APPLIED FINANCE LETTERS

VOLUME 13, 2024

POWER OF CSAD-BASED TEST ON HERDING BEHAVIOUR

HAORAN ZHANG[1*]

1. Manhattan College, USA.

* Corresponding Author: Haoran Zhang Present address: O'Malley School of Business, Manhattan College, 4513 Manhattan College Parkway, Riverdale, NY, USA, 10471.
✉ hzhang05@manhattan.edu

Abstract

This study aims to answer the question of whether the cross-sectional absolute deviation (CSAD)-based test is powerful enough to detect herding behaviour in financial markets. Using US stocks as the main sample, I investigate the power of the CSAD-based test as a herding detection method, with a focus on two dimensions: the self-consistency of the method and the power of t-tests used in the method. I find that conducting the tests with a large number of stocks over extended time periods is likely to provide consistent conclusions on whether herding behaviour exists in the stock market. These findings support the CSAD-based test as a herding detection method. However, with an overall mean of 59.37%, the estimated power of t-tests can be as low as 37.62%, indicating low testing power. Therefore, researchers should be careful when using the CSAD-based test as a herding detection method, especially when $\boldsymbol{R}^2$s are low.

Keywords: herding behaviour, CSAD-based herding detection method

1. Introduction

Since Chang et al. (2000) proposed the herding measure based on cross-sectional absolute deviation (CSAD), researchers have used this method to study herding behaviour worldwide. As a result, it has been determined that herding behaviour exists in different financial markets around the world. The CSAD-based method has a strong theoretical framework built on the capital asset pricing model. However, features of this method have not been fully discussed in the literature. Of the undiscussed features, the power of the herding tests is a significant one. The power of herding tests can be decomposed into two dimensions: the self-consistency of the method and the power of t-tests used in the method.

Table 1 below outlines selected research studies in which the CSAD-based method was used to detect herding behaviours and summarises the sample used in each study. The sample size ranges from 6 to 912. The first question to consider is whether a sample of 6 and a sample of 912 form consistent conclusions on herding behaviour under similar market conditions. If not, then it is important to determine how many stocks should be considered for the studies. Ideally, the results obtained through the CSAD-based method should exhibit convergence towards a stable level as the stock sample size increases. However, there is a lack of evidence to support this expectation. This raises concerns about the method's accuracy and reliability when samples of different sizes lead to different conclusions regarding herding behaviour under similar market conditions.

Table 1: Sample of Selected Studies Using CSAD-Based Methods

Article	Target market(s)/area(s)	Sample size	Sample period	Data frequency
Espinosa-Méndez and Arias (2021)	Stock markets in France, Germany, Italy, the United Kingdom, and Spain	30 to 100	January 3, 2000, to June 19, 2020	Daily
Yarovaya at el. (2021)	USD, EUR, JPY, and KRW cryptocurrency markets	6 to 12	January 1, 2019, to March 13, 2020	Hourly
Bernales et al. (2020)	US equity options market	-*	January 1996, to December 2012	Daily
Youssef and Mokni (2018)	Stock markets in six Gulf Cooperation Council countries	-*	January 5, 2003, to May 28, 2017	Weekly
Kabir and Shakur (2018)	Stock markets in eight Asian and four Latin American countries	171 to 912	January 1, 1995, to December 31, 2014	Daily
Pochea et al. (2017)	Central and East European stock markets	9 to 331	January 2, 2003, to December 31, 2013	Daily
Philippas et al. (2013)	US REIT market	112 to 152	January 2004, to December 2011	Daily
Economou et al. (2011)	Portuguese, Italian, Spanish, and Greek stock markets	49 to 337	January 1998, to December 2008	Daily
Chiang and Zheng (2010)	Stock markets in 18 countries	53 to 155	May 25, 1988, to April 24, 2009	Daily
Tian et al. (2008)	Stock market and its submarkets in China	54 to 746	July 12, 1994, to December 31, 2003	Daily, weekly, monthly

*No sample size is specified in the studies.

