against a forecast mean squared error of 0.1162 for the agnostic forecast rule.

These results confirm that the number of common short sellers has useful information to forecast stock return correlation, beyond the information contained in the controls used in our model.

C Trading strategy

We evaluate the returns from a trading strategy exploiting the signal of common short selling. We followed the strategy proposed by Antón and Polk (2014), who develop a strategy exploiting common mutual fund holdings. The strategy works as follows: If a high (low) performing stock is connected, through one or more common short sellers, to other high (low) performing stocks, then we sell (buy) that stock. The main presumption behind the strategy is that short sellers can predict future price reversals, so that stocks targeted by common short sellers should earn negative abnormal returns, whereas stocks avoided by common short sellers will earn positive abnormal returns (Boehmer et al., 2010).

Every month, we ordered stocks based on the performance of their past three-month cumulative returns and sorted them into quintiles. Independently, we also ordered stocks based on the past three-month cumulative returns of their connected portfolio, which we constructed following Antón and Polk (2014). Specifically, we built the connected portfolio using the orthogonalised variable, $NSSO^*$, and NSS^{**}, where

$$NSS^{**}_{ij,t} = rank(NSSO^*_{ij,t}) \qquad \text{if} \qquad NSS_{ij,t} > 0$$
$$NSS^{**}_{ij,t} = 0 \qquad \text{if} \qquad NSS_{ij,t} = 0.$$

The return on the connected portfolio of stock i is computed as the weighted sum of all the connected stocks of i through one or more common short sellers, where the weights depend on the number of common short sellers as follows:

$$r_{iC,t} = \frac{\sum_{j=1}^{J} NSS^{**}_{ij,t-1} r_{j,t}}{\sum_{j=1}^{J} NSS^{**}_{ij,t-1}}.$$

Next, we computed the cumulative abnormal returns for an equally-weighted portfolio of stocks that were in the lowest quintile of own-returns and lowest quintile of connected returns. We called this the *low* portfolio. Similarly, we computed the cumulative abnormal returns for the equally-weighted portfolio of stocks that were in the high own-return and high connected return quintile, and called this the *high* portfolio. We computed the abnormal returns by regressing the time-series of returns from these portfolios on the four factors of Fama and French (1993) and Carhart (1997), and the short-term reversal factor.[8]

Figure 2 shows the cumulative abnormal returns over 12 months of the *low* portfolio, the *high* portfolio, and a portfolio that is short *high* and long *low*. The figure shows that the strategy can earn cumulative abnormal returns of over 9% after 12 months. These cumulative returns are gross of transaction costs, including the cost of short selling, which can be particularly high.

Notice that the cumulative returns of the *low* portfolio are negative and decreasing during the first two months of the strategy. These returns could be curbed by short-run momentum-based traders and could be evidence of possible predatory behaviour of short sellers. Successively, the *low* portfolio gains increasing positive returns, making positive cumulative returns after five months. Quite to the contrary, the high portfolio earns negative cumulative abnormal returns of almost 5% after 10 months.

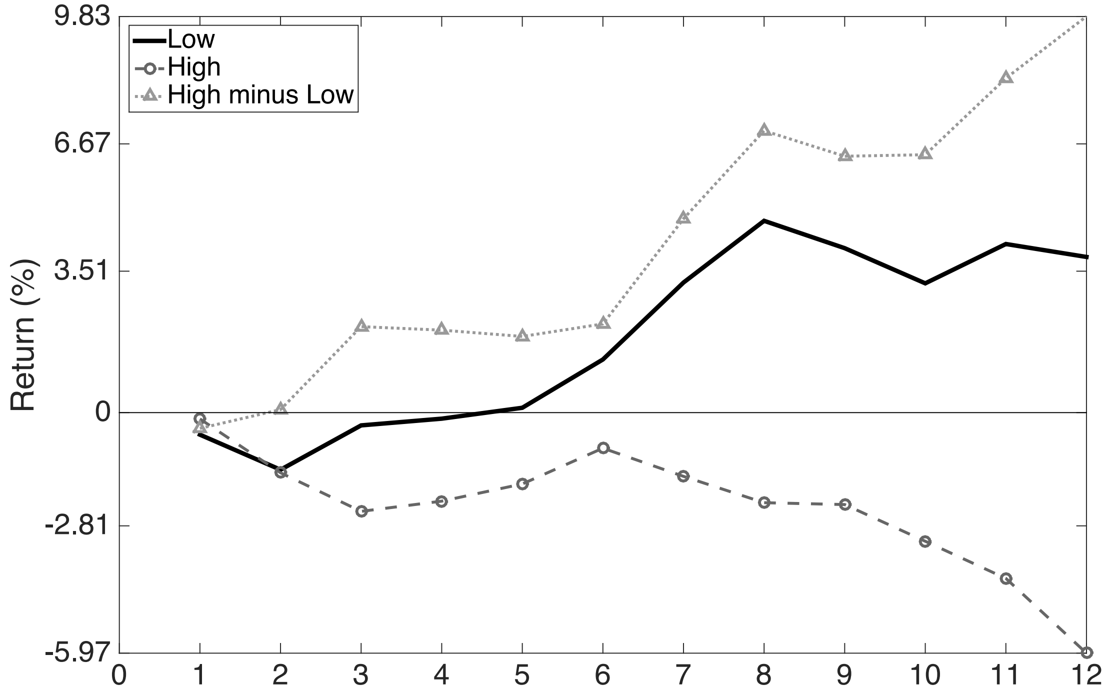
The results of this section show that common short selling information can be used to construct trading strategies that can earn positive cumulative abnormal returns for several months. Stocks that have had negative returns in the past and which are connected to stocks that also have had negative past returns, will revert to having positive returns. Similarly, stocks which have had high returns in the past and are connected to stocks which also had high past returns, are expected to revert to having negative returns in the future. The trading strategy exploits this observation to gain a positive margin which continues to hold for the following 24 months.

V An explanation

In this section, we discuss three possible mechanisms that might explain the uncovered positive relationship between common short selling and future excess correlation. The first two, which we identify as the negative price pressure mechanism and liquidity provision mechanism, are covered in Subsection A. The third mechanism, which we call the informative short selling mechanism, is covered in Subsection B.

[8]We collected the short-term reversal factor also from Ken French's website.

Figure 2: Cumulative excess returns for a trading strategy exploiting common short selling



The figure shows the cumulative excess returns of a trading strategy exploiting common short selling and the results of Table 2. We sorted stocks based on their own three-month return portfolio and on their three-month connected portfolio return. We defined a stock's connected return as the return on the portfolio of stocks that are short sold in common with that stock for the quarter in question. The connected portfolio weights are proportional to the number of common short sellers. The figure shows the cumulative abnormal returns on the stocks that are in the lowest quintile in the own-return and connected-return portfolio (which we called the *low* portfolio), stocks that are in the highest quintile of the own-return and connected-return portfolio (which we called the *high* portfolio), and the difference between the *low* and *high* abnormal returns. Abnormal returns are calculated using the Fama and French (1993) and Carhart (1997) four-factor model augmented with the short-term reversal factor.

A Negative price pressure and liquidity provision

Brunnermeier and Pedersen (2005) suggest that predatory trading can cause market contagion, in the sense of Forbes and Rigobon (2002). Furthermore, Cont and Wagalath (2013) show that a common short seller can induce high correlation by applying negative price pressure across several stocks. In both theoretical models, the price impact of short selling inversely depends on market depth. Hence, the more illiquid the underlying assets, the greater should be the impact of common short sellers on the correlation of the underlying stocks.

