



LUND UNIVERSITY

Revisiting US Herding – Accounting for the Dynamics

Almgren, Pelle

2026

Document Version:
Other version

[Link to publication](#)

Citation for published version (APA):

Almgren, P. (2026). *Revisiting US Herding – Accounting for the Dynamics*. (Working Papers; No. 2026:9).

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2026:9

Department of Economics
School of Economics and Management

Revisiting US Herding – Accounting for the Dynamics

Pelle Almgren

September 2026



LUND
UNIVERSITY

Revisiting US Herding – Accounting for the Dynamics*

Pelle Almgren[†]

August, 2025

Abstract

I revisit the US stock market during COVID-19 and investigate the effect of news sentiment and the pandemic on herding behavior in an extended, dynamic, version of the model introduced in Chang et al. (2000). This extension is motivated both econometrically and theoretically and the results indicate that treating herding as a dynamic phenomena may reveal insights into the mechanisms driving this behavior. Evidence of herding is identified in the full sample (12th December 2017 to 18th February 2022) and during the COVID-19 period (20th January 2020 to 18th February 2022), whereas no evidence of herding is found in the pre-COVID-19 sample (12th December 2017 to 19th January 2020). Sentiment is further shown to influence herding in different directions and with different dynamics in the two regimes. While negative sentiment is shown to contemporaneously increase herding during COVID-19, the opposite reaction is found for the pre-pandemic period. Before COVID-19, the initial response to sentiment changes over the week.

Keywords: Herding behavior, Time Series Analysis, COVID-19, Dynamic Effects

JEL Classification numbers: C32, C53, G01, G40

*I am grateful to Joakim Westerlund for his guidance and support throughout this project. I also want to thank Anders Vilhelmsson for valuable discussion, insights and comments.

[†]Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. E-mail address: pelle.almgren@nek.lu.se

1 Introduction

Considerable efforts have been devoted to understanding investor behavior and factors influencing it. Following a period of financial disruptions and increases in volatility related to financial crises such as the Global Financial Crisis (September 2008 to May 2010), the European Debt Crisis (May 2010 to July 2013), and more recently, the COVID-19 pandemic (March 2020 to May 2023), one field within financial economics that has received much attention is herding behavior. This refers to when investors, to a large extent, transact in the same direction, a behavior that may have significant impact on financial markets. When investors intentionally trade based on the action of others rather than fundamental information, it may induce excess volatility and systemic risk (Bikhchandani and Sharma, 2000), and lead to market bubbles and crashes (Lux, 1995).

Herding may be influenced by several mechanisms, the most prominent being the dynamics of investor reputation and compensation schemes (Scharfstein and Stein, 1990; Bikhchandani and Sharma, 2000), and a general disregarding of private information in favor of the actions of seemingly better informed investors (Welch, 2000). The first two relate to the inner dynamics of the investment market where investors are constantly evaluated against their peers and compensated based on their relative performance. This creates motives not to deviate too far from the market consensus (Scharfstein and Stein, 1990; Lux, 1995), why this is often referred to as rational herding. The third mechanism, disregarding private information to follow seemingly better informed investors is commonly denoted irrational herding (Chang et al., 2000), pressure for which often increases during periods of elevated fear and information cascades (Economou et al., 2018; Bikhchandani and Sharma, 2000).

Over the last couple of years, increased effort has been devoted to determining conditions and factors that affect these mechanisms, and consequently influence herding behavior. Such conditions and factors often include market volatility, cross-market spillovers and financial crises (Economou et al., 2018; Chang et al., 2020; Duygun et al., 2021; Ferreruela and Mallor, 2021). Additionally, since 2020, many studies have focused on the effect of COVID-19 on herding behavior (see e.g. Wu et al., 2020; Espinosa-Méndez and Arias, 2021; Yarovaya et al., 2021; Jia et al., 2022; Mandaci

and Cagli, 2022; Ampofo et al., 2023). Much attention has also been directed at the effect of sentiment on herding (see e.g. Clements et al., 2017; Economou et al., 2018; Choi and Yoon, 2020; Hudson et al., 2020; Garcia, 2021; Kim et al., 2022).

Herding, especially towards a market consensus, is commonly studied on the basis of individual stocks. However, following the historical and current importance of mutual funds (Elton and Gruber, 2013; Ben-David et al., 2017; Investment Company Institute, 2025) and the rapid growth, and characteristics, of exchange-traded funds (ETFs) (Poterba and Shoven, 2002; Ben-David et al., 2018; Oliver Wyman, 2023), it may be important to study herding at the index level directly. This is because a majority of ETFs and mutual funds have broad indices as underlying assets (Oliver Wyman, 2023; ETFGI, 2025; Investment Company Institute, 2025), and trading in ETFs and mutual funds exerts pressure on the prices of the underlying securities (Ben-David et al., 2018). While for ETFs this mainly works through an (no) arbitrage channel (Poterba and Shoven, 2002; Ben-David et al., 2017, 2018), the channel through which mutual fund flows impact prices of the underlying securities is more complex and may have asymmetric effects among fund constituents, see Hau and Lai (2017) for details. Still, trading in the overarching assets induce trading in the underlying securities.

Further, demand shocks in the overarching instruments have been shown to propagate to the underlying securities (Ben-David et al., 2017). This propagation is especially strong during events of market stress which affect the underlying assets by impacting liquidity and by injecting nonfundamental volatility into market prices of the underlying assets (Ben-David et al., 2017, 2018). This is interesting from a herding perspective as a demand shock induced by market stress on the mutual funds and ETFs market is expected to propagate to a demand shock in the underlying assets, which together with the market stress may induce herding in the underlying broad indices directly.

Additionally, herding in the overarching market likely propagate to herding in the underlying assets through this very channel. Thus, by using indices as a proxy for mutual funds and ETFs and study herding at the index level directly, it would be possible to provide further knowledge of the link between mutual funds and ETFs, and their underlying assets.

Herding can in practice, and as noted above, occur at many different levels and,

consequently, different models have been proposed to study this behavior. Given its historical importance and increased recent attention, the focus of this paper is herding towards a market consensus, i.e. market-level herding. The models of Christie and Huang (1995) and Chang et al. (2000), henceforth referred to as the CH and CCK model respectively, are widely used in this strand of the literature. To investigate herding on the buy-, or sell-side in individual assets, the Lakonishok et al. (1992) model, henceforth referred to as the LSV model, and extensions thereof, is commonly used. Herding in analysts' recommendations is often studied in the context of analysts' earnings forecast dispersion (e.g. Garcia, 2021). As this paper examines herding towards the market consensus I will use an appropriate market-level herding model. Specifically, the CCK model will be used as it improves on the bias that may arise in the CH model from outliers in the data (Chang et al., 2000). Herding in this model is measured in terms of the cross-sectional absolute dispersion (CSAD) of returns (Chang et al., 2000). Since indices behave like diversified portfolios, extreme individual returns tend to cancel out on normal trading days. Computing CSAD based on index returns is therefore expected to yield a smaller cross-sectional dispersion compared to calculations based on individual stock returns. Hence, any conclusion about herding behavior drawn from an analysis based on such a measure should consequently be regarded as conservative with respect to the herding in the underlying securities.

The CCK model is traditionally a univariate static model where CSAD is regressed on two covariates capturing market returns. With static, I refer to a model in which there are no lags, or other time series components, although the data varies over time. After including variables hypothesized to influence herding, e.g. sentiment and COVID-19 interactions, this model is commonly estimated with OLS using heteroscedasticity and autocorrelation consistent (HAC) standard errors (see e.g. Chang et al., 2000; Chen, 2013; Choi and Yoon, 2020; Jia et al., 2022)¹.

At the asset level, persistency of institutional herding on the buy or sell side has been studied within the LSV model (e.g. Chattopadhyay et al., 2018; Merli and Roger, 2013). Garcia (2021) considers lags of the explanatory variables in a study of sentiment

¹Some studies additionally consider quantile regression approaches to explore heterogeneity along the CSAD distribution, yet in a static CSAD setting (e.g. Choi and Yoon, 2020; Ampofo et al., 2023).

effects on herding in analysts recommendations. Chang et al. (2020) is, however, to the best of my knowledge, the only paper to consider lags in the standard univariate CCK model. They do this by studying the effect of (lagged) cross-market herding spillovers in the energy market. Consequently, empirical evidence of persistency in this framework is scarce and the effect of dynamics remain largely unexplored. Yet, the literature on behavioral economics is often suggestive of persistence. For example, habit formation makes households less likely to change portfolio holdings (Muñoz, 2006) and habit persistence is found in liquid assets (Fleissig and Swofford, 2023). Psychological biases are further shown to lead investors to load on recent well performing stocks or stocks with abnormally high trading volume (Barber et al., 2009).

Additionally, the previously discussed mechanisms hypothesized to drive herding behavior should arguably be considered as dynamic phenomena. This suggests that a period of high (low) levels of herding is likely followed by another period of high (low) levels of herding. Therefore, including lags in the model may be important as it can help better capture this persistence and thus gain new insights of the existence, and drivers, of investor herding in a more realistic, i.e. dynamic, setting.

As with the variety of available measures of herding, many different definitions of sentiment have been used. Following the growing literature on textual analysis, sentiment measures derived from news are competing with the market-based measures traditionally used in financial research. Currently, state-of-the-art news-based sentiment measures rely on deep learning algorithms as in Garcia (2021). Market-based measures refer to volatility indexes such as the CBOE VIX (see e.g. Economou et al., 2018; Jia et al., 2022), and sentiment proxies in the spirit of Baker and Wurgler (2006) (see e.g. Blasco et al., 2018; Jia et al., 2022). Research indicates that news-based sentiment measures offer several advantages over market-based ones. For example, recent publications confirm the idea that news influence and shape both investor and consumer sentiment (Boubaker et al., 2021; Adämmer et al., 2024) and sentiment derived from news is shown to influence herding in analysts recommendations (Garcia, 2021). Non-market based sentiment measures are considered exogenous to the financial market (Kaplanski and Levy, 2017), thus providing a more objective measure.