The sample periods are also critical. Most articles listed in Table 1 have sample periods of more than 10 years. For example, Bernales et al. (2020) used the CSAD-based method to study a daily dataset spanning from January 1996 to December 2012 and provided approximately 4,215 daily observations. Conversely, Youssef and Mokni (2018) evaluated a weekly dataset and offered approximately 751 weekly observations. In theory, more observations are associated with higher testing power, which raises the question of whether 751 observations are sufficient.

Another open question concerns the power of the t-tests used in the method. The power of the t-tests directly impacts the power of the CSAD-based method, setting the upper bound of its testing power. However, previous studies show little regard for the statistical power of tests in the field of finance. Kim and Jin (2015) conducted a survey on 161 articles published in four journals, Journal of Finance, Journal of Financial Economics, Journal of Financial and Quantitative Analysis, and The Review of Financial Studies, in 2012 and found only one article discussing the power of tests. None of the previous works discuss the statistical power of the CSAD-based method. In this study, I address these gaps in the literature and provide insights into the unanswered questions.

The remainder of this study is organised as follows: Section 2 describes the sample and data, Section 3 discusses the self-consistency of the method, Section 4 discusses the power of the t-tests used in the CSAD-based test, and Section 5 concludes the paper.

2. Sample and data

In this study, I selected the S&P 500 stock universe as of June 30, 2023, for the main sample pool. I retrieved daily gross returns, including distributions, of S&P 500 stocks from Bloomberg. The sample period spans from 2016 to 2022. I excluded any stocks that did not have consecutive daily returns from the full sample period, resulting in a pool of 454 stocks for the sample.

3. Self-consistency of the CSAD-based method

I followed the method proposed by Chang et al. (2000) to detect herding behavior in the stock market. I constructed CSAD for each day using Equation (1):

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|r_{it} - \bar{r}_t|, \tag{1}$$

where N is the number of sample stocks, r_{it} is the return of stock i at time t, and $\bar{r}_t$ is the equally weighted average return of all sample stocks at time t.

$$CSAD_t = \alpha + \beta_1|\bar{r}_t| + \beta_2\bar{r}_t^2 + \varepsilon_t. \tag{2}$$

If herding behaviour exists in the market, β_2 should be negative and significant in the regression. The t value of the CSAD-based test for H_α: $\beta_2 < 0$ was the main variable of interest in this study. I also provided results on estimated H_α: β_2 when the t value was not appropriate to draw a conclusion.

3.1 Number of sample stocks and convergence of β_2

Whether sample pools of different sizes reach the same conclusion under identical market conditions is the first topic to address in this research. The method's power could be low if, despite including sufficient sample stocks, pools of different sizes yield divergent conclusions. Furthermore, researchers must consider the question of how many stocks are necessary for an acceptable sample size. If we can use 10 stocks to provide a solid conclusion on herding behaviour in the market, including 3,000 stocks in the sample would be pointless. I provided answers to the above questions based on simulations. The simulations to estimate the t values involve the following steps:

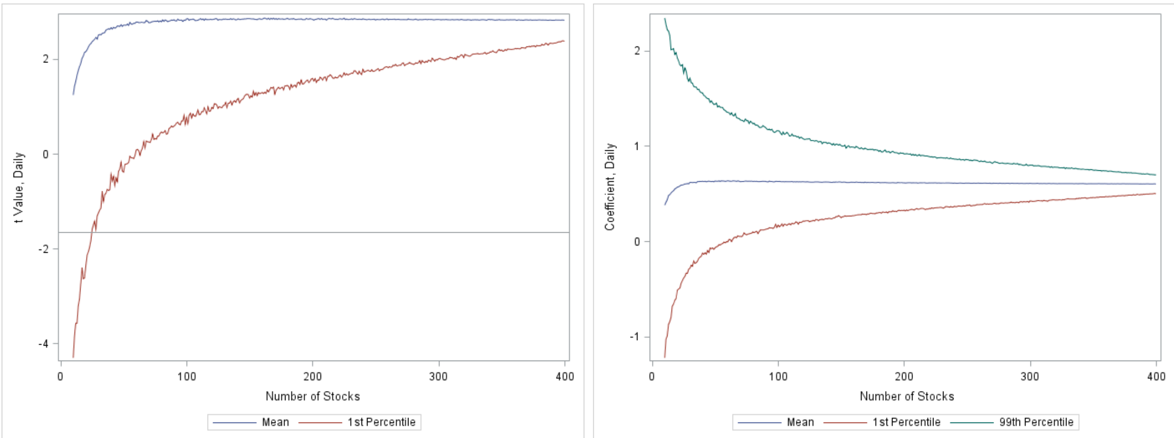
i. Sample N stocks from the sample pool. N is the number of sample stocks in this simulation.
ii. Obtain returns of the stocks in the sample period.
iii. Calculate CSADs and equally weighted "market" returns based on the sample stocks [Equation (1)].
iv. Estimate the t value based on Equation (2).
v. Repeat the process 5,000 times[1].

[1] The number of simulations is determined by separate tests. Please see details in Appendix

For each N, or each set of simulations, I estimated 5,000 t values and β_2s and used the results to draw conclusions on herding behaviour. N increases from 10 to 400[2]. Figure 1(a) shows the mean t value and 1st percentile of the t values estimated from the set of simulations for each N. The reference line at the bottom is the critical value of a 5% significance level given the number of observations, 1,821. As N increases, the mean t value stabilises to a level well above the critical value. The important indicator, 1st percentile of t values, is also consistently above the critical value after N moves beyond about 30. This means that as the number of sample S&P 500 stocks increases beyond about 30, at least 99% of simulations in a set have t values that indicate a failure to reject the null hypothesis (H0: $\beta_2 \geq 0$), consistently demonstrating that herding behaviour does not exist in this sample period in the US stock market. The results imply that we may not need 500 sample stocks to provide a solid conclusion on herding behaviour, although 10 or 20 stocks may also be insufficient. The exact necessary number of sample stocks may not be evident, but the findings suggest that more than 50 stocks could provide a consistent answer on herding behaviour in the stock market.

The conclusions were drawn based on one assumption: as the number of sample stocks increases, the estimated β_2 should converge to a stable level. If the estimated β_2 changes dramatically without a stable terminal level, this could indicate that the CSAD-based method is not reliable in detecting herding. However, Figure 1(b) shows that the mean β_2 converges, validating the assumption and demonstrating the self-consistency of the CSAD-based method.

Figure 1: Change in the Variables of Interest as the Number of Sample Stocks Increases



(a) (b)

3.2 Sample period length

Sample period length is another important factor. As shown in Table 1, the sample periods of previous studies vary. Sample period length determines the number of observations in the regression described by Equation (2), and affects test results and conclusions on herding. In this section, I