A second possible mechanism that could explain our result is the liquidity provision mechanism. Short sellers have been found to act as voluntary liquidity providers (Boehmer and Wu, 2013). Accordingly, short sellers help provide liquidity for illiquid stocks, by selling in periods of high buying pressure. Diether et al. (2009) argue that subsequent negative returns might be seen as compensation for short sellers voluntarily providing liquidity to the market. If this mechanism were in place, we should be able to observe a higher correlation for those short-sold stocks which are low on liquidity.

We can test these two mechanisms empirically, as both mechanisms predict that the association between the number of common short sellers and correlation is stronger for illiquid assets. We collected free float market capitalisation data for the stocks in our sample from Thompson Reuters Eikon and constructed a new variable, $FLOAT$, which captures the total free float market capitalisation (in £ sterling) of stock pairs. As with the other variables in our regression analysis, we worked with the cross-sectional normalised ranking of $FLOAT$, denoted $FLOAT^*$. We augment our regression analysis of Table 2 with a term capturing the interaction between NSS^* and our measures of pair liquidity, $FLOAT^*$.[9] A negative coefficient on the interaction term would provide evidence in support of either the price pressure mechanism or the liquidity provision mechanism (or both).

To further verify our results, we repeated the same analysis using the illiquidity dummy variable, $DFLOAT$, instead of $FLOAT^*$. We let $DFLOAT$ capture highly illiquid stock pairs. We set $DFLOAT_{ij,t}$ equal to one if, in the quarter preceding period t, both i and j are in the lowest cross-sectional quintile in terms of free float market capitalisation. Otherwise, $DFLOAT_{ij,t}$ is equal to zero. In this case, a positive coefficient on the interaction term would provide evidence in favour of one of the two mechanisms (or both) taking place.[10]

Table 3 shows the results of the additional regressions. Panel A of Table 3 shows the results for the regression using $FLOAT^*$. Except for the second specification, the regression coefficient on the interaction term, $NSS^* \times FLOAT^*$, is positive as would be expected if the negative price pressure mechanism or liquidity provision mechanism were in place.

[9] For completeness, we also included in each regression analysis the measure of pair liquidity on its own.

[10] Again, we also included in each regression analysis the pair illiquidity dummy variable on its own.

However, across all specifications except for the first, we cannot confirm that the coefficient is statistically different from zero. Across all specifications except for the third, the coefficient on NSS^* continues to remain significant at the 10% level or better. All remaining regression coefficients are given in the Internet Appendix.

Table 3: Common short selling and excess correlation: The effect liquidity

Panel A: Interaction with liquidity measure ($FLOAT^*$)

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.07063	0.08466	0.08711	0.08343
	(15.65)	(7.18)	(6.95)	(5.15)
NSS^*	0.0054	0.00147	0.00106	0.00142
	(6.06)	(1.94)	(1.41)	(2.41)
$NSS^* \times FLOAT^*$	0.00121	-0.00038	0.00054	0.00084
	(2.78)	(-0.65)	(0.87)	(1.36)

Panel B: Interaction with illiquidity dummy ($DFLOAT$)

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.07118	0.08426	0.083	0.07329
	(15.74)	(6.98)	(6.64)	(4.58)
NSS^*	0.00548	0.00155	0.00126	0.00164
	(5.9)	(1.92)	(1.58)	(2.54)
$NSS^* \times DFLOAT$	-0.0031	0.00306	0.00327	0.0023
	(-1.54)	(1.53)	(1.46)	(0.94)

Other controls reported in the Internet Appendix

The table reports Fama and MacBeth (1973) cross-sectional regressions of monthly abnormal realized correlation on the number of common short sellers and several stock pair control variables. Panel A shows the result of adding $FLOAT^*$ and the interaction of $FLOAT^*$ and NSS^* to the regressions carried out in Panel A of Table 2. $FLOAT^*$ captures the total free float capital of a stock pair. As with the other independent variables, $FLOAT^*$ has been cross-sectionally ranked-normalised. Panel B shows the result of using $DFLOAT$ instead of $FLOAT^*$ in our regression analysis. $DFLOAT$ is a dummy variable that is equal to one if both stocks are in the lower cross-sectional quintile in terms free float capital. Estimates for the remaining control variables can be found in the Internet Appendix. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

Panel B of Table 3 shows the same analysis as Panel A, except that we used the illiquidity dummy, $DFLOAT$, instead of $FLOAT^*$. Results show that, for all four specifications considered, the coefficient on the interaction term, $NSS^* \times DFLOAT^*$, is insignificant at the 10% significance level. Given these results, we were only able to find weak evidence in support of the hypothesis implied by the price impact mechanism and the liquidity-provision mechanism.

Two considerations are in order, one per mechanism. Regarding the price impact mechanism, it might be that this is visible only at very high frequencies—e.g., at the daily or even intra-daily frequency. In that case, the lag structure in our model is unlikely to capture it and the explanation for the observed relationship between common short selling and future correlation must occur through some other mechanism.

Our results regarding the liquidity provision mechanism seem to corroborate those of Diether et al. (2009). They studied short selling for NYSE and NASDAQ stocks and found that the liquidity provision mechanism could not explain the predictive relationship between short selling and future price declines for individual stocks. In that case, the liquidity provision mechanism would be unable to explain correlation in stock pairs. In the next section, we propose that the uncovered relationship might be due to informative short selling.

B Informative short selling

There is considerable evidence in the literature that short sellers are informed traders. Boehmer, Jones, and Zhang (2008) and Diether et al. (2009), for example, show that short sellers can correctly predict future returns. If short sellers trade according to information that stock prices should decline in the future, this will result in higher future correlation across the shorted stocks. Thus, the association between common short selling and future correlation might be the result of informed trading by short sellers. In this section, we verify this mechanism by making use of several indicators from financial statement analysis.

Dechow et al. (2001) show that short sellers target stocks that are overpriced according to several fundamental-to-price ratios. Moreover, Curtis and Fargher (2014) show that short sellers target the stock of companies with high accruals and high growth rates. Following these studies, we collected and constructed three fundamental-to-price measures. Additionally, we constructed a measure based on accruals, and a measure based on asset growth. We then use these measures to develop an indicator of stock value and recover the common short selling that is due to informative trading.

In terms of fundamental-to-price measures, we used book-to-market, earnings yield, and value-to-price. The lower these fundamental-to-price measures, the more likely a stock is overpriced (Basu, 1983, Fama and French, 1992, Dechow et al., 2001).

We obtained quarterly book-to-market values for the last fiscal year from Thompson Reuters Eikon. We computed earning yield ratio as the reciprocal of the P/E ratio, which we also obtained from Thompson Reuters Eikon. Finally, we computed the value-to-price ratio using the residual income model of Ohlson (1995).[11]

Higher accruals are associated with lower future returns (Sloan, 1996), and therefore also provide an indicator for overpriced stocks. Following the cash-flow approach of Hribar and Collins (2002), we defined accruals as the difference between net income before extraordinary items and cash flows from operating activities, all scaled by total current assets. We obtained all three variables from Thompson Reuters Eikon.

High growth stocks are associated with lower future returns (Cooper, Gulen, and Schill, 2008). Hence, we expect informed short sellers, who are following value-based trading strategies, to focus on stock pairs with high asset growth. We use the year-on-year growth rate of total assets as an indicator of overpriced stocks.