On the methodological side, sentiment classification methods relying on dictionar-

ies and linguistic parsing (as used in Clements et al., 2017 and Kim et al., 2022) have been criticized for their inability to capture contextual nuances and their dependence on predefined word lists (Leippold, 2023). In contrast, machine learning algorithms, particularly Large Language Models (LLMs) like FinBERT, substantially outperform dictionary approaches in accuracy and other evaluation metrics (Huang et al., 2023).

This section has identified three main gaps in the existing literature. First, although well-founded in theory, the empirical evidence of persistency remain largely unexplored in the CCK model, i.e. in a model that concerns herding towards a market consensus. Second, with the link between mutual funds and ETFs and the underlying assets as a backdrop, it is interesting to study herding at the index level directly. To the best of my knowledge, Ampofo et al. (2023) is the only paper to study herding based on indices, although in a static CCK model. Third, the existing papers studying sentiment effects using news-based measures in this model rely on out-dated sentiment classification algorithms which are likely biased (see Leippold, 2023). Therefore, in this paper I measure the sentiment of news headlines with FinBERT, a state-of-the-art LLM pre-trained on financial text, and extend the static CCK model to allow for lags of both cross-sectional return dispersion and the news-based sentiment variable. To have comparability, I revisit the setting recently considered by Ampofo et al. (2023), who study how COVID-19 influence herding in two developed markets, the US and the UK. I focus on the American market and extend the scope to also consider the effect of sentiment on herding.

While evidence of herding is found in the entire sample and during COVID-19, I find no such evidence in the period before COVID-19. This is consistent with Duygun et al. (2021); Espinosa-Méndez and Arias (2021); Ampofo et al. (2023). The results indicate that the lags are an important addition to the model as these have significant effects on the standard regressors in the CCK framework and shed light on the mechanisms influencing herding. Their inclusion reveal differences in how investors are influenced by sentiment in the COVID-19 period viz a viz the period prior to the pandemic. Although no sign of herding is detected before COVID-19, positive news sentiment is shown to contemporaneously decrease return dispersion. The results further suggest that this effect changes over the business week, indicating

a complex relationship in this period. During COVID-19, however, evidence suggest that investors payed more attention to, and acted on, the sentiment conveyed by news media as this signal is immediately absorbed into return dispersion. Here, negative news is found to decrease return dispersion, thus to increase herding. This suggests that, during COVID-19, investors herd on negative news while they held on to private information in the face of positive news. The opposite reaction to news sentiment is found for the period before the COVID-19 pandemic.

The rest of the paper is structured as follows: Section 2 presents the data, Section 3 the standard CCK model, my extension and discuss estimation. Monte Carlo evidence in support of my estimation strategy is presented in Section 4 while the empirical results are presented in Section 5. Section 6 provides a discussion and concludes.

2 Data

The price and news data used in this study was downloaded from Capital IQ and Global Newsstream, respectively. As in the comparison paper, Ampofo et al. (2023), the sample period considered spans from the 5th of December 2017 to the 28th of February 2022. After removing observations from non-trading days, the data consists of 1065 observations. However, as this study makes use of lags of up to five days, see Section 3.2 for a discussion on the modeling strategy, the first five observations are automatically removed meaning that the five last observations has to be excluded to keep the balance between observations in each regime. Performing this trimming leaves a final sample of 1055 observations for the empirical analysis, spanning the 12th of December 2017 to the 18th of February 2022.

The acquired stock data consists of closing prices adjusted for dividend payout of three major US stock indices on which many ETFs and mutual funds have been issued, the S&P 500, the Dow Jones Industrial Average and the NASDAQ Composite². In equation (1), daily returns (in per cent) are computed as the logarithmic difference

²Note that while the three indices included in the paper partially cover the same markets, there are important differences in the weighting scheme used to construct the indices. While the S&P 500 and the NASDAQ Composite are both based on market capitalization, the weights in the S&P 500 account for publicly traded shares only. No such distinction between public and private shares are made for the weights in the NASDAQ Composite. The DJIA is further a price-weighted index. Due to differences in weighting, and differences among the constituents, these three indices provide sufficiently different information to justify their inclusion in the paper.

between the day t and $t - 1$ closing prices for index i . This difference is scaled by 100.

$$R_{i,t} = \ln \left(\frac{P_{i,t}}{P_{i,t-1}} \right) * 100 \quad (1)$$

The prices and returns on the different indexes are plotted in Figure 1. The first confirmed US case of COVID-19 was reported the 20th of January 2020 (Holshue et al., 2020; CDC, 2023), after which all indexes face initial sharp declines and increases in volatility. However, within a short period, all prices steadily increased far beyond previous levels. This is suggestive of time dependence in the price data.

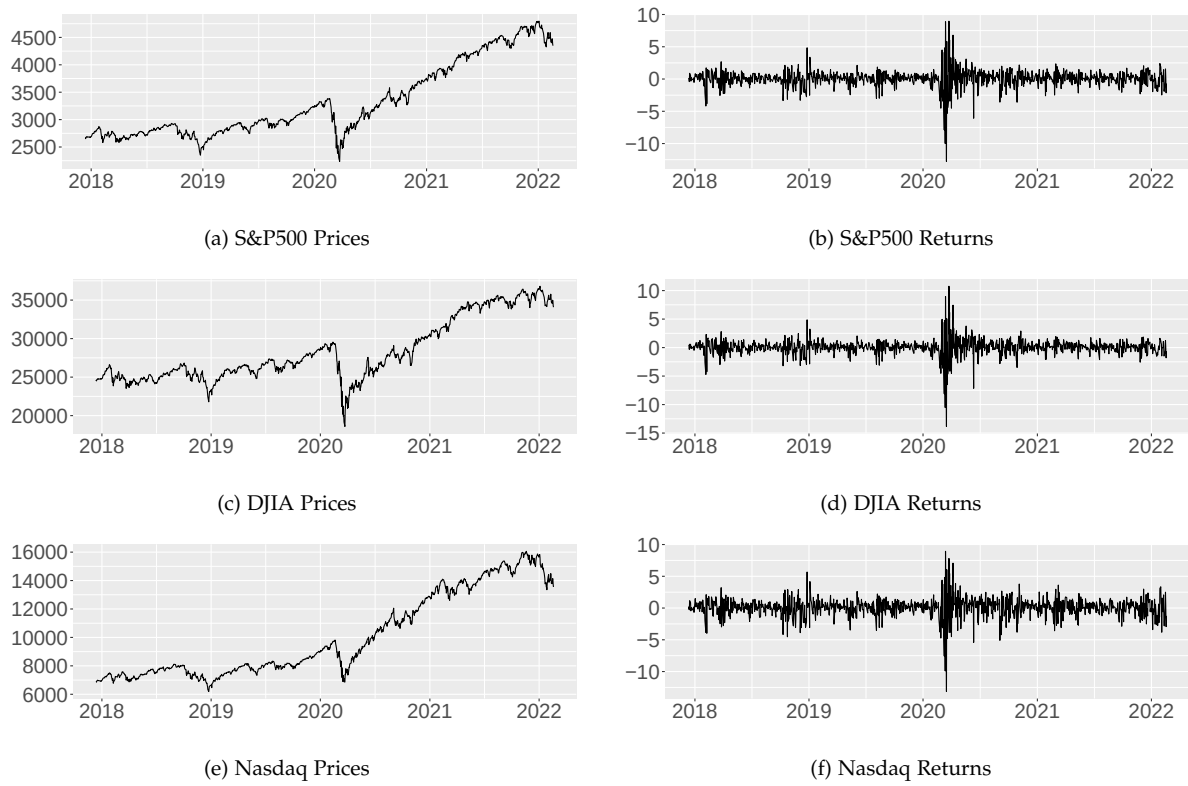


Figure 1: Prices and Returns for all Indexes

Further, as expected in financial data, the index returns plotted in the right column of Figure 1 reveal signs of volatility clustering for all three indexes. This is the case for both the period before and after the initial phase of COVID-19.

I use the cross-sectional absolute dispersion (*CSAD*), introduced in Chang et al. (2000), to study herding. This measure captures dispersion around the market mean

and is calculated as follows for each day t :

$$CSAD_t = \frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_t^M|, \quad (2)$$

where N refers to the number of indices used and R_t^M is the market return of an equally weighted portfolio of all indices. $CSAD$ and the market return, both in percent, are plotted below. The COVID-19 pandemic induced higher levels and variation of $CSAD$ which is visible in the left panel in Figure 2. Visual inspection hints at a possibly different relationship in the factors influencing $CSAD$ in the period before and during COVID-19. The right panel of Figure 2 also suggest a time-varying clustering of the variance of returns.