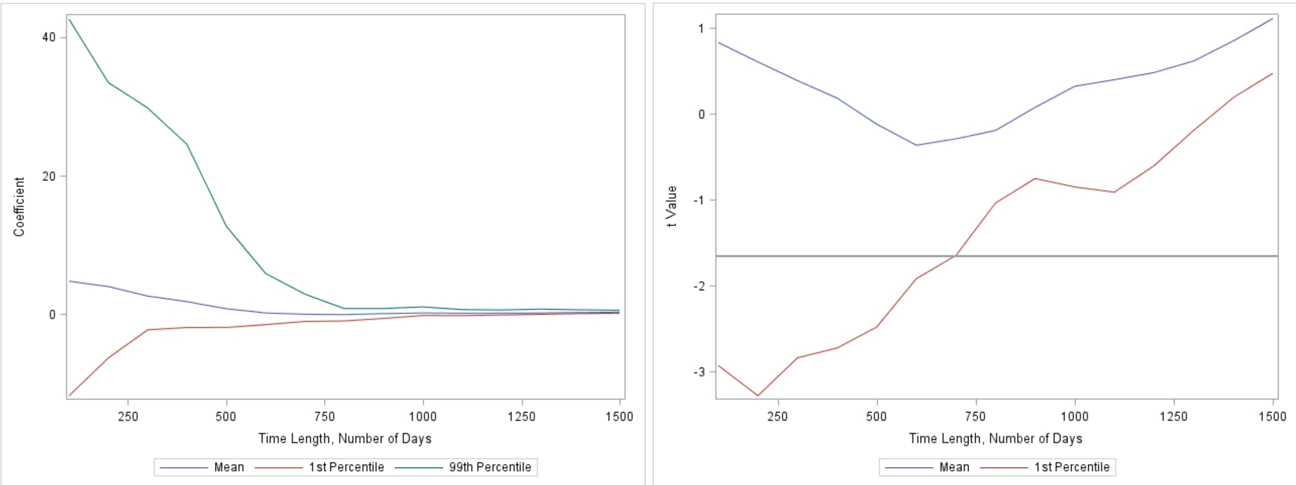
[2] *N* should not be close to the total number of stocks, 454, for this sample pool, or the simulations will provide almost identical *t* values.

provided evidence on the self-consistency of the method with respect to change in time length. I modified the first step of the procedure described in Section 3.1 as follows:

i. Sample *T* consecutive trading days from the whole period, 2016 to 2022, as the sample period of this simulation. All stocks are used as sample stocks. The start day of the sample period may not be the first trading day of 2016.

For each T, 5,000 t values and β_2s were estimated. T increases from 100 to 1,500. Figure 2 shows that the mean β_2 converges as the length of the sample period increases. After T reaches about 700, more than 99% of simulations have t values that indicate consistent no-herding conclusions. These results suggest that a sample period of at least 700 days can help researchers avoid inconsistent conclusions.

Figure 2: Change in the Variables of Interest as the Sample Period Length Increases



(a) (b)

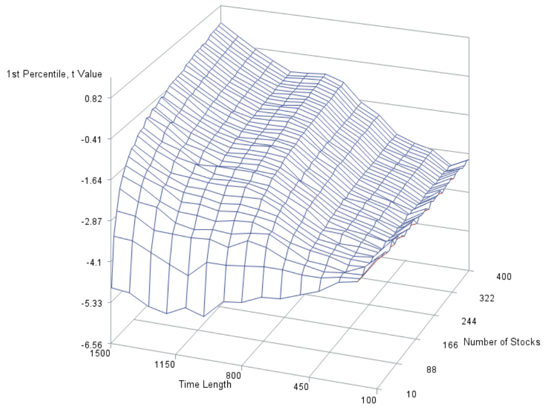
3.3 Number of sample stocks versus length of sample period

In the sample formatting process, stocks with missing values were excluded from the sample pool, creating a trade-off between the number of sample stocks and the length of the sample period. A longer sample period is likely associated with fewer stocks in the sample pool in many cases. This is especially an issue for stock markets without large trading volumes or solid trading records. If extending the sample period results in a lower number of sample stocks, one may question how the sample can be formed to maximise the consistency and power of the test. To study this topic, I modified the first step of the procedure described in Section 3.1 as follows:

i. Sample N stocks from the sample pool as sample stocks and T consecutive trading days from the whole period as the sample period in this simulation.

For each N and T, 5,000 t values were estimated. N ranged from 10 to 400 by 10, and T ranged from 100 to 1,500 by 100. As shown in Figure 3, the 1st percentile of t values increases as N and T increase. When the number of days and the number of sampled stocks are large, the 1st percentile of t values is well above the critical values, which are around 1.64. However, when the sample period is sufficiently long, the growth rate of the 1st percentile relative to the number of sample stocks is higher compared to its growth rate relative to the time length, provided the number of sample stocks is sufficiently large. The evidence indicates that when there is a conflict, it's more important to focus on increasing the number of stocks in the sample rather than extending the length of the sample period.