For each stock-period observation, we collected these five indicators in the vector, $y_{it} = [BM_{it}, EY_{it}, V_{it}, ACCR_{it}, AG_{it}]'$, where BM_{it}, EY_{it}, V_{it}, $ACCR_{it}$, AG_{it} represent, respectively, book-to-market ratio, earnings yield, value-to-price, accruals, and asset growth for stock i at time t. We then estimated the following static factor model,

$$y_{it} = \lambda F_{it} + e_{it} \qquad e_{it} \sim \mathcal{N}(0, \Psi), \quad \forall t = 3 \times t_q, \tag{2}$$

where F_{it} represents the common factor driving the data for the stock-time observation (i, t), λ represents the loading on the common factor, and $t_q = 1, \ldots, T_q$ represents quarters in our sample. We estimated the model by standard maximum likelihood techniques pooling panel observations. Table 4 shows the estimates of the loadings.

Table 4: Estimated loading for the factor model on financial statement variables

	BM	EY	V	$ACCR$	AG
$\hat{\lambda}$	0.9966	0.8694	0.9801	0.0087	0.0108

The table reports the factor loadings estimated for Equation 2. BM is the book-to-market ration, EY is the earning yields, V is the value-to-price ratio computed according to Ohlson (1995), $ACCR$ is accruals, and AG is asset growth.

The estimates show that the factor loads positively and extensively on BM, EY, and V, and to a lesser extent on $ACCR$ and AG. This means that the common underlying factor mostly represents fundamental-to-price ratios. Since BM, EY, and V are positively associated to future returns and the loadings on these variables are positive, then the factor

[11]Precisely, we followed the one-period income model:

$$Vf(1)_t = b_t + \max\left(\frac{f(1)_t - rb_t}{r - g}, 0\right)$$

Following Curtis and Fargher (2014), we set r equal to the monthly rate on the 10-year gilt plus an equity premium of 6% and g equal to 3%. $f(1)$ is the average 1-year forecast earnings (we used the average of all broker estimates of earnings per share obtained from the I/B/E/S database) and b_t is the book value.

should also be positively associated with future returns. Hence, we interpret the common factor as an indicator of the value of a given stock i at a given period t. The lower the realization of the factor for a given stock, the more likely it is that the stock is overpriced relative to value. Hence, we expect informed short sellers to target these stocks, as they would expect their price to decrease in the future.

Next, using our estimated model, we recover the common factor score, $\hat{F}_{i,t}$, which represents the estimated value of the underlying common factor for stock i at time t. We use the factor score to recover the portion of $NSS^*_{ij,t}$ which is due to informative trading, by running the following regression,

$$NSS^*_{ij,t} = a + b_1 \widehat{F1}^*_{ij,t} + b_2 \widehat{F2}^*_{ij,t} + \epsilon_{ij,t}, \qquad \forall t = 3 \times t_q \tag{3}$$

where $\widehat{F1}^*$ represents the percentile rank of the larger factor score between i and j. Similarly, $\widehat{F2}^*$ represents the percentile rank of smaller factor score between i and j. Both $\widehat{F1}^*$ and $\widehat{F2}^*$ have been normalized and rank-transformed for each cross-section, to be consistent with our analysis in previous sections.

We run the regression in Equation 3 cross-sectionally and use the cross-sectional regression coefficient estimates to recover the part of common $NSS_{ij,t}$ that is due to $\widehat{F1}^*_{ij,t}$ and $\widehat{F2}^*_{ij,t}$. We call this fitted variable $\widehat{INSS}_{ij,t}$ and we assume that it captures the number of common short sellers that trade informatively according to stock value. This might be a simplistic assumption. However, it is useful to achieve our goal of separating informed short selling from NSS^*.

We rank-transformed and standardised the projected variable, $\widehat{INSS}$, and denoted it $\widehat{INSS}^*$. We then used $\widehat{INSS}^*$ in place of NSS^* and repeat the analysis of Table 2. If the effect of informative short selling is not important for forecasting future correlation, then we should find insignificant coefficients attached to $\widehat{INSS}^*$. We show the results in Table 5. Again, because of space concerns, we present the coefficient estimates on the constant and main covariate only. The remaining estimates are given in the Internet Appendix.

Table 5 shows that the part of common short selling that is due to informed trading is highly predictive of future excess correlation. Across all four specifications, the coefficient on $\widehat{INSS}^*$ is significant at the 10% significance level or better. Notice that the size of the coefficient has also increased. In the fourth specification, for example, a standard deviation increase in informed common short selling is now associated with a increase 3% of the average excess correlation. This effect is substantially larger than that found using simple NSS^*. Fitted values range from an average minimum of 0.0583 to an average maximum of 0.0926, around an average abnormal correlation of 0.0754.[12]

[12]To calculate the range of fitted values, we first orthogonalise $\widehat{INSS}^*$ with respect to all the controls. We

Table 5: Informed common short selling and excess correlation

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.08989	0.11818	0.10938	0.07542
	(17.58)	(7)	(6.96)	(3.24)
$\widehat{INSS}^*$	0.00789	0.00248	0.00323	0.0023
	(6.39)	(1.95)	(3.02)	(1.77)
Other controls reported in the Internet Appendix				
$R2$	0.12252	0.16954	0.18213	0.19521
No. Obs.	18348	13336	12465	12465

The table shows the Fama and MacBeth (1973) cross-sectional regressions of the monthly realized correlation of abnormal returns on informed common short selling and several stock pair control variables. The dependent variable is the realized correlation of a stock pair 4-factor Fama and French (1993) and Carhart (1997) abnormal returns in month $t+1$. The independent variables are updated quarterly and include $\widehat{INSS}^*$, which represents the part of common short selling (NSS^*) that is due to informed trading. $\widehat{INSS}^*$ is constructed from the fitted values from the regression of NSS^* on $\widehat{F1}^*_{ij,t}$ and $\widehat{F2}^*_{ij,t}$, which are, respectively, the larger and smaller normalised ranked-tranforms of the factor scores from the model given in Equation 2. Estimates for the remaining controls can be found in the Internet Appendix. t-statistics (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

Overall, our results show that the part of NSS^* that can be related to informative short selling is a significant predictor of the future correlation of excess returns. This evidence supports the hypothesis that the association between NSS^* and correlation is due to informative trading of short sellers, who attempt to predict future returns using financial statement analyses data.

VI The determinants of common short selling

We verify whether informed trading is in fact taking place in our sample, controlling for several determinants of common short selling, including transaction costs for short sellers.

We follow Curtis and Fargher (2014) and attempt to distinguish informed, value-based, short selling from uninformed, momentum-based, short selling in NSS^*. Momentum-based trading relies on the idea that prices will follow their most recent trend. That is, if two stocks have been undergoing losses in the past, momentum-based short sellers will short both stocks and attempt to ride the negative price trend. On the other hand, value-based short selling relies on the idea that if prices are far from their fundamental value, they

then forecast the realized correlation of 4-factor excess returns using the orthogonalised $\widehat{INSS}^*$ and save the minimum and maximum forecast for each cross-section. We then average these values across cross-sections.

should ultimately revert. Both strategies are based on the idea that future prices will go down, making short selling a profitable trade. However, the motivations underlying the two strategies are intrinsically different.

In order to measure value-based short selling we use the indicators described in the previous section i.e., book-to-market, earnings yield, value-to-price, accruals, and asset growth. To measure momentum-based short selling, we identify those stocks that have had negative buy-and-hold returns over the last 12 months. We define the dummy variable $DECLINE_{ij,t}$ as equal to one if both stocks i and j have had a negative buy-and-hold returns in the 12-month period preceding t.