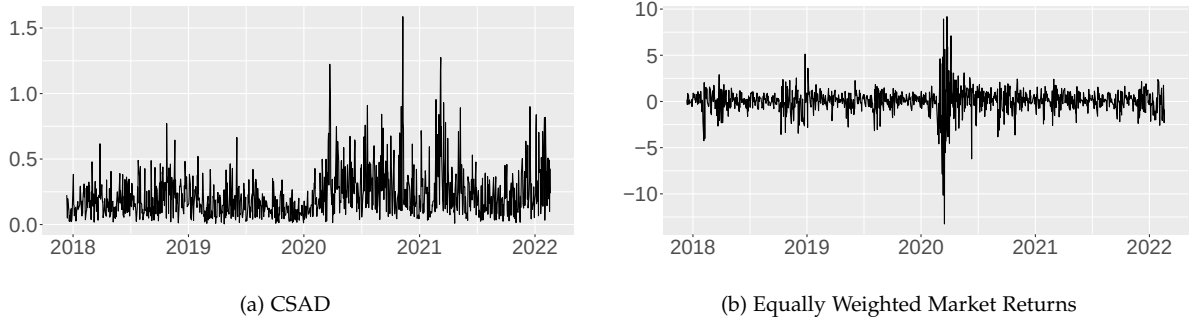


Figure 2: CSAD and Market Returns during 2017-12-05 - 2022-02-28

Regarding news sentiment, Andres-Escayola et al. (2024) find evidence that maximizing the breadth of press coverage when constructing newspaper-based indices helps in reducing the bias that may arise from using news from a single source. Therefore, news data was collected for the four largest daily US newspapers measured in terms of circulation. A recent survey show that these are the Wall Street Journal, The New York Times, USA Today and The Washington Post (Kiely, 2023). Hence, all headlines for these four newspapers were collected for the period 2017-12-05 - 2022-02-28. Here I use the headlines appearing on the front page of all editions of the daily newspapers. The *Sentiment* variable is constructed by first categorising the sentiment of each daily headline with FinBERT at the newspaper level, then forming daily newspaper averages for each of the four newspapers by the following formula:

$$S_{i,t} = \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} Sent_j, \quad (3)$$

where $N_{i,t}$ is the number of headlines in newspaper i at day t and $Sent_j$ is the FinBERT sentiment classification for each of the $N_{i,t}$ headlines. Following this, the daily sentiment index, as expressed by the news coverage, is created in (4) as the equally weighted average of the four newspaper sentiments:

$$Sentiment_t = \frac{1}{M} \sum_{i=1}^M S_{i,t}, \quad (4)$$

where M refer to the number of newspapers used to create the index, four in the present study. Figure 3 plots the resulting sentiment index. The spread of COVID-19 is clearly visible in the decline of sentiment during the first half of 2020.

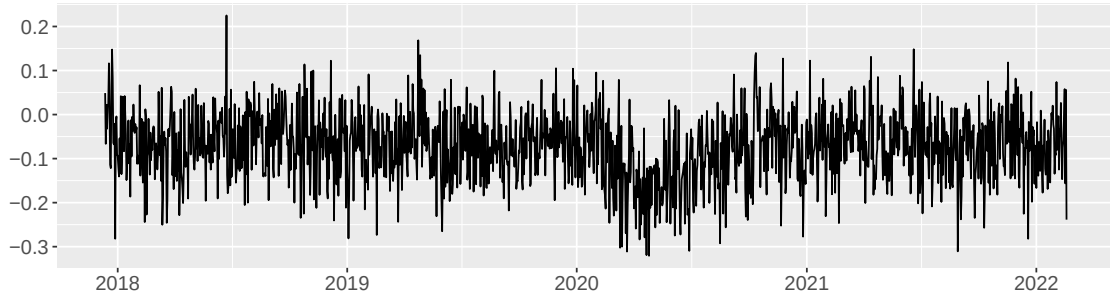


Figure 3: Sentiment Index during 2017-12-05 - 2022-02-28

Table 1 present descriptive statistics and unit root tests for the main variables. While the average value of $CSAD$ is 0.224, the median is 0.174, suggesting a right-skewed distribution for the measure of herding. *Sentiment* appears to be symmetric with the mean and median equal to -0.07 . The dispersion, on the other hand, is approximately proportional to the mean in both series. Two standard tests for unit root were conducted, the Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) test. Panel B report the results of these tests, which unanimously identify both variables as stationary. For descriptive statistics and unit root tests in COVID-19 and the pre-pandemic period, see Table 4 in Appendix B.

As discussed in Section 1, $CSAD$ is expected to be smaller when computed based

Table 1: Descriptive Statistics and Unit Root tests.

Panel A: Descriptive Statistics						
	N	Min	Median	Mean	St. Dev.	Max
<i>CSAD</i>	1,055	0.005	0.173	0.224	0.183	1.587
<i>Sentiment</i>	1,055	−0.319	−0.070	−0.069	0.075	0.169

Panel B: Unit Root tests				
	ADF(1)	ADF(2)	PP(1)	PP(2)
<i>CSAD</i>	−4.316***	−4.742***	−307.885***	−951.618***
<i>Sentiment</i>	−3.992***	−3.986***	−729.998***	−1169.780***

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ADF denotes the Augmented Dickey-Fuller test and PP the Phillips-Perron test for unit root. The tailing (1) and (2) refer to using a constant and trend in the test regression, respectively.

on indices compared to individual stocks. Comparing against the descriptive statistics reported in for example Table 1 in Chang et al. (2000), the numbers reported here are indeed smaller, although they suggest a similar shape of the *CSAD* distribution. Again, this means that the results from the subsequent analysis are conservative with regards to the magnitude of herding effects in the underlying securities.

3 Methodology

3.1 The Standard Approach

To detect herding behavior towards a market consensus, researchers commonly estimate the CCK model (Chang et al., 2000), which is an extension of the CH model of (Christie and Huang, 1995). The former improves on the bias that may arise in the latter due to outliers in the data, and is formulated as:

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \epsilon_t . \quad (5)$$

The CCK model in equation (5) is derived from the notion that rational asset pricing models predict a linear relation between market return and cross-sectional

absolute dispersion (Chang et al., 2000). Evidence from the behavioral finance literature suggests that extreme market conditions cause the dispersion between individual and market returns to decrease, or increase at a decreasing rate (Pochea et al., 2017). That is, the linear relation hypothesized in rational asset pricing models break down when market participants abandon their private information and instead follow aggregate market movements. Hence, a negative and statistically significant $\hat{\gamma}_2$ is empirical evidence of herding behavior. If, at the same time, $\hat{\gamma}_1$ is positive and significant, then the cross-sectional dispersion increase at a decreasing rate.

Since herding is measured in terms of *CSAD*, if the estimated parameter of an additional regressor included in (5), say $\hat{\gamma}_i$, is negative and significant, the regressor is interpreted as decreasing the dispersion and thus increasing the herding (Garcia, 2021). On the contrary, if $\hat{\gamma}_i$ is positive and significant the regressor is said to imply reverse herding. This refer to a situation in which there is less consensus on appropriate prices among investors where they, to a large extent trust, and act on, their own private information rather than abandoning it in favor of seemingly better informed investors (Choi and Yoon, 2020).

The CCK model is commonly estimated with OLS and HAC robust standard errors, see e.g. Chang et al. (2000); Chen (2013); Choi and Yoon (2020), and is often extended to include exogenous variables hypothesized by economic theory to influence herding. As seen in Section 1, much recent work has been devoted to understanding the impact of sentiment and COVID-19 on herding behaviour. In this case, the right-hand side of the CCK model is simply augmented with a measure of sentiment or COVID-19 interactions (see e.g. equation (7) in Choi and Yoon (2020) and equation (4) in Ampofo et al. (2023), respectively).

3.2 Modelling Dynamics and the Effect of Sentiment

From theory is it known that investor behavior is persistent, which here means that a period of high (low) levels of herding is likely followed by another period of high (low) herding levels. Empirical evidence of persistency in the CCK model is, however, scarce and the effect of lags remain largely unexplored in this framework. This is important

from both an econometric and a behavioral finance perspective as persistence is present in the data. The persistence of CSAD is evident from visual inspection of Panel A in Figure 2 and formally illustrated in Figure 4 by means of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the data.

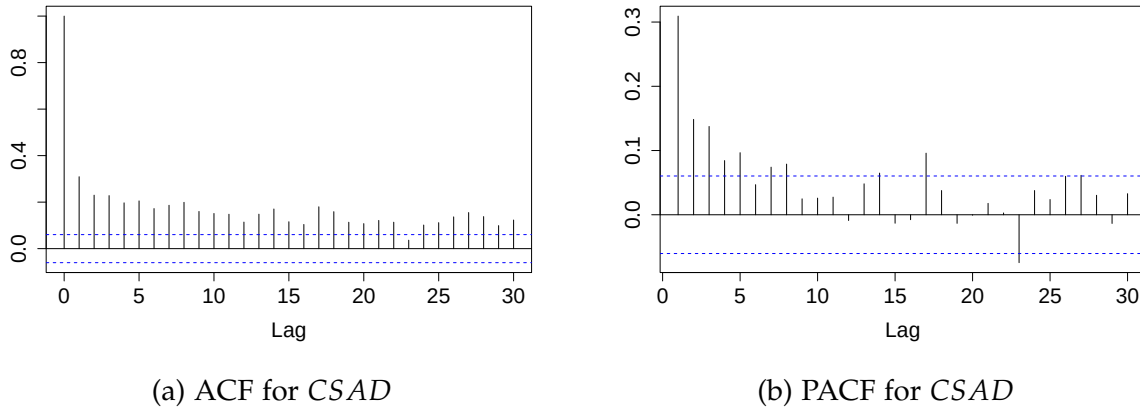


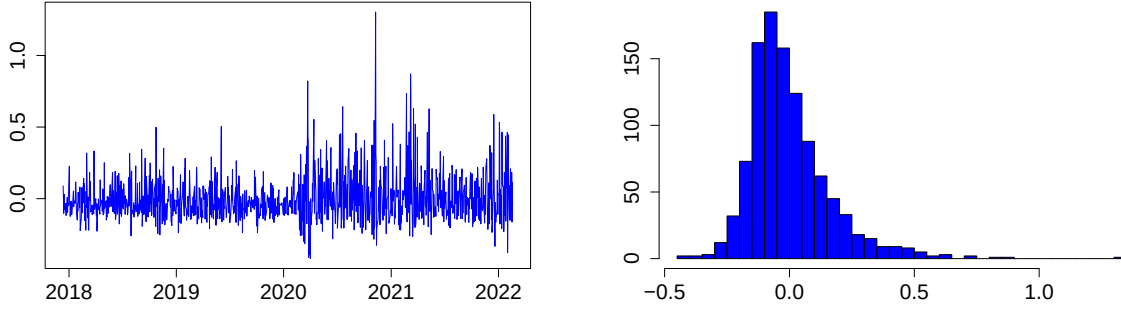
Figure 4: (Partial) Autocorrelation Functions for CSAD

A dynamic extension of the CCK model may thus, *ex ante*, provide a better fit. Hence, to model persistence in herding behavior and to capture how it changes with, for example, the spread of COVID-19, I extend the standard CCK model in equation (5) with lagged values in both *CSAD* and any exogenous variable in the model.