Figure 3: Change in 1st Percentile of t Values as the Sample Period Length and Number of Sample Stocks Increase



4. Power of the t-tests

The main results on herding are derived from the t-test for Hα: $\beta_2 < 0$. As a result, the power of the CSAD-based test should not be greater than the power of the t-test. A potential problem is that the power of t-tests in the regressions may be too low, making it unlikely to reject the false null hypothesis. Therefore, there is a significant probability that herding exists, but the model may not be capable of detecting it. The homogenous no-herding conclusions in the simulations may not be the results of no-herding conditions in the market; instead, they may be driven by the weak power of the t-tests. To address this issue, I followed the method described by Cohen (1988) and used Equations (3) to (5) to estimate the power of the t-tests in the regressions.

$$Cohen's\ f^2 = \frac{R^2_{with\ \bar{r}_t^2} - R^2_{without\ \bar{r}_t^2}}{1 - R^2_{with\ \bar{r}_t^2}}, \tag{3}$$

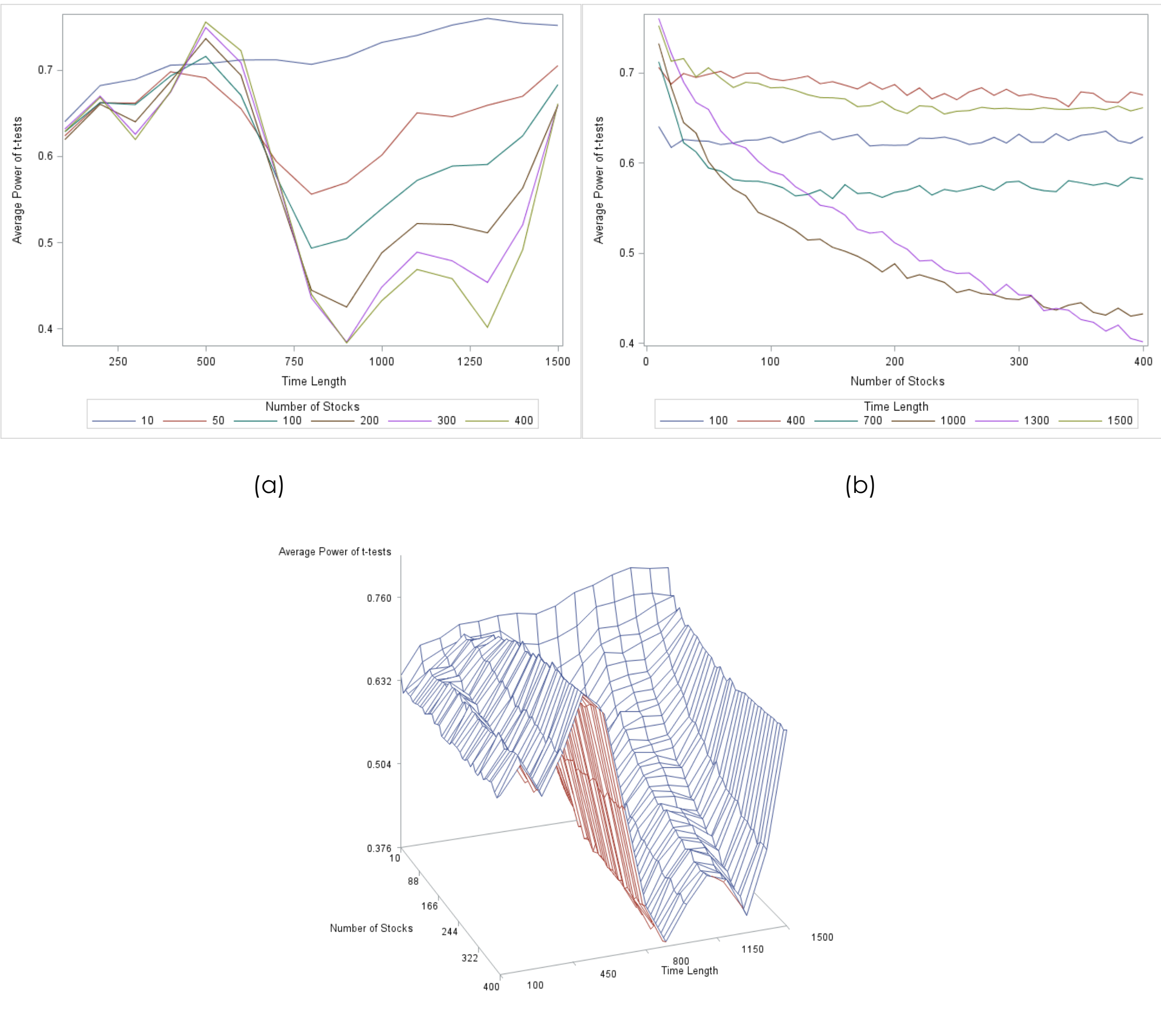
$$\lambda = Cohen's\ f^2 T, \tag{4}$$

$$Power = F(\lambda, u, v), \tag{5}$$

where $R^2_{with\ \bar{r}_t^2}$ and $R^2_{without\ \bar{r}_t^2}$ are the R^2s of the full model [Equation (2)] and the model without $\bar{r}_t^2$, respectively. T is the number of observations in regressions. $F(\lambda, u, v)$ is the cumulative probability given F-value, λ, and degrees of freedoms, u, v. Because there is only one variable of interest, $\bar{r}_t^2$, u is set to 1. Lastly, v is T- u -1. I used the procedure in Section 3.3 but focused on the power of t-tests instead of the t value. The number of observations, which determines the degree of freedom for the t-test, is a result of the sample period length and observation frequency in this section. The number of sample stocks does not directly impact the number of observations because the regression described by Equation (2) is a pure time-series regression.