We defined the dummy variable $OVERPRICED_{ij,t}$ as equal to one if, according to our value indicators, stocks i and j are both overpriced in period t. In particular, for the fundamental-to-price indicators (book-to-market, earning yields, and value-to-price), $OVERPRICED_{ij,t}$ is equal to one if, at time t, for both i and j, the indicator in question is jointly in the lower quintile of its cross-sectional sample. On the other hand, for accruals and asset growth, we set $OVERPRICED_{ij,t}$ equal to one if, for both i and j, the indicator is jointly in the highest quintile of its cross-sectional sample.

We then ran the following regression,

$$\begin{aligned} NSS^{*}_{ij,t} = \quad & a + b_1 \times OVERPRICED_{ij,t} + b_2 \times DECLINE_{ij,t} \\ & + \sum_{k}^{K} b_k \times CONTROLS_{ij,k} + \epsilon_{ij,t}, \quad \forall t = 3 \times t_q. \end{aligned} \tag{4}$$

Notice that NSS^{*} and all variables on the right-hand-side of Equation 4 are updated quarterly. To avoid serial correlation, we ran Equation 4 at the quarterly frequency, t_q.

We expect b_2 to be positive if common short sellers concentrate on stocks that have had declining returns during the past 12 months. On the other hand, the sign of b_1 will depend on common short sellers motivation and on the indicator. If short sellers are trading according to fundamentals, we expect b_1 to be positive, such that stock pairs with lower fundamentals are associated with a higher number of common short sellers.

We controlled for three of the determinants of the number of common short sellers, related to the cost of short selling.

First, the availability of stocks to borrow is an important determinant for the cost of short selling. Geczy, Musto, and Reed (2002) suggest that this should be related to size. Thus, we included $SIZE1_{ij,t}$ and $SIZE2_{ij,t}$, defined in Section III.B.

Second, institutional owners are among the main lenders of stocks that are then short sold (Faulkner, 2007). For this reason, we controlled for the percentage of shares of the two stocks that are held by strategic investors, which include corporations, holding companies, individuals and government agencies. Precisely, we defined $INST1_{ij,t}$ and $INST2_{ij,t}$ as, respectively, the larger and smaller cross-sectional percentile rank of institutional ownership of the pair.

Finally, since short sellers are also required to repay dividends to the original owner during a short sale, we included dividend earnings, $DIV1_{ij,t}$ and $DIV2_{ij,t}$, as controls in our regression. $DIV1_{ij,t}$ denotes the larger dividend earnings cross-sectional percentile rank of the pair, whereas $DIV2_{ij,t}$ denotes the smaller dividend earnings cross-sectional percentile rank of the pair.

All controls are rank-transformed and normalised. We estimate Equation 4 using Fama and MacBeth (1973) cross-sectional regressions and averaging the time-series of the coefficients. We notice that autocorrelation in the residuals is limited to one lag. Hence, we use Newey and West (1987) robust standard errors to control for autocorrelation up to one quarter.

Panel A of Table 6 displays the result of the cross-sectional regression analysis. For book-to-market and value-to-price, the coefficient was positive and significant, meaning that overpriced pairs were associated with higher presence of common short sellers. For earnings yield, the coefficient was also positive but insignificant, whereas it was significantly negative for accruals. Overall, there appears to be evidence that overpriced stock pairs are targeted by short sellers, at least according to fundamental-to-price ratios such as book-to-market.

In terms of price declines, stock pairs which had negative buy-and-hold returns over the past year were associated with a higher number of common short sellers. This means that common short sellers in our sample were targeting stock pairs with declining returns. One possibility, put forward by Curtis and Fargher (2014), is that short sellers target overpriced stocks with declining prices because it is excessively costly to short overpriced stocks that are rallying. We verified this possibility below.

Precisely, we analysed the subsample of stock pairs that have suffered price declines (i.e., for which $DECLINE_{ij,t} = 1$). We regressed NSS^* on the various indicators of $OVERPRICED$. We also included $UNDERPRICED_{ij,t}$, a dummy variable for stock pairs that are underpriced according to our five indicators. For book-to-market, earnings yield, and our measure of value-to-price, $UNDERPRICED_{ij,t}$ is equal to one if both stocks i and j are in the upper quintile of the cross-sectional sample of the indicators. Contrarily, for accruals and asset growth, $UNDERPRICED_{ij,t}$ is equal to one if both stocks i and j are in their lower cross-sectional quintile of these measures.

Table 6: The determinants of common short selling

Panel A: Full sample

Indicator	Dependent Variable: NSS^*				
	Book-to-market	Earnings yield	Value-to-price	Accruals	Asset growth
Constant	-0.00835	-0.00634	-0.0088	-0.0047	-0.00556
	(-0.52)	(-0.39)	(-0.55)	(-0.3)	(-0.34)
$DECLINE$	0.25044	0.25055	0.2511	0.25001	0.25006
	(10.87)	(10.67)	(10.77)	(10.72)	(10.68)
$OVERPRICED$	0.08451	0.02075	0.0741	-0.036	-0.01283
	(2.97)	(0.97)	(4.33)	(-2.67)	(-0.48)
Other controls reported in the Internet Appendix					

Panel B: Subsample for declining stock pairs ($DECLINE_{ij,t} = 1$)

Indicator	Dependent Variable: NSS^*				
	Book-to-market	Earnings yield	Value-to-price	Accruals	Asset growth
Constant	0.27492	0.28455	0.28627	0.27954	0.2911
	(9.12)	(10.11)	(9.32)	(10.12)	(9.54)
$OVERPRICED$	0.117	0.03302	0.07848	0.13772	0.03726
	(2.56)	(0.44)	(1.3)	(1.93)	(0.46)
$UNDERPRICED$	0.11715	-0.05954	-0.043	0.08159	-0.11086
	(2.17)	(-0.82)	(-0.87)	(2.66)	(-4.09)
Other controls reported in the Internet Appendix					

This table reports the results of the Fama and MacBeth (1973) cross-sectional regressions investigating the determinants of the number of common short sellers for a stock pair, $NSS^*_{ij,t}$. Panel A reports results for the full sample. The independent variables include $OVERPRICED_{ij,t}$, a dummy variable equal to one if the stock pair is overpriced according to our value indicators (and zero otherwise), and $DECLINE_{ij,t}$, a dummy variable equal to one if the stock pair returns have had negative buy-and-hold returns over the last 12 months (and zero otherwise). Panel B reports results for the subsample of declining stock pairs (i.e, for which $DECLINE_{ij,t} = 1$). In addition to $OVERPRICED_{ij,t}$, the independent variables include $UNDERPRICED_{ij,t}$, a dummy variable equal to one if the stock pair is overpriced according to our value indicators (and zero otherwise). All variables for regressions in Panel A and B are updated quarterly and therefore regressions are carried out at the quarterly frequency, for $t = 3 \times t_q$, $t_q = 1, \ldots, T_q$. For both Panel A and Panel B, only the coefficient on the constant and the main indicators of interest are shown. The remaining coefficients are reported in the Internet Appendix. t-values (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to one quarter.

The regression model for the subsample of stock pairs with price declines is:

$$\begin{aligned} NSS^*_{ij,t} = \quad & a + b_o \times OVERPRICED_{ij,t} + b_u \times UNDERPRICED_{ij,t} \\ & + \sum_k^K b_k \times CONTROLS_{ij,k} + \epsilon_{ij,t}, \quad \forall t = 3 \times t_q. \end{aligned} \tag{5}$$

Again, we included controls for dividend yields, size, and institutional ownership, and, as done previously, we ran the Fama and MacBeth (1973) regressions cross-sectionally. For inference, we used Newey and West (1987) robust standard errors, accounting for autocorrelation up to one quarter.