After using up to 5 lags of *CSAD* in the standard CCK model, which amounts to a business week worth of lags, there is still heteroscedasticity in the residuals from such an extended model. These residuals are plotted in Figure 5(a) which reveal a nontrivial dependence structure. Utilizing the conditional information in the time series data enables estimation of this structure.

Additionally, in Figure 5(b), the residuals appear right skewed rather than symmetric which indicates that the normal distribution may not be an appropriate distribution to impose when estimating these parameters. Indeed, when contrasting the empirical distribution of these residuals with theoretical quantities in Q-Q plots, the non-normality typically encountered in financial data is evident. Rather, a Skew Student-t distribution³ seems an appropriate approximation, see Figure 8 in Appendix A.

³I use the Skewed Student-t distribution proposed in Fernández and Steel (1998). The R package *rugarch* jointly estimates the shape and skewness parameters of this distribution with the other model parameters via maximum likelihood.



(a) Residuals over Time

(b) Histogram of Residuals

Figure 5: Visualization of Residuals from Equation (6)

Thus, to investigate herding and the role of sentiment, the following model is used:

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \sum_{i=1}^k \theta_i \text{Sentiment}_{t-i+1} + \sum_{i=1}^l \psi_i CSAD_{t-i} + \epsilon_t, \quad (6)$$

$$\text{var}(\epsilon_t | \mathcal{I}_{t-1}) = h_t = \omega + \sum_{p=1}^P \alpha_p \epsilon_{t-p}^2 + \sum_{q=1}^Q \beta_q h_{t-q}, \quad (7)$$

where $\mathbb{E}[\epsilon_t] = 0$ and a Skew Student-t distribution is imposed on the innovations. In addition to modeling the intra-week persistency in herding, my dynamic specification allows for empirical analysis of short- and long-run effects. This is not possible in the static CCK model. Long-run effects refer here to long-run multipliers, defined as:

$$\hat{\delta} = \frac{\sum_{i=1}^k \hat{\theta}_i}{1 - \sum_{j=1}^l \hat{\psi}_j}.$$

For estimating the mean and conditional heteroscedasticity equations in (6) and (7) respectively, I use the Maximum Likelihood Estimator (MLE). The motivation for this approach is twofold: first, since there is evidence of heteroscedasticity in the residuals of (6), a MLE that accounts for (7) is expected to be more efficient than OLS based on (6) alone. That is, the dependence structure in the residuals is information that can be used to improve estimation, see Section 4 for Monte Carlo evidence. The relative gains in efficiency of MLE over OLS in settings with dynamics in the data has been well documented (see e.g. Engle, 1982; Pantula, 1988; Iglesias and Phillips, 2001).

Second, since there is considerable autocorrelation and heteroscedasticity in the

data, fitting the standard CCK model in (5) with OLS require the use of HAC robust standard errors. Although defaulting to HAC in the face of autocorrelated and/or heteroscedastic data is standard practice, such standard errors typically have poor finite-sample properties. The reason is that, to achieve consistency, the standard asymptotic theory for HAC assumes that the lag truncation grows slower than the sample size. In practice, this condition is never met (Neave, 1970), leading to oversized test statistics. While this is well documented for (zero-) mean processes with ARMA(1,1) errors, see e.g. Jansson (2004); Kiefer and Vogelsang (2005); Sun et al. (2008), Section 4.2 presents results from a Monte Carlo based on the more relevant AR-GARCH family. The simulation study suggest that the problem caused by HAC can be avoided by using lags to capture persistence in the mean equation and by accounting for the GARCH-structure present in the innovations. Although strategies for improving the performance of HAC robust tests exists, such as using the fixed-b asymptotic framework, I opt to adjust the model rather than the HAC estimator as the former allows me to capture the dynamics which influence herding behavior.

In the mean equation in my model, I set $k = 5$ and $l = 6$ to allow a window of observations corresponding to the business week to influence the time t herding. This is in line with e.g. Kim et al. (2022) who allow sentiment information of one week in their model. For the GARCH terms, I minimize the Bayesian-Schwartz Information Criterion (BIC) over a grid search of various combinations (p, q) for a $GARCH(p, q)$ ⁴.

3.3 Herding before and during COVID-19

As discussed in Section 1, much attention has been devoted to understanding the effect of COVID-19 on herding. The standard approach to account for COVID-19 in the literature is to include a COVID-19 dummy on each slope variable in the CCK model, see Espinosa-Méndez and Arias (2021); Mandaci and Cagli (2022); Ampofo et al. (2023). Additionally, visual inspection of the data in Figure 2(a) motivates effectively splitting the sample at the onset of COVID-19. Therefore, to explore the effect of COVID-19 on

⁴I use a maximum lag length of 10 for both p and q and test all possible combinations (p, q) .

herding in a dynamic setting, I extend model (6) accordingly and estimate:

$$\begin{aligned}
CSAD_t = & \gamma_0 + \gamma_1^{Covid} D \cdot |R_{m,t}| + \gamma_2^{Covid} D \cdot R_{m,t}^2 + \sum_{i=1}^k \theta_i^{Covid} D \cdot Sentiment_{t-i+1} \\
& + \sum_{i=1}^l \psi_i^{Covid} D \cdot CSAD_{t-i} + \gamma_1^{NoCovid} (1-D) \cdot |R_{m,t}| + \gamma_2^{NoCovid} (1-D) \cdot R_{m,t}^2 \\
& + \sum_{i=1}^k \theta_i^{NoCovid} (1-D) \cdot Sentiment_{t-i+1} + \sum_{i=1}^l \psi_i^{NoCovid} (1-D) \cdot CSAD_{t-i} + \epsilon_t \quad (8)
\end{aligned}$$

$$\text{var}(\epsilon_t | \mathcal{I}_{t-1}) = h_t = \omega + \sum_{p=1}^P \alpha_p \epsilon_{t-p}^2 + \sum_{q=1}^Q \beta_q h_{t-q}, \quad (9)$$

where $\mathbb{E}[\epsilon_t] = 0$ and parameters (k, l, p, q) are as in Section 3.2. In model (8), $D = 1$ for the days on and after the 20th of January 2020, the date of the first reported confirmed US case of COVID-19, and 0 for all observations preceding it.

4 Monte Carlo Simulation

As discussed in Section 3.2, estimating my dynamic extension of the standard CCK model with MLE offer advantages in terms of efficiency and size control over OLS and HAC on the static model. This section presents a Monte Carlo which illustrates these advantages. More concretely, I show the consequences in terms of efficiency and test size of estimating a static model with HAC robust standard errors rather than modeling the dynamics, when dynamics exist in the data. Thus, from an econometric point of view, motivating my extension of the standard CCK model.

Since the point I am trying to make here in the Monte Carlo is a general one, I will not limit myself to the specific model considered earlier. Thus, I disregard exogenous regressors and focus on estimation of the mean parameter. To mimic the empirical data, I make draws from a process with characteristics similar to that described in Section 2 and 3. That is, a process which depend on both autoregressive and heteroscedastic components. Specifically, I assume the following data generating process (DGP):

$$y_t = \mu + u_t, \quad (A1)$$

$$u_t = \gamma u_{t-1} + \epsilon_t = \gamma u_{t-1} + v_t \sqrt{h_t}, \quad v_t \sim \mathcal{N}(0, 1), \quad (A2)$$

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \quad (A3)$$

The *static model* is, for all simulations, an estimation of (A1) with OLS and HAC robust standard errors while the *dynamic model* is a Maximum Likelihood estimation of (A1)-(A3). When computing the HAC standard errors I follow the convention in the literature on herding and use the standard Newey-West, or Bartlett, kernel.

In this section, I use $T = 300$ which is comparable to samples often used in empirical papers studying the effect of crisis periods on herding, see Appendix E. Further, when not otherwise stated, I use the rule of thumb $M = 0.75T^{1/3}$ (Stock and Watson, 2011) for the HAC bandwidth and $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.1, 0.8)$ in the GARCH specification. For a more detailed description of tests, results and the robustness of these results under alternative specifications and sample sizes, see Appendix D.

4.1 Efficiency

It is well known that the Maximum Likelihood estimator achieves the Cramer-Rao lower bound, i.e. estimating a MLE on both the mean and variance equations (A1)-(A3) is expected to be more efficient than an OLS on the mean equation alone. This can be illustrated by estimating μ with both approaches and comparing the variance of the point estimates. More concretely, by comparing $\mathbb{V}(\hat{\mu}^{OLS})$ with $\mathbb{V}(\hat{\mu}^{MLE})$ where $\mathbb{V}(\cdot)$ denotes the sample variance defined as

$$\mathbb{V}(\hat{\mu}^j) = \frac{1}{N-1} \sum_{n=1}^N (\hat{\mu}_n^j - \tilde{\mu}^j)^2, \quad j \in \{OLS, MLE\}, \quad (A4)$$

where $\hat{\mu}_n^j$ refer the point estimate from the n^{th} MC replication estimated with $j \in \{OLS, MLE\}$ while $\tilde{\mu}^j$ is the average of those estimates of the mean over all N MC replications. Below I plot $\mathbb{V}(\hat{\mu}^{OLS})$, $\mathbb{V}(\hat{\mu}^{MLE})$ and the relative efficiency of MLE for

$\gamma \in [0.1, 0.9]$ and $\mu = 0$. The relative efficiency of MLE, RE_{MLE} , is defined as the ratio between $\mathbb{V}(\hat{\mu}^{MLE})$ and $\mathbb{V}(\hat{\mu}^{OLS})$ where $RE_{MLE} > 1$ means that MLE is more efficient.