The average power of each set of simulations for different N and T was reported in Figure 4. As a rule of thumb, it is believed that 80% is an acceptable level of testing power (see, for example, De Winter, 2019; Serdar et al., 2021). The average power ranges from 37.62% to 75.98% with a mean of 59.37%, indicating low testing power in this study. Surprisingly, the number of observations in regressions does not play a larger role in the testing power. As shown in Figure 4(a), the testing power first decreases significantly and then increases as the number of days increases. The pattern observed with fewer sample stocks exhibits a flatter curve. In theory, a higher number of observations, namely, days in this study, should be associated with higher testing power. However, the evidence does not support this expectation, indicating potential problems with the model specifications. Furthermore, Figure 4(b) indicates that as the number of sample stocks increases, the testing power decreases. Additionally, longer sample length is associated with a steeper dive in average power as the number of sample stocks increases. As discussed in Section 3.1, a higher number of sample stocks should make the testing results more stable. However, the findings in this section are not consistent with this expectation. The declining testing power raises another concern: when sample sizes are large, it may be that the consistent conclusions about herding are driven by the low testing power, which questions the model's ability to accurately detect herding. The joint effect shown in Figure 4(c) confirms the findings.

Figure 4: Change in Average Power of T-tests as the Sample Period Length and Number of Sample Stocks Increase



(a) (b)

(c)

Two factors may contribute to the low power of the t-tests. First, stock returns are featured in a high level of noise and are influenced by many market and fundamental factors. The dependent variable, CSAD, is itself a deviation measure, which could be noisy. Second, the model includes only two independent variables, which may not be sufficient to explain the whole variance of the dependent variable. As a result, R^2s of the full and reduced models are low. The average R^2 is about 35.89% in this section, and some previous studies document even lower R^2 (see, for example, Espinosa-Méndez and Arias, 2021; Ukpong et al., 2021). The low R^2 leads to a relatively low $Cohen's\ f^2$ and low power of a t-test. Future research in this field should take the power of the herding test into consideration, especially when R^2 of the model is low.