If common short sellers trade according to value, we would expect the coefficient on $OVERPRICED$ to be positive and significant. By contrast, if common short sellers adopt momentum-based strategies, we expect them to concentrate on stocks that are underpriced relative to value. This would be consistent with a positive and significant coefficient on $UNDERPRICED$.

Panel B of Table 6 shows the results for this subsample regression. We found that across all specifications the coefficient on $OVERPRICED$ is positive. It was significant for our indicator based on book-to-market and accruals. Hence, there is evidence that short sellers target stock pairs with declining prices but which are overpriced according to these indicators.

The coefficients for our indicator of underpriced stock pairs gave mixed results according to the measures under consideration. The coefficient was negative and significant for asset growth, which indicates that there is evidence of common short sellers avoiding stock pairs with declining prices that are underpriced. On the other hand, for book-to-market and accruals, we found that the coefficient on $UNDERPRICED$ is positive and significant, showing evidence of momentum-based short selling.

Overall, our analysis showed that both informed short selling and momentum-based short selling is occurring for the stocks in our data. However, it might be that our measure of common short selling, NSS, prevalently captures informed short selling because it is built using large disclosed short positions. Easley and O'Hara (1987) and Avramov, Chordia, and Goyal (2006) classify large trades as informed. If this were also true for short sales, then this would further reinforce the results found in Section V.B. Our results in this section seem to point in this direction, even if we cannot exclude the possibility that some momentum trading is taking place.

VII Conclusion

We have connected stocks according to common short sellers, and have studied the relationship between the number of common short sellers and realized excess return correlation. We have found that the number of common short sellers is positively associated with higher future correlation of excess returns.

Having excluded two possible mechanisms—the price impact mechanism and the liquidity provision mechanism—as explanations for the association we have uncovered, we verified whether the association could be explained by informative short selling. We have found that informed common short selling—i.e., the number of common short sellers that can be linked to informative trading—is strongly associated to future excess correlation. We also verified whether informed short selling is occurring among the common short sellers in our sample, whilst controlling for several factors capturing short selling costs. We have found evidence of both informed and uninformed (momentum-based) short selling.

Our study has several policy implications. First, for investors and risk managers, our study offers a new way to interpret the publicly available data on short selling disclosures. We show that this data can be used to forecast future correlation and potentially to make portfolio investment decisions.

Second, for regulators, our study uncovers a relationship between common short selling and asset correlation that should be taken into consideration for financial stability policy. Our results suggest that this relationship is driven by informative short selling, thus confirming the sophistication of short sellers and their proven importance for market efficiency and price informativeness (Boehmer and Wu, 2013). On the other hand, our results do not allow us to dismiss the possibility that also non-informative momentum-based short selling is occurring in our sample. The good news is that we did not find evidence of a potentially detrimental price-impact effect of common short selling for illiquid stock, which is the sort of predatory effect that regulators often fear.

References

Antón, M., and C. Polk. "Connected stocks." *The Journal of Finance*, 69 (2014), 1099–1127.

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

Basu, S. "The relationship between earning yields, market value, and return for nyse common stocks: further evidence." *Journal of Financial Economics*, 12 (1983), 129–156.

Billio, M., and L. Pelizzon. "Contagion and Interdependence in Stock Markets: Have They Been Misdiagnosed?" *Journal of Economics and Business*, 55 (2003), 405–426.

Boehmer, E.; Z. R. Huszar; and B. D. Jordan. "The Good News in Short Selling." *Journal of Financial Economics*, 96 (2010), 80–97.

Boehmer, E.; C. M. Jones; and X. Zhang. "Which Shorts Are Informed?" *Journal of Finance*, 63 (2008), 491–527.

Boehmer, E., and J. J. Wu. "Short Selling and the Price Discovery Process." *Review of Financial Studies*, 26 (2013), 287–322.

Brunnermeier, M. K., and M. Oehmke. "Predatory Short Selling." *Review of Finance*, 18 (2014), 2153–2195.

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

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

Chen, H. J.; S. J. Chen; Z. Chen; and F. Li. “Empirical investigation of an equity pairs trading strategy.” *Management Science*.

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

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

Cooper, M. J.; H. Gulen; and M. J. Schill. “Asset growth and the cross-section of stock returns.” *Journal of Finance*, 63 (2008), 1609–1651.

Corsetti, G., and M. P. M. Sbracia. “Śome Contagion, Some Interdependence ́ More Pitfalls in Tests of Financial Contagion.” *Journal of International Money and Finance*, 24 (2005), 1177–1199.

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

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

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

Easley, D., and M. O’Hara. “Price, trade size, and information in securities markets.” *Journal of Financial Economics*, 19 (1987), 69–90.

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

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

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

Faulkner, M. C. “An Introduction to Securities Lending.” Tech. rep., Spitalfields Advisors (2007). Fourth Edition.

Forbes, K., and R. Rigobon. “No contagion, only interdependence: Measuring stock market comovements.” *Journal of Finance*, 62 (2002), 2223–2261.

Geczy, C. C.; D. K. Musto; and A. V. Reed. "Stocks Are Special Too: An Analysis of the Equity Lending Market." *Journal of Financial Economics*, 66 (2002), 241–269.

Hribar, P., and D. W. Collins. "Error in estimating accruals: Implications for empirical research." *Journal of Accounting Research*, 40 (2002), 105–134.

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

Lynch, A.; B. Nikolic; X. Yan; and H. Yu. "Aggregate Short Selling, Commonality, and Stock Market Returns." *Journal of Financial Markets*, 17 (2013), 199–229.

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

Ohlson, J. A. "Earnings, book values, and dividends in equity valuation." *Contemporary Accounting Research*, 11 (1995), 661–687.

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

Sloan, R. G. "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *Accounting Review*, 71 (1996), 289–315.

Stattman, D. "Book values and stock returns." *The Chicago MBA: A Journal of Selected Papers*, 4 (1980), 25–40.

Yang, S., and S. E. Satchell. "Endogenous correlation." *Macroeconomic Dynamics*, 11 (2007), 124–153.

Appendix for Short Selling and Excess Return Correlation

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

[†] *Cambridge-INET Institute, University of Cambridge*

[‡] *CeReFiM, Université de Namur*

[⋆] *Vlerick Business School and University of Ghent*

NOT FOR PUBLICATION

This version: December 9, 2017

Abstract

This appendix contains additional results of the paper.

Table A1: Common short selling originations and excess correlation.

This table reports complete results of the Fama and MacBeth (1973) cross-sectional regressions carried out in Table 2. Panel A reports all the coefficient estimates for the regression analysis given in Panel A, Table 2. Panel B and Panel C reports, respectively, the coefficient estimates for the fourth specification of Panel B and Panel C of Table 2. Results for the remaining specifications are available upon request. The dependent variable is the realized correlation of a stock pair 4-factor Fama and French (1993) and Carhart (1997) abnormal returns in month $t+1$. The independent variables are updated quarterly and include NSS^*, which is the past number of common short sellers with an open short position on the stock pair. The remaining controls are defined as in the text. All independent variables (except the dummy variables) are rank-transformed and normalised (to have zero mean and unit standard deviation). t-values (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

Panel A: Full sample (Jan. 2013 - Jun. 2017)