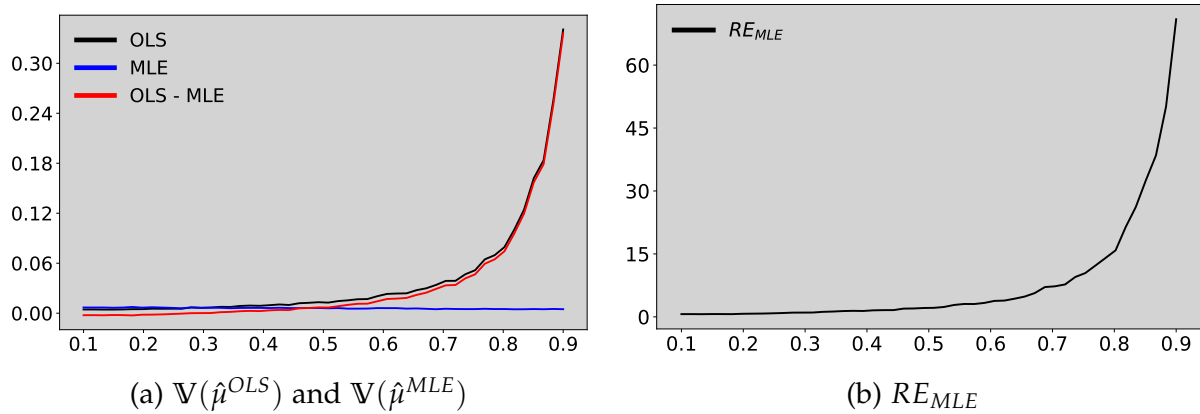


Figure 6: Efficiency of MLE and OLS over a Grid of Autoregressive Parameters.

From Figure 6 it is clear that for a fixed bandwidth on the HAC standard errors, the efficiency gain from estimating μ with MLE increases with γ . Further, it is negligible when γ is small and sizable even when the degree of autocorrelation is moderate. For example, in the simulation, $RE_{MLE} \approx 2.27$ when $\gamma = 0.5$, i.e. the variance in $\hat{\mu}$ is more than twice as large when estimated with OLS compared to MLE. In general, the relative efficiency of MLE grows with the sample size, $RE_{MLE} \approx 2.38$ when $\gamma = 0.5$ for $T = 750$. This is also true for alternative GARCH specifications, see Appendix D.2.

So, in terms of efficiency, a MLE on both the mean and variance equation offer advantages over OLS with HAC standard errors when dynamics exist in the data.

4.2 Oversized Tests

This section considers test size, which is the probability of falsely rejecting a null hypothesis. To assess the size of the HAC robust tests, I borrow the setup from Kiefer and Vogelsang (2005). That is, I generate data according to (A1)-(A3) with $\mu = 0$ and compute rejection probabilities for the null hypothesis $H_0 : \mu \leq 0$ against the alternative $H_1 : \mu > 0$ using the standard HAC robust t-statistic for μ :

$$t_{HAC} = \frac{\sqrt{T}(\hat{\mu} - \mu)}{\sqrt{\hat{\sigma}_{HAC}^2}}, \quad (A5)$$

for all possible bandwidths. I use the standard normal critical value 1.645 for all simulations. Note that a two-sided test is qualitatively similar since the distribution of the test statistic is symmetrical, see Kiefer and Vogelsang (2005) for additional details.

To assess size of the MLE tests, I use the same null and alternative as defined above and compute rejection probabilities using the standard t statistics for $\hat{\mu}$ from a MLE on (A1)-(A3). Again, specifying $T = 300$ and $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.1, 0.8)$. For larger samples and alternative specifications on the GARCH-parameters, see Appendix D.3.

Figure 7 illustrates the rejection rates from both OLS with HAC corrections and MLE for the true null hypothesis for a variety of values on the autoregressive parameter. The HAC rejection rates are depicted as solid blue lines while the MLE rejections are plotted in orange. The nominal 5% level is indicated with the dashed black line.

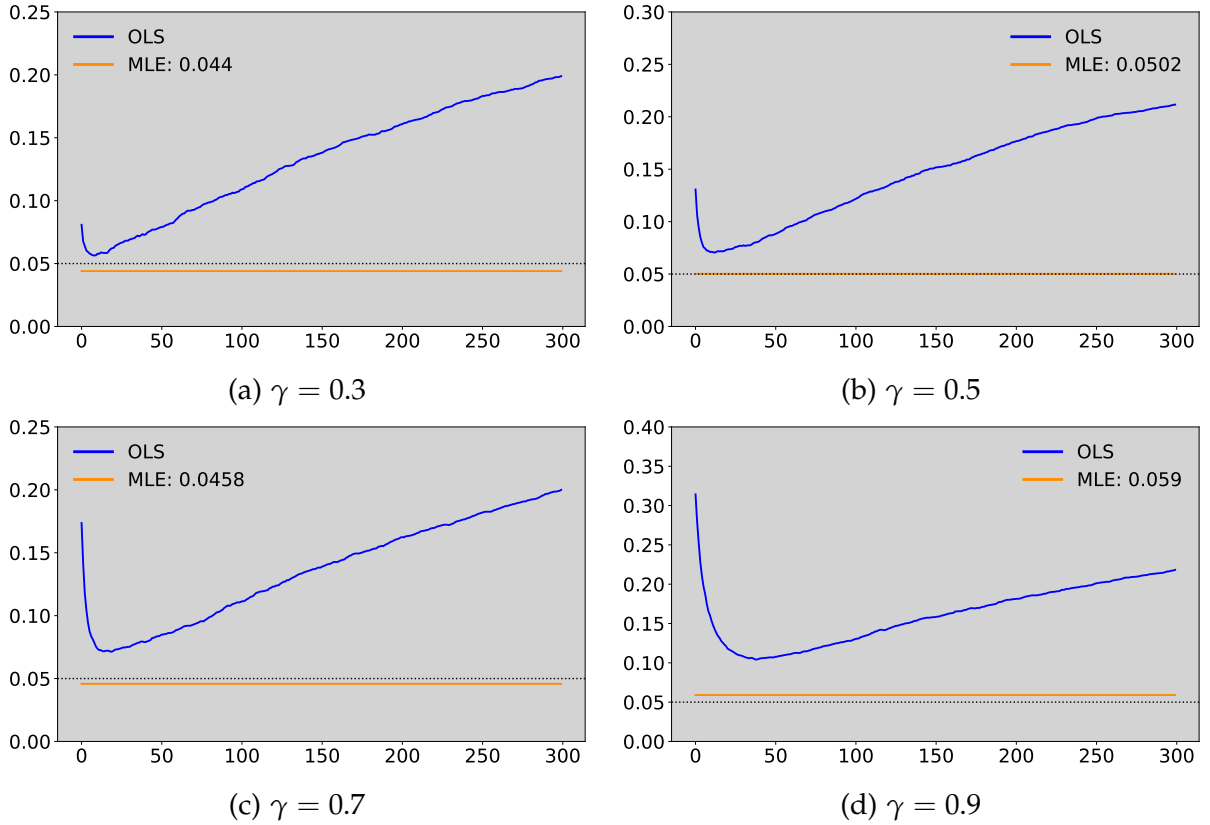


Figure 7: Rejection Rates over the Bandwidth.

It is clear that the degree with which the HAC robust tests overreject the null depend both on the autoregressive parameter and the bandwidth. Further, for each value γ , the overrejection rates vary nonlinearly with the bandwidth. For a larger

autoregressive parameter, the more the rejection rates are characterized by a L-shape. The reason for this pattern is a nontrivial trade-off between bias and variance in $\hat{\sigma}_{HAC}^2$ that depends on the bandwidth, M , see Appendix D.3.2 for an explanation.

Additionally, the larger the autoregressive parameter the more is the graph shifted northeast, note the scale in the y-axis in Figure 7(a)-7(d). This means that the performance of HAC robust tests deteriorates with increases in γ , for all choices of the bandwidth parameter. For example, when $\gamma = 0.9$, HAC rejection rates vary between approximately 10% and 30%. That is, for this level of persistency, there is at minimum a 10% probability of rejecting the true null, twice the nominal level. Note that even for the less persistent cases, as seen in Figure 7(a) and 7(b), HAC robust tests produce rejection rates that exceed the nominal 5% level for all bandwidths. With the bandwidth being chosen with the rule of thumb, the rejection rates are 5.92%, 7.98%, 9.44% and 21.04% for $\gamma \in \{0.3, 0.5, 0.7, 0.9\}$, respectively. For a more detailed representation of these results, see Table 6 in Appendix D.3.2.

In contrast, the rejection rates from the MLE on (A1)-(A3) are suggestive of size control for all values γ , as expected. While the MLE slightly overreject compared to the nominal 5% level for the most persistent case in this sample of $T = 300$, the issue vanishes asymptotically, see Table 6 in Appendix D.3.2. In that section, it is evident that the problem for HAC robust tests are still present in larger samples.

The results presented in this section are also robust to alternative GARCH specifications, see Appendix D.3.3. Together, this indicates that the issue with the HAC robust tests is the presence of autoregressive rather than heteroscedastic components which again underlines the preference for using a model and estimator that accounts for dynamics in the data, when such dynamics exist.