5. Conclusion

The CSAD-based method provides a convenient way for researchers to detect herding behaviour in a market. Therefore, it is important to know about the power and features of the method. In this study, I find that the CSAD-based method provides consistent results if the number of stocks included in the tests is more than 30 and/or the sample period is more than 700 trading days. Moreover, evidence shows that as additional stocks and trading days are included in the tests, the method tends to produce more consistent conclusions on herding behaviour. These findings support the CSAD-based method; however, evidence also demonstrates that the power of the t-tests used in the method is low overall, indicating that the method may suffer from low testing power problems. Researchers should take the testing power into consideration when conducting herding tests based on CSAD.

References

Bernales, A., Verousis, T., & Voukelatos, N., 2020. Do investors follow the herd in option markets? Journal of Banking & Finance, 119, 104899.

Chang, E.C., Cheng, J.W., & Khorana, A., 2000. An examination of herd behavior in equity markets: An international perspective. Journal of Banking & Finance, 24(10), 1651-1679.

Chiang, T.C., & Zheng, D., 2010. An empirical analysis of herd behavior in global stock markets. Journal of Banking & Finance, 34(8), 1911–1921.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Routledge.

De Winter, J.C., 2019. Using the student's t-test with extremely small sample sizes. Practical Assessment, Research, and Evaluation, 18(1), 10.

Economou, F., Kostakis, A., & Philippas, N., 2011. Cross-country effects in herding behaviour: Evidence from four south European markets. Journal of International Financial Markets, Institutions and Money, 21(3), 443–460.

Espinosa-Méndez, C., & Arias, J., 2021. COVID-19 effect on herding behaviour in European capital markets. Finance Research Letters, 38, 101787.

Kabir, M.H., & Shakur, S., 2018. Regime-dependent herding behavior in Asian and Latin American stock markets. Pacific-Basin Finance Journal, 47, 60–78.

Kim, J.H., & Ji, P.I., 2015. Significance testing in empirical finance: A critical review and assessment. Journal of Empirical Finance, 34, 1-14.

Philippas, N., Economou, F., Babalos, V. & Kostakis, A., 2013. Herding behavior in REITs: Novel tests and the role of financial crisis. International Review of Financial Analysis, 29, 166–174.

Pochea, M.M., Filip, A.M., & Pece, A.M., 2017. Herding behavior in CEE stock markets under asymmetric conditions: A quantile regression analysis. Journal of Behavioral Finance, 18(4), 400–416.

Serdar, C.C., Cihan, M., Yücel, D., & Serdar, M.A., 2021. Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochemia medica, 31(1), 27-53.

Tian, L., Chiang, T.C., Mason, J.R., & Nelling, E., 2008. Herding behavior in Chinese stock markets: An examination of A and B shares. Pacific-Basin Finance Journal, 16(1–2), 61–77.

Ukpong, I., Tan, H., and Yarovaya, L., 2021. Determinants of industry herding in the US stock market. Finance Research Letters, 43, 101953.

Yarovaya, L., Matkovskyy, R., & Jalan, A., 2021. The effects of a "black swan" event (COVID-19) on herding behavior in cryptocurrency markets. Journal of International Financial Markets, Institutions and Money, 75, 101321.

Youssef, M., & Mokni, K., 2018. On the effect of herding behavior on dependence structure between stock markets: Evidence from GCC countries. Journal of Behavioral and Experimental Finance, 20, 52–63.

Appendix A: Choosing the number of simulations in each simulation set

To find a reasonable number of simulations per set for the main results, I ran sets of simulations to check when the variables of interest were stabilised. The procedure is similar to the one described in Section 3.1. The numbers of sample stocks from the sample pool were fixed at 10 and 400. The number of simulations in each set increased from 10 to 10,000 by 10. Figure A.1 shows the results. After the number of simulations in each set reached 5,000, the mean β_2 and the mean t value converged to stable levels. The number of sample stocks in the sampling process is relevant, but the results were consistent. In this study, I have selected 5,000 simulations for each set (as indicated in Figure A.1) to guarantee comprehensive and reliable results.

Figure A.1: Change in Variables of Interest as the Number of Simulations Increase