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
$Constant$	0.07001 (15.6)	0.08627 (7.2)	0.08747 (7.02)	0.07893 (4.92)
NSS^*	0.00544 (5.97)	0.0016 (1.98)	0.0013 (1.65)	0.00168 (2.61)
A^*		0.0035 (6.78)	0.00284 (5.85)	0.00311 (6.65)
$SAMESIZE^*$		0.01946 (1.3)	0.02155 (1.14)	0.02887 (1.39)
$SAMEBM^*$		0.00086 (2.02)	-0.00013 (-0.27)	-0.10995 (-2.22)
$SAMEMOM^*$		0.01212 (8.27)	0.01126 (7.4)	-0.01714 (-0.89)
$NUMNACE^*$		0.00874 (15.36)	0.00745 (16.53)	0.0075 (16.2)
$SIZE1^*$		0.02041 (1.19)	0.02304 (1.01)	0.03216 (1.3)
$SIZE2^*$		0.02247 (1.26)	0.01722 (0.8)	0.00863 (0.36)
$SIZE1^* \times SIZE2^*$		-0.01037 (-1.77)	0.01715 (10.6)	-0.12092 (-2.08)
$RETCORR^*$			0.00356 (10.6)	0.01563 (-2.08)
$ROECORR^*$			0.00343 (4.62)	0.00334 (2.27)
$VOLCORR^*$			-0.00212 (3.54)	0.00372 (-1.13)
$DIFFGROW^*$			-0.00064 (-2.08)	-0.00241 (1.68)
$DIFFLEV^*$			-0.00123 (-0.49)	-0.00234 (9.69)
$DIFFPRICE^*$			-0.01313 (-1.45)	-0.00122 (4.5)
$DCITY$			-0.00184 (-1.81)	-0.00153 (4.05)
$DINDEX$			-0.00449 (-1.35)	-0.00333 (-2.62)
$(SAMESIZE^*)^2$		-0.00977 (-1.58)	-0.01072 (-1.35)	-0.01364 (-1.56)
$(SAMESIZE^*)^3$		0.00155 (1.05)	0.00188 (1)	0.00245 (1.19)
$(SIZE1^*)^2 \times (SIZE2^*)^2$		-0.00076 (-1.06)	-0.00067 (-0.74)	-0.00077 (-0.8)
$(SIZE1^*)^2$		-0.00395 (-1.45)	-0.00388 (-1.55)	-0.00401 (-1.6)
$(SIZE2^*)^2$		0.00721 (3.24)	0.00791 (3.03)	0.00885 (3.42)

Continued on next page

. . . continued from previous page

Panel A: Full sample (2013 - 2015)				
	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
$(SIZE1^*)^2 \times SIZE2^*$		-0.00049 (-0.16)	-0.00007 (-0.02)	0.00154 (0.37)
$SIZE1^* \times (SIZE2^*)^2$		-0.00569 (-1.94)	-0.0075 (-1.9)	-0.00922 (-2.15)
$BM1$				-0.12092 (-2.08)
$BM2^*$				0.13241 (2.27)
$MOM1^*$				-0.02573 (-1.13)
$MOM2^*$				0.03506 (1.68)
$BM1^* \times BM2^*$				0.02633 (2.66)
$(SAMEBM^*)^2$				0.02629 (2.3)
$(SAMEBM^*)^3$				-0.00443 (-2.03)
$(BM1^*)^2 \times (BM2^*)^2$				0.00139 (1.16)
$(BM1^*)^2$				0.00377 (1.01)
$(BM2^*)^2$				0.01215 (2.87)
$(BM1^*)^2 \times BM2^*$				-0.01815 (-2.7)
$BM1^* \times (BM2^*)^2$				0.01491 (2.28)
$MOM1^* \times MOM2^*$				0.01707 (3.34)
$(SAMEMOM^*)^2$				0.01173 (1.73)
$(SAMEMOM^*)^3$				-0.00243 (-1.66)
$(MOM1^*)^2 \times (MOM2^*)^2$				0.00306 (3.46)
$(MOM1^*)^2$				-0.0078 (-4.45)
$(MOM2^*)^2$				-0.00537 (-2.57)
$(MOM1^*)^2 \times MOM2^*$				-0.00857 (-2.81)
$MOM1^* \times (MOM2^*)^2$				0.00593 (1.78)
$R2$	0.0787 (9.98)	0.13 (11.75)	0.14368 (12.41)	0.15407 (13)
No. Obs.	46,093 (73.56)	27,023 (225.82)	20,589 (86.76)	20,589 (86.76)

Table A2: Common short selling originations and excess correlation: Liquidity results.

The table reports the results for the fourth specification of Fama and MacBeth (1973) cross-sectional regressions carried out in Table 3, in which NSS is intereacted with liquidity indicators $FLOAT^*$ and $DFLOAT$. $FLOAT^*$ captures the total float of the stock pair, whereas $DFLOAT$ is a dummy variable that is equal to one if both stocks are in the lower cross-sectional quintile in terms free float capital. All independent variables, except for dummy variables, have been cross-sectionally rank-transformed and normalised. t-values (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month)

	Dependent Variable: Correlation of 4F Residuals							
	FLOAT				*DLOAT*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant	0.07063	0.08466	0.08711	0.08343	0.07118	0.08426	0.083	0.07329
	(15.65)	(7.18)	(6.95)	(5.15)	(15.74)	(6.98)	(6.64)	(4.58)
NSS^*	0.0054	0.00147	0.00106	0.00142	0.00548	0.00155	0.00126	0.00164
	(6.06)	(1.94)	(1.41)	(2.41)	(5.9)	(1.92)	(1.58)	(2.54)
$FLOAT^*$	0.01967	0.02853	0.0266	0.02342				
	(7.03)	(6.49)	(6.83)	(5.71)				
$DFLOAT^*$					-0.03031	0.00525	0.00892	0.01107
					(-5.06)	(1.84)	(2.6727)	(3.3782)
$NSS^* \times FLOAT^*$	0.00121	-0.00038	0.00054	0.00084				
	(2.78)	(-0.65)	(0.87)	(1.36)				
$NSS^* \times DFLOAT^*$					-0.0031	0.00306	0.00327	0.0023
					(-1.54)	(1.53)	(1.46)	(0.94)
A^*		0.00334	0.00265	0.00295		0.00352	0.00286	0.00313
		(6.59)	(5.43)	(6.26)		(6.78)	(5.87)	(6.68)
$SAMESIZE^*$		0.0134	0.01776	0.02453		0.01624	0.01456	0.02029
		(0.89)	(0.92)	(1.16)		(1.09)	(0.77)	(0.97)
$SAMEBM^*$		0.00063	-0.00031	-0.10748		0.00086	-0.00013	-0.11024
		(1.49)	(-0.66)	(-2.18)		(2.03)	(-0.27)	(-2.22)
$SAMEMOM^*$		0.01204	0.01116	-0.00821		0.01212	0.01126	-0.01745
		(8.11)	(7.24)	(-0.43)		(8.28)	(7.41)	(-0.91)
$NUMNACE^*$		0.00879	0.00743	0.00748		0.00873	0.00745	0.0075
		(15.4)	(15.93)	(15.75)		(15.38)	(16.55)	(16.22)
$SIZE1^*$		-0.01089	-0.00369	0.00727		0.01688	0.01527	0.02261
		(-0.57)	(-0.15)	(0.27)		(0.98)	(0.66)	(0.9)
$SIZE2^*$		0.02248	0.01507	0.00788		0.0258	0.02458	0.01765
		(1.3)	(0.7)	(0.33)		(1.44)	(1.13)	(0.73)
$RETCORR^*$			0.01729	0.01581			0.01716	0.01564
			(10.91)	(10.01)			(10.59)	(9.69)
$ROECORR^*$			0.00359	0.00337			0.00356	0.00334
			(4.68)	(4.58)			(4.62)	(4.5)
$VOLCORR^*$			0.00337	0.00363			0.00343	0.00372
			(3.52)	(4.04)			(3.54)	(4.05)
$DIFFGROW^*$			-0.00193	-0.00238			-0.00215	-0.00245
			(-1.97)	(-2.69)			(-2.12)	(-2.68)
$DIFFLEV^*$			-0.001	-0.00269			-0.00062	-0.00233
			(-0.77)	(-2.19)			(-0.47)	(-1.94)
$DIFFPRICE^*$			-0.00095	-0.00089			-0.00122	-0.00121
			(-1.12)	(-1.01)			(-1.44)	(-1.38)
$DCITY$			-0.00081	-0.00048			-0.00181	-0.00151
			(-0.42)	(-0.26)			(-1)	(-0.86)
$DINDEX$			-0.00686	-0.00551			-0.00464	-0.00352
			(-2.42)	(-1.93)			(-1.62)	(-1.22)
$(SAMESIZE^*)^2$		-0.00762	-0.00972	-0.01221		-0.00914	-0.00911	-0.01166
		(-1.22)	(-1.18)	(-1.36)		(-1.49)	(-1.15)	(-1.33)
$(SAMESIZE^*)^3$		0.00078	0.00142	0.00188		0.0014	0.0015	0.00199
		(0.51)	(0.73)	(0.88)		(0.94)	(0.79)	(0.95)
$(SIZE1^*)^2 \times (SIZE2^*)^2$		-0.00005	-0.0001	-0.00024		-0.00082	-0.00073	-0.00086
		(-0.07)	(-0.1)	(-0.23)		(-1.13)	(-0.8)	(-0.89)
$(SIZE1^*)^2$		-0.00589	-0.00508	-0.00512		-0.00345	-0.00291	-0.00282
		(-2.22)	(-1.99)	(-2)		(-1.24)	(-1.16)	(-1.13)
$(SIZE2^*)^2$		0.00499	0.00583	0.00683		0.0079	0.00919	0.01044
		(2.2)	(2.11)	(2.47)		(3.35)	(3.45)	(3.96)