5 Empirical Results

5.1 Herding over the entire sample

Table 2 reports the estimates from the dynamic CCK model (MLE on equation (6) and (7)). BIC was used to select $GARCH(1,1)$ as the model for conditional heteroscedasticity.

Results from an OLS on equation (6) only are reported in column 5.

A negative and statistically significant estimate of γ_2 is evidence of herding in the US market since this corresponds to a breakdown of the linear relationship between dispersion and returns in the CAPM-framework, as argued in Chang et al. (2000). Since herding is measured in terms of dispersion around the market return, a negative and significant coefficient of any additional regressor is interpreted as decreasing the dispersion and thus increasing the herding. On the contrary, a positive significant coefficient increases the dispersion which is interpreted as a variable that causes reverse herding.

Table 2: Estimates of US herding. Sample period is 2017-12-12 - 2022-02-18.

Parameter	Variable	MLE	MLE SE	OLS
γ_0	<i>Intercept</i>	0.196***	0.0103	0.074***
γ_1	$ R_{m,t} $	0.023***	0.0072	0.059***
γ_2	R_{mt}^2	-0.003***	0.0010	-0.005
ψ_1	$CSAD_{t-1}$	0.100***	0.0222	0.194***
ψ_2	$CSAD_{t-2}$	0.049**	0.0217	0.070*
ψ_3	$CSAD_{t-3}$	0.053***	0.0199	0.094*
ψ_4	$CSAD_{t-4}$	0.048***	0.0178	0.047
ψ_5	$CSAD_{t-5}$	0.055***	0.0168	0.090**
θ_1	<i>Sentiment_t</i>	0.039	0.0398	0.164**
θ_2	<i>Sentiment_{t-1}</i>	-0.040	0.0369	0.002
θ_3	<i>Sentiment_{t-2}</i>	0.007	0.0403	-0.076
θ_4	<i>Sentiment_{t-3}</i>	-0.003	0.0340	-0.018
θ_5	<i>Sentiment_{t-4}</i>	-0.016	0.0345	-0.044
θ_6	<i>Sentiment_{t-5}</i>	0.036	0.0331	0.007
ω	<i>Intercept</i>	0.004**	0.0019	NA
α_1	ϵ_{t-1}	0.053***	0.0121	NA
β_1	h_{t-1}	0.789***	0.0818	NA

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

This table presents MLE and OLS-HAC estimation results of model (6), the dynamic extension of the standard CCK model, augmented with news sentiment. A Skew Student-t is used as conditional distribution for the innovations.

As $\hat{\gamma}_2$ is negative and significant, there is evidence of herding towards the market consensus in the US data. Since $\hat{\gamma}_1$ is positive and significant it means that dispersion increased at a decreasing rate over the sample period. This result is consistent with for

example Chen (2013) and the view that the periods of elevated market stress induces herding, see Section 5.2, although found in a model able to capture the dynamics in which herding occur. Persistency of dispersion is important since all autoregressive components ($\hat{\psi}_1 - \hat{\psi}_5$) are highly significant and including them in the model results in an estimate of γ_2 half in magnitude compared to the estimate reported in Ampofo et al. (2023). Further, this finding suggest that the mechanisms proposed to drive herding behavior indeed should be considered dynamic rather than static phenomena.

For comparison, column 5 in the above table report the results from an OLS on the mean equation of the extended CCK model, equation (6), dealing with the dependence in the data by imposing HAC robust standard errors. This correspond to simply defaulting to the standard estimation approach outlined in Section 3.1. With this strategy, $\hat{\gamma}_2$ is insignificant meaning that there is no evidence of herding. This result is in contrast to the clear evidence of herding at the index level found when estimating the model with MLE making use of the conditional information available in the data, as well as the results found in the limited existing literature. This is likely due to the inefficiency of OLS relative to MLE when dynamics exist in the data, see Section 4.1.

The null hypothesis of no effect cannot be rejected for any of the parameters that capture news sentiment. That is, when allowing parameters that capture the persistence in return dispersion in the model, and making use of conditional information, news sentiment has no additional effect on herding in this sample. Recall that the HAC robust tests was shown to overreject a true null while the MLE did not, see Section 4.2. Hence, the positive and significant $\hat{\theta}_1$ presented in column 5 should be interpreted carefully. This means that positive sentiment likely did not induce reverse herding during the sample, rather there was no additional effect of sentiment on herding after controlling for lags in return dispersion and using the conditional information. The reason for this is likely related to forcing the sentiment dynamics before, and after, the spread of COVID-19 to be captured by a single set of coefficients although the relationship between sentiment and herding may be different in the regimes, see figure 2(a). More on this in Section 5.2.

As discussed in earlier sections, the model (6) specification allow for estimation of long-run effects, in addition to the short-run effects listed in Table 2. However,

since there are no short-run effects of sentiment on herding, there can be no long-run effects. Further, from Table 2, the sum of the GARCH components is well below unity, indicating a stable conditional heteroscedasticity process.

5.2 Herding during COVID-19

The results of model (8) and (9) are presented in Table 3. For this specification, the error structure is parameterized by a $GARCH(2,2)$ where, again, the pair (p,q) was chosen by minimizing the BIC over a grid search of possible combinations. For comparison, results obtained with the standard OLS and HAC approach are reported in column 5.

When allowing a different relationship between the variables in the period before and during COVID-19, much of the results presented in Section 5.1 are only valid for the COVID-19 period. This showcase the importance of accounting for possible exogenous shifts in the data, such as the COVID-19 period.

As in the previous subsection, there is evidence of herding in the US market, however, only in the COVID-19 period as the MLE of γ_2^{Covid} is negative and significant while that of $\gamma_2^{NoCovid}$ is not. This result is consistent with the literature, see for example Duygun et al. (2021); Espinosa-Méndez and Arias (2021). Also, dispersion increase at a decreasing rate during COVID-19 while there is no such tendency in the pre-COVID part of the sample. Similar to the full sample specification, an OLS with HAC standard errors on the mean equation alone finds no evidence of herding as both $\hat{\gamma}_2^{Covid}$ and $\hat{\gamma}_2^{NoCovid}$ in column 5 are insignificant. Since $\hat{\gamma}_2^{Covid}$ has the same sign regardless of the estimator, the insignificance in column 5 may be related to efficiency, as discussed in Section 5.1. Further, the MLE and OLS estimate of γ_2^{Covid} are of similar magnitude here as in the full sample specification. This again suggest that the COVID-19 period seems to drive the results presented in Table 2.

Along the same lines, the autoregressive features of return dispersion discussed in Section 5.1 are only present during COVID-19 as $\hat{\psi}_1^{Covid}$, $\hat{\psi}_3^{Covid}$, $\hat{\psi}_4^{Covid}$ and $\hat{\psi}_5^{Covid}$ are all positive, significant and of similar magnitude to the estimates presented in Table 2. During COVID-19, return dispersion is thus persistent and high (low) levels of $CSAD$ feed into future high (low) levels of $CSAD$ via positive coefficients. No such evidence

Table 3: Estimates of US herding before and during Covid-19. Sample period is 2017-12-12 - 2022-02-18.

Parameter	Variable	MLE	MLE SE	OLS
γ_0	<i>Intercept</i>	0.143***	0.0219	0.114***
γ_1^{Covid}	$D * R_{m,t} $	0.017***	0.0028	0.060***
γ_2^{Covid}	$D * R_{m,t}^2$	-0.002***	0.0004	-0.005
ψ_1^{Covid}	$D * CSAD_{t-1}$	0.122***	0.0278	0.198***
ψ_2^{Covid}	$D * CSAD_{t-2}$	-0.021	0.0283	0.041
ψ_3^{Covid}	$D * CSAD_{t-3}$	0.057**	0.0242	0.095
ψ_4^{Covid}	$D * CSAD_{t-4}$	0.082***	0.0237	0.028
ψ_5^{Covid}	$D * CSAD_{t-5}$	0.083***	0.0201	0.075
θ_1^{Covid}	$D * Sentiment_t$	0.147***	0.0428	0.348***
θ_2^{Covid}	$D * Sentiment_{t-1}$	0.048	0.0536	0.047
θ_3^{Covid}	$D * Sentiment_{t-2}$	-0.067	0.0471	-0.194
θ_4^{Covid}	$D * Sentiment_{t-3}$	-0.028	0.0517	0.031
θ_5^{Covid}	$D * Sentiment_{t-4}$	0.001	0.0459	-0.111
θ_6^{Covid}	$D * Sentiment_{t-5}$	-0.036	0.0528	-0.057
$\gamma_1^{NoCovid}$	$(1 - D) * R_{m,t} $	0.008	0.0107	0.055***
$\gamma_2^{NoCovid}$	$(1 - D) * R_{m,t}^2$	0.005	0.0028	-0.001
$\psi_1^{NoCovid}$	$(1 - D) * CSAD_{t-1}$	0.061	0.0483	0.074
$\psi_2^{NoCovid}$	$(1 - D) * CSAD_{t-2}$	0.030	0.0319	0.067
$\psi_3^{NoCovid}$	$(1 - D) * CSAD_{t-3}$	0.047	0.0404	0.005
$\psi_4^{NoCovid}$	$(1 - D) * CSAD_{t-4}$	0.034	0.0272	0.012
$\psi_5^{NoCovid}$	$(1 - D) * CSAD_{t-5}$	-0.004	0.0311	0.059
$\theta_1^{NoCovid}$	$(1 - D) * Sentiment_t$	-0.087*	0.0464	0.039
$\theta_2^{NoCovid}$	$(1 - D) * Sentiment_{t-1}$	-0.076	0.0475	0.050
$\theta_3^{NoCovid}$	$(1 - D) * Sentiment_{t-2}$	0.026	0.0455	0.065
$\theta_4^{NoCovid}$	$(1 - D) * Sentiment_{t-3}$	0.009	0.0385	-0.006
$\theta_5^{NoCovid}$	$(1 - D) * Sentiment_{t-4}$	0.019	0.0447	0.052
$\theta_6^{NoCovid}$	$(1 - D) * Sentiment_{t-5}$	0.079**	0.0383	0.120
ω	<i>Intercept</i>	0.002***	0.0004	NA
α_1	ϵ_{t-1}^2	0.052***	0.0110	NA
α_2	ϵ_{t-2}^2	0.000	0.0169	NA
β_1	h_{t-1}	0.006	0.0327	NA
β_2	h_{t-2}	0.873***	0.0157	NA

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

This table presents MLE and OLS-HAC estimation results of model (8), the dynamic extension of the COVID-19 CCK model, augmented with news sentiment. A Skew Student-t is used as conditional distribution for the innovations.

is found for the pre-COVID-19 period. Importantly, this may shed a different light on the mechanisms that influence herding behavior in the US market than was revealed in the previous section. In section 5.1, no distinction between the period before and during COVID-19 was made and hence, persistence in *CSAD* could be attributed to mechanisms such as the structure of investor reputation and compensation schemes. Recall that these mechanisms refer to investors being constantly evaluated against one another and compensated based on their relative performance making them inherently motivated to not deviate too far from the market consensus (Scharfstein and Stein, 1990; Lux, 1995). As there is little reason to expect the inner dynamics of the investor market to have changed over the course of this sample, this suggest that the results in Table 3 are more likely driven by investors disregarding their own private information in the face of the Black Swan event, COVID-19. Although a statistical test of such an hypothesis is however not feasible in model (8)-(9), this is consistent with the view that fear and uncertainty influence herding (Pochea et al., 2017; Economou et al., 2018). Much of the *CSAD* dynamics, and thus the insights into the mechanisms influencing herd behavior, are lost when estimating the model without the use of conditional information and appropriate estimation strategy, see column 5.

While the model that effectively pooled the effects across the entire sample revealed no effects of sentiment on herding, allowing for the relationship to differ between the pre-COVID period and the COVID-19 period indicate that sentiment indeed is influential. The results presented in Table 3 show that sentiment has different effects in the different regimes as $\hat{\theta}_1^{Covid}$ and $\hat{\theta}_6^{NoCovid}$ are significantly positive while $\hat{\theta}_1^{NoCovid}$ is significantly negative. This underlying heterogeneity of effects with respect to COVID-19 is missed in the specification discussed in Section 5.1. This likely drives the finding of no sentiment effect discussed there as $\hat{\theta}_1^{NoCovid}$ and $\hat{\theta}_6^{NoCovid}$ are both significant, similar in magnitude but of opposite sign.

The significant effect of sentiment on herding is consistent with the view that investors abandon their private information and follow seemingly better informed investors during periods of elevated fear and information cascades (Bikhchandani and Sharma, 2000; Economou et al., 2018; Ferreruella and Mallor, 2021) as news media is a platform through which such information is transmitted.

As $\hat{\theta}_1^{Covid}$ is positive and significant, positive sentiment had an immediate effect and caused reverse herding during COVID-19. In this regime, positive sentiment was transferred into higher levels of return dispersion, resulting in a decrease of herding.

In the period before COVID-19, the dynamics were different. As $\hat{\theta}_1^{NoCovid}$ is negative and significant, positive news before COVID-19 had the immediate effect of decreasing the return dispersion, indicating a tendency of investors trading towards the market consensus although no evidence of herding is found. In contrast, negative news meant an increase in return dispersion in this period. This suggest that investors, before COVID-19 responded to positive news by moving positions towards the market consensus while, in the face of negative news, this was not the case. Facing negative news, investors were more resilient, holding to their private information to a larger extent before COVID-19, compared to what they did during COVID-19. During the pandemic, investors rather held on to their private information in the face of positive news and abandoned it to follow seemingly more informed investors during episodes of negative sentiment. This is not surprising as the pandemic meant a shift for the content traditionally published in the news as this media largely became a channel through which public health and policy communication was transmitted (Mach et al., 2021). Moreover, the significantly positive estimate of $\theta_6^{NoCovid}$ further highlights the different impact newspaper sentiment has on herding in the two regimes. During COVID-19, the effect of newspaper sentiment is quickly absorbed into return dispersion since it is only the sentiment parameter capturing the contemporaneous effect that is significant. In the period before the COVID-19 outbreak, however, newspaper sentiment has both a contemporaneous effect of opposite sign compared to the COVID-period, as well as an effect with five periods lag of similar magnitude to the contemporaneous effect. This result suggest that people payed closer attention to, and acted on, the sentiment conveyed by news media during the COVID-19 period than before.

Regarding the GARCH model, $\hat{\alpha}_1$ and $\hat{\beta}_2$ are significant. Their sum is below the stationarity condition which imply a stable conditional heteroscedasticity process.

In contrast to the previous section, as there are short-run effects of news sentiment on herding, long-run effects of this relation can be tested. The estimated long-run

multiplier is 0.0436 and -0.0140 during and before COVID-19, respectively. However, the null of no long-run effect could not be rejected, see Appendix C for more details.

6 Discussion and Conclusion

This paper empirically investigate herding towards the market consensus and the effect of news sentiment and the COVID-19 pandemic on this behavior in the US. The dataset used throughout this study span from the 12th of December 2017 to the 18th of February 2022 and includes daily price data for three major US stock indexes; the S&P500, the Dow Jones Industrial Average and the NASDAQ Composite. FinBERT, a Large Language Model adapted to the financial domain, was used to classify the sentiment of headlines published in the four largest US daily newspapers; the Wall Street Journal, the New York Times, USA Today and the Washington Post. The data was downloaded from Capital IQ and Global Newsstream, respectively.

In this study, I compute CSAD based on major stock indices on which many ETFs and mutual funds have been issued. This allows me to proxy trading in the funds instruments themselves which represents an increasingly popular trading strategy among both institutional and retail investors. It has been shown that trading and demand shocks in the overarching assets propagate to trading and demand shocks in the underlying securities. This propagation is especially strong during events of market stress. Since market stress is a factor often related to herding behavior, the propagation channel is interesting from a herding perspective. This since a demand shock induced by market stress on the mutual funds and ETF market is expected to propagate to a demand shock in the underlying assets, it may, as documented in Section 5 of this paper, induce herding in the underlying broad indices directly.

I further contribute to the existing literature by extending the standard static CCK model to include lags of CSAD and news sentiment. Exploratory analysis suggest the presence of heteroscedasticity in the residuals of such a model, hence the conditional information in the time series is used for estimating the GARCH-structure. This allows me to model the persistency in behavior hypothesized in theory and present in the data. Although behavioral biases such as habit formation and investors

commonly loading on recently well performing stocks (Muñoz, 2006; Barber et al., 2009) are established concepts, empirical evidence of persistency is scarce and the effect of dynamic components remain largely unexplored in the CCK framework. Such modeling is important as the mechanisms suggested in the literature to drive herding behavior should arguably be considered dynamic, not static. To the best of my knowledge, this has not been done in the standard single-equation CCK model before.

In addition to the conceptual reasons for modeling the persistency in herding behavior, as expressed in the behavioral finance literature, there are also econometric reasons for considering a dynamic rather than a static model when dynamics exist in the data. First, a model that accounts for the dynamic structure in the data is shown, in a Monte Carlo, to be more efficient than a model that does not. This is true for a range of alternative GARCH specifications even when the degree of autocorrelation is moderate, see Section 4. This means that estimating (6)-(7) with MLE rather than (5) with OLS and HAC robust standard errors is expected to be more efficient as the former makes use of the information available in the residuals of the latter. Secondly, and more importantly, HAC robust tests in a static model are shown to cause size distortions when autocorrelation and heteroscedasticity exist in the data. This holds for all levels of autocorrelation, GARCH specifications, and sample sizes considered and is, *ceteris paribus*, more pronounced the stronger the autocorrelation. Further, Monte Carlo results suggest that the problem with size distortion in HAC robust tests can be avoided by using lags to capture autocorrelation in the mean equation and by accounting for conditional heteroscedasticity in the variance equation. This means that a MLE on my dynamic extension of the standard CCK model is expected to produce tests of correct size relative to the standard CCK model estimation procedure.

The empirical results indicate that herding is found in the full sample (12th December 2017 to 18th February 2022) and during COVID-19 (20th January 2020 to 18th February 2022) in the US. No such evidence is found for the period before the COVID-19 pandemic (12th December 2017 to 19th January 2020). While these results are consistent with those in for example Chen (2013); Duygun et al. (2021); Ampofo et al. (2023), they also highlight that the dynamic components are an important addition to the model as including these parameters has a significant impact on the

standard regressors in the CSAD framework. For example, Table 2 report an estimate of γ_2 half in magnitude compared to the corresponding estimate reported in Ampofo et al. (2023). Recall that since indices behave like well diversified portfolios where extreme returns cancel out on normal days, computing CSAD based on index returns is expected to yield smaller cross-sectional dispersion compared to calculations based on individual stocks. Any evidence for herding in indices should hence be regarded as conservative with respect to the herding in the underlying securities. Table 2 further reveal that all autoregressive components of *CSAD* ($\hat{\psi}_1 - \hat{\psi}_5$) are highly significant, thus that the mechanisms proposed to drive herding should be considered dynamic and not static. Employing the standard estimation strategy, OLS with HAC, reveal no evidence for herding in this specification which is in contrast to the existing literature.