Continued on next page

... continued from previous page

	Dependent Variable: Correlation of 4F Residuals							
	FLOAT				*DLOAT*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$(SIZE1^*)^2 \times SIZE2^*$		-0.00087 (-0.29)	0.00005 (0.01)	0.00142 (0.33)		-0.00057 (-0.19)	-0.00043 (-0.11)	0.0011 (0.26)
$SIZE1^* \times (SIZE2^*)^2$		-0.00427 (-1.4)	-0.00662 (-1.6)	-0.00815 (-1.82)		-0.00519 (-1.81)	-0.00642 (-1.63)	-0.00787 (-1.82)
$BM1^*$				-0.11749 (-2.03)				-0.12114 (-2.08)
$BM2^*$				0.12917 (2.23)				0.13277 (2.27)
$MOM1^*$				-0.01525 (-0.67)				-0.02609 (-1.16)
$MOM2^*$				0.02453 (1.17)				0.03543 (1.7)
$BM1^* \times BM2^*$				0.02826 (2.86)				0.02651 (2.67)
$(SAMEBM^*)^2$				0.02658 (2.33)				0.02638 (2.3)
$(SAMEBM^*)^3$				-0.00467 (-2.15)				-0.00445 (-2.03)
$(BM1^*)^2 \times (BM2^*)^2$				0.00096 (0.8)				0.00141 (1.17)
$(BM1^*)^2$				0.00242 (0.65)				0.00366 (0.97)
$(BM2^*)^2$				0.0118 (2.82)				0.01217 (2.86)
$(BM1^*)^2 \times BM2^*$				-0.01883 (-2.81)				-0.01825 (-2.71)
$BM1^* \times (BM2^*)^2$				0.0153 (2.35)				0.01498 (2.29)
$MOM1^* \times MOM2^*$				0.0143 (2.71)				0.01713 (3.38)
$(SAMEMOM^*)^2$				0.00806 (1.16)				0.01183 (1.75)
$(SAMEMOM^*)^3$				-0.00156 (-1.03)				-0.00246 (-1.69)
$(MOM1^*)^2 \times (MOM2^*)^2$				0.0028 (3.24)				0.00306 (3.47)
$(MOM1^*)^2$				-0.00774 (-4.59)				-0.00778 (-4.44)
$(MOM2^*)^2$				-0.00568 (-2.74)				-0.00534 (-2.56)
$(MOM1^*)^2 \times MOM2^*$				-0.00714 (-2.28)				-0.00861 (-2.83)
$MOM1^* \times (MOM2^*)^2$				0.00443 (1.29)				0.00597 (1.79)
$R2$	0.09042 (9.91)	0.13263 (11.8)	0.1465 (12.46)	0.15696 (13.05)	0.07928 (10)	0.13018 (11.76)	0.14387 (12.42)	0.15424 (13.01)
No. Obs.	45042 (65.54)	26791 (224.84)	20399 (87)	20399 (87)	46093 (73.56)	27023 (225.82)	20589 (86.76)	20589 (86.76)

Table A3: Common short selling originations and excess correlation.

The table shows the complete results of the Fama and MacBeth (1973) cross-sectional regressions carried out in Table 5. The dependent variable is the realized correlation of a stock pair 4-factor Fama and French (1993) and Carhart (1997) abnormal returns in month $t+1$. The independent variables are updated quarterly and include $\widehat{INSS}^*$, which represents the part of common short selling (NSS) that is due to informed trading. $\widehat{INSS}^*$ is constructed from the fitted values from the regression of NSS^* on $\widehat{F1}^*_{ij,t}$ and $\widehat{F2}^*_{ij,t}$, which are, respectively, the larger and smaller normalised ranked-tranforms of the factor scores from the model given in Equation 2. t-values (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
Constant	0.08989 (17.58)	0.11818 (7)	0.10938 (-0.91)	0.07542 (3.24)
$\widehat{INSS}^*$	0.00789 (6.39)	0.00248 (1.95)	0.00323 (-1.87)	0.0023 (1.77)
A^*		0.00439 (6.82)	0.00382 (11.96)	0.0041 (7.08)
$SAMESIZE^*$		0.05867 (2.72)	0.0514 (2.49)	0.05474 (2.44)
$SAMEBM^*$		-0.00078 (-0.99)	-0.00088 (-0.78)	-0.06264 (-0.91)
$SAMEMOM^*$		0.0142 (7.74)	0.01185 (-0.76)	-0.04664 (-1.87)
$NUMNACE^*$		0.00844 (14.92)	0.00735 (0.98)	0.00729 (11.96)
$SIZE1^*$		0.06837 (2.76)	0.06013 (-2.05)	0.06443 (2.49)
$SIZE2^*$		-0.01936 (-0.78)	-0.0169 (2.52)	-0.02021 (-0.78)
$BM1^*$				-0.06241 (-0.76)
$BM2^*$				0.07924 (0.98)
$MOM1^*$				-0.06132 (-2.05)
$MOM2^*$				0.07114 (2.52)
$RETCORR^*$			0.0196 (10.31)	0.01769 (9.57)
$ROECORR^*$			0.0062 (5.62)	0.00516 (5.16)
$VOLCORR^*$			0.00403 (4.12)	0.00444 (4.88)
$DIFFGROW^*$			-0.00389 (-3.64)	-0.00523 (-5.27)
$DIFFLEV^*$			-0.00063 (-0.38)	-0.00282 (-1.89)
$DIFFPRICE^*$			0.0002 (0.22)	-0.00004 (-0.04)
$SIZE1^* \times SIZE2^*$		-0.03099 (-4.17)	-0.03137 (-4.36)	-0.03245 (-4.05)
$(SAMESIZE^*)^2$		-0.03084 (-3.64)	-0.02874 (-3.6)	-0.0299 (-3.31)
$(SAMESIZE^*)^3$		0.00751 (3.59)	0.00665 (3.32)	0.00666 (2.98)
$(SIZE1^*)^2 \times (SIZE2^*)^2$		-0.00124 (-0.94)	-0.00088 (-0.6)	-0.00091 (-0.62)
$(SIZE1^*)^2$		-0.00488 (-1.49)	-0.00256 (-0.83)	-0.00315 (-1.01)
$(SIZE2^*)^2$		0.01249 (3.05)	0.01384 (3.59)	0.01476 (3.68)
$(SIZE1^*)^2 \times SIZE2^*$		0.00762	0.00814	0.0091

Continued on next page

. . . continued from previous page

	Dependent Variable: Correlation of 4F Residuals			
	(1)	(2)	(3)	(4)
		(1.81)	(2.07)	(2.1)
$SIZE1^* \times (SIZE2^*)^2$		-0.01651 (-3.99)	-0.01726 (-4.41)	-0.01811 (-4.14)
$BM1^* \times BM2^*$				0.01997 (1.5)
$(SAMEBM^*)^2$				0.01508 (0.94)
$(SAMEBM^*)^3$				-0.00271 (-0.86)
$(BM1^*)^2 \times (BM2^*)^2$				-0.00096 (-0.56)
$(BM1^*)^2$				-0.00197 (-0.39)
$(BM2^*)^2$				0.01244 (2.44)
$(BM1^*)^2 \times BM2^*$				-0.01484 (-1.62)
$BM1^* \times (BM2^*)^2$				0.01005 (1.14)
$MOM1^* \times MOM2^*$				0.02796 (3.91)
$(SAMEMOM^*)^2$				0.02453 (2.65)
$(SAMEMOM^*)^3$				-0.00564 (-2.72)
$(MOM1^*)^2 \times (MOM2^*)^2$				0.00519 (4.47)
$(MOM1^*)^2$				-0.00856 (-4.25)
$(MOM2^*)^2$				-0.00655 (-2.57)
$(MOM1^*)^2 \times MOM2^*$				-0.01411 (-3.43)
$MOM1^* \times (MOM2^*)^2$				0.01198 (2.76)
$DCITY$			0.00112 (0.48)	0.00015 (0.07)
$DINDEX$			-0.00781 (-1.68)	-0.00769 (-1.55)
$R2$	0.12252 (11.64)	0.16954 (12.78)	0.18213 (13.37)	0.19521 (13.98)
No. Obs.	18348 (105.29)	13336 (46.23)	12465 (58.36)	12465 (58.36)

Table A4: The determinants of common short selling.

This table reports the complete results of the Fama and MacBeth (1973) cross-sectional regressions for the determinants of NSS^*, carried out in Table 6. Panel A reports the complete results of Panel A of Table 6. The independent variables include $OVERPRICED_{ij,t}$, a dummy variable equal to one if the stock pair is overpriced according to our indicators (and zero otherwise), and $DECLINE_{ij,t}$, a dummy variable equal to one if the stock pair returns have had negative buy-and-hold returns over the last 12 months (and zero otherwise). Panel B reports the complete results of Panel B of Table 6, which were obtained for the subsample of declining stock pairs (i.e, for which $DECLINE_{ij,t} = 1$). In addition to $OVERPRICED$, the independent variables include $UNDERPRICED_{ij,t}$, a dummy variable equal to one if the stock pair is overpriced according to our indicators (and zero otherwise). All other controls are defined as in the text. All variables for regressions in Panel A and B are updated quarterly. t-values (in parentheses) are computed using Newey and West (1987) robust standard errors, accounting for autocorrelation up to 1 lag (one month).

Panel A: Full sample

	Dependent Variable: NSS^*				
Indicator	Book-to-market	Earnings yield	Value-to-price	Accruals	Asset growth
Constant	-0.0083519 (-0.52)	-0.0063428 (-0.39)	-0.0088018 (-0.55)	-0.0047006 (-0.3)	-0.0055573 (-0.34)
$DECLINE$	0.25044 (10.87)	0.25055 (10.67)	0.2511 (10.77)	0.25001 (10.72)	0.25006 (10.68)
$OVERPRICED$	0.084507 (2.97)	0.020745 (0.97)	0.074096 (4.33)	-0.036001 (-2.67)	-0.01283 (-0.48)
$SIZE1$	-0.079306 (-6.77)	-0.078451 (-6.63)	-0.077811 (-6.55)	-0.078761 (-6.63)	-0.078554 (-6.61)
$SIZE2$	0.11928 (6.08)	0.1196 (6.13)	0.12004 (6.11)	0.11984 (6.11)	0.11978 (6.13)
$DIV1$	0.059374 (6.91)	0.059117 (6.83)	0.060926 (7.26)	0.059249 (6.86)	0.058894 (6.88)
$DIV2$	0.021833 (1.49)	0.021922 (1.51)	0.022343 (1.54)	0.021793 (1.49)	0.02159 (1.49)
$INST1$	-0.028529 (4.13)	-0.028032 (4.23)	-0.028596 (4.24)	-0.027887 (4.22)	-0.02769 (4.13)
$INST2$	0.025191 (4.13)	0.025572 (4.23)	0.025673 (4.24)	0.025603 (4.22)	0.025549 (4.13)
R^2	0.03222 (13.57)	0.03196 (13.26)	0.03213 (13.08)	0.03193 (13.33)	0.03205 (13.14)
N. Obs.	40391 (31.47)	40391 (31.47)	40391 (31.47)	40391 (31.47)	40391 (31.47)

Continued on next page

. . . continued from previous page

Panel B: Subsample for declining stock pairs ($DECLINE_{ij,t} = 1$)					
	Dependent Variable: NSS^*				
Indicator	Book-to-market	Earnings yield	Value-to-price	Accruals	Asset growth
$Constant$	0.27492 (9.12)	0.28455 (10.11)	0.28627 (9.32)	0.27954 (10.12)	0.2911 (9.54)
$OVERPRICED$	0.117 (2.56)	0.03302 (0.44)	0.07848 (1.3)	0.13772 (1.93)	0.03726 (0.46)
$UNDERPRICED$	0.11715 (2.17)	-0.05954 (-0.82)	-0.043 (-0.87)	0.08159 (2.66)	-0.11086 (-4.09)
$SIZE1$	-0.13013 (-5.9)	-0.13257 (-6.11)	-0.13113 (-6.22)	-0.1329 (-6.16)	-0.13383 (-6.21)
$SIZE2$	0.24754 (12)	0.24481 (11.74)	0.24585 (11.79)	0.24746 (11.91)	0.24238 (11.4)
$DIV1$	0.05317 (2.5)	0.05154 (2.4)	0.05417 (2.54)	0.05253 (2.44)	0.05015 (2.33)
$DIV2$	0.04002 (1.48)	0.03771 (1.4)	0.03754 (1.37)	0.03813 (1.39)	0.03704 (1.37)
$INST1$	-0.05703 (1.79)	-0.05916 (1.82)	-0.06004 (1.88)	-0.05967 (1.81)	-0.0586 (1.87)
$INST2$	0.04027 (1.79)	0.04087 (1.82)	0.04266 (1.88)	0.04096 (1.81)	0.04165 (1.87)
R^2	0.07063 (12.14)	0.06987 (10.83)	0.07049 (11.34)	0.0699 (11.04)	0.0699 (11.04)
N. Obs.	7665 (4.74)	7665 (4.74)	7665 (4.74)	7665 (4.74)	7665 (4.74)

References

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

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

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

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