For the second part of this study, I empirically test the effect of COVID-19 on herding. These results are presented in Table 3 and show that the persistency of *CSAD* is present only in the COVID-19 period. Indeed, much of the results presented in Table 2 hold only for this period. This result is important as it may help shed light on the mechanisms influencing herding behavior. In the first specification, the persistence in *CSAD* could be attributed to mechanisms such as the structure of investor compensation and a disregarding of private information in a period of elevated fear. Since investors are commonly evaluated against their peers and compensated based on relative performance, they are inherently motivated not to deviate too far from the market consensus (Scharfstein and Stein, 1990; Lux, 1995). However, there is little reason to expect dramatic changes in the inner dynamics of investor markets during this period which would suggest that herding in this period is more likely driven by investors disregarding their private information in the face of the Black Swan event, COVID-19. While this is not statistically testable in the current model which warrant further research, this view is supported by the result that negative sentiment contemporaneously increase herding during COVID-19, $\hat{\theta}_1^{Covid}$ is positive. This can be explained by the shift in content traditionally transmitted via the news following the spread of the pandemic as news became an important channel through which public health and policy communication was broadcast (Mach et al., 2021). This result is of importance for investors and policymakers as herding pose

challenges related to excess volatility and market inefficiency and may even lead to market bubbles and crashes that may further amplify the negative consequences of pandemics on society.

Including dynamic components in the model further reveal differences with which investors are influenced by sentiment in the COVID-19 period *viz a viz* the period prior to the pandemic. Although no sign of herding is detected before COVID-19, positive news sentiment is shown to contemporaneously decrease return dispersion during this period, an effect that is reversed during a business week. During COVID-19 however, evidence suggests that investors paid more attention to, and acted on, the sentiment conveyed by news media as this signal is immediately absorbed into return dispersion. Here, negative news is found to decrease return dispersion, thus to increase herding. This suggests that investors, during COVID-19, herd on negative news while they held on to private information in the face of positive news. The opposite reaction to news sentiment is found for the period before the COVID-19 pandemic.

While the results discussed in this section would suggest that herding in the sample was likely driven by a fear factor prompting investors to disregard private information to follow seemingly better informed investors, however convincing, this hypothesis cannot be tested in the dynamic CCK model which motivates further study. In addition to disentangling the mechanisms driving herding, future research may continue studying the various insights that may be drawn from modeling herding behavior as a dynamic, rather than static phenomenon. For example, an investigation of how the increased geopolitical tensions, caused by the Russia-Ukraine and Israel-Palestine conflicts, influence herding dynamics may provide insight into potential market reactions in the wake of a potentially escalating situation in the Middle East.

Appendices

A Empirical Residual Distribution

The normal distribution does not provide a good fit as the sample quantiles indicate right-skewed fat tails. The Skew Student-t distribution, however, provide a satisfactory overlap between empirical and theoretical quantiles.

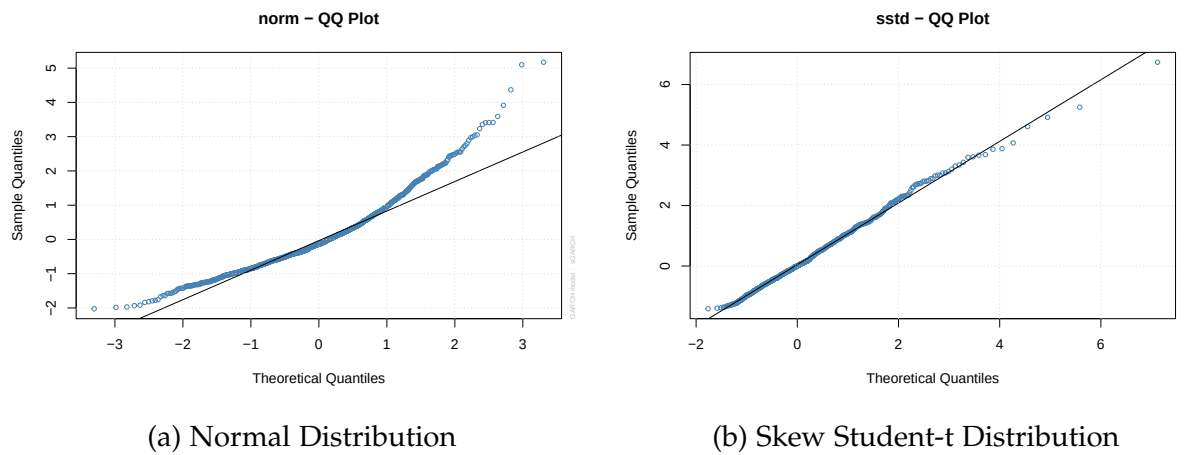


Figure 8: Q-Q plots of Residuals from Equation (6)

B Descriptive Statistics and Unit Root Tests

The below table report summary statistics and unit root tests based on the data split in a pre-COVID-19, and a COVID-19 period as outlined in Section 3.3. Importantly, Panel B show that a majority of the unit root tests suggests that both series in both regimes are stationary time series.

Table 4: Descriptive Statistics and Unit Root tests.

Panel A: Descriptive Statistics						
	N	Min	Median	Mean	St. Dev.	Max
<i>During Covid</i>						
CSAD	527	0.010	0.222	0.280	0.217	1.587
Sentiment	527	−0.319	−0.076	−0.076	0.075	0.140
<i>Pre Covid</i>						
CSAD	528	0.005	0.142	0.168	0.117	0.772
Sentiment	528	−0.282	−0.062	−0.062	0.073	0.169
Panel B: Unit Root tests						
	ADF(1)	ADF(2)	PP(1)	PP(2)		
<i>During Covid</i>						
CSAD	−4.191***	−4.387***	−143.032***	−455.127***		
Sentiment	−2.493	−3.281*	−252.287***	−552.214***		
<i>Pre Covid</i>						
CSAD	−2.709*	−3.348*	−140.7818***	−522.9378***		
Sentiment	−8.301***	−8.304***	−431.401***	−533.103***		

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ADF denotes the Augmented Dickey-Fuller test and PP the Phillips-Perron test for unit root. (1) and (2) refer to using a constant and trend in the test regression, respectively.

C Long-run effects

Long-run effects refer here to long-run multipliers in ADL-models defined as

$$\delta = \frac{\sum_{i=1}^6 \theta_i}{1 - \sum_{j=1}^5 \psi_j},$$

which is a non-linear transformation of the parameter estimates of model (6), hence, the delta method is used for inference. That is, the transformation of the random variables are expanded around their mean with a first-order Taylor approximation

after which the variance is computed. For a generic transformation, $G(X)$:

$$G(X) \approx G(\mu_X) + \nabla G(\mu_X)^T (X - \mu_X)$$

$$Var(G(X)) \approx \nabla G(\mu_X)^T Cov(X) \nabla G(\mu_X).$$

Where in this case $G(X) = \delta$, so $Var(G(X))$ is a scalar.

The long-run multiplier for model (8) is computed by the logic presented above. Here, however, discriminating between the COVID-19 period and the observations prior to the outbreak of the pandemic.

Table 5: Long-run effects of Sentiment on herding during and before COVID-19.

Regime	Parameter	Estimate	Standard Error	Z-Value	p-value
COVID-19	δ	0.0436	0.157	0.278	0.390
Pre-COVID-19	δ	-0.0140	0.124	-0.113	0.545

Notes: Long-run Effects of Sentiment on Herding.

The estimated long-run multiplier is 0.0436 and -0.0140 during and before COVID-19. As the resulting p-values are above conventional levels, the null hypothesis cannot be rejected and no long-run effects of sentiment on herding are thus found.

D Monte Carlo Simulation

D.1 Setup, Definitions and Preliminaries

This section presents a simulation study that aims to highlight the benefits of modeling the dynamics explicitly, rather than defaulting to heteroscedasticity and autocorrelation consistent (HAC) standard errors, when dynamics exist in the data. To recap, the CCK model (Chang et al., 2000) which is commonly used to empirically test for market-level herding, is defined as

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \epsilon_t,$$

and the convention in the literature is to estimate this model with OLS and HAC standard errors. Note that the above is a time series regression model where the time dimension is not at all used. Therefore, the standard CCK model will be referred to as a static model. However, the empirical data as well as theoretical results both suggest there may be persistencies and dynamics in herding behavior, see discussions in Section 1 and 2. The static nature of the standard CCK model simply cannot capture these concepts meaning that all dynamics in the data will be put in the error term, ϵ_t .

Yet, there exist several advantages in modeling the dynamics explicitly, rather than keeping it in the residual terms. In Section 3.2 and 3.3, I propose treating herding as a time series problem, i.e. extending the static CCK model to a model which can capture dynamics both in the mean and variance equation. This section will focus on motivating this extension from a statistical point of view. Concretely, I will focus on efficiency of the estimator and size control of the tests.

Since the point I am trying to make here with the Monte Carlo is a general one, I will not limit myself to the specific model considered earlier. Thus, I disregard exogenous regressors and focus on estimating the mean parameter. To mimic the empirical setting, I make draws from a process which depend on both autoregressive and heteroscedastic components. That is, a process with similar characteristics as the empirical data described in Section 2. More specifically, I assume the following data generating process (DGP):

$$y_t = \mu + u_t, \tag{B1}$$

$$u_t = \gamma u_{t-1} + \epsilon_t = \gamma u_{t-1} + v_t \sqrt{h_t}, \quad v_t \sim \mathcal{N}(0, 1), \tag{B2}$$

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{B3}$$

Where I, as is the case with the real data, see Panel B in Table 1 and 4, assume a stationary DGP: $|\gamma| < 1, \alpha_1 < 1, \beta_1 < 1, \alpha_1 + \beta_1 < 1$. When generating data from this process, I set $y_0 = 0$ and h_0 as the unconditional variance, $h_0 = \sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$. ϵ_0 is then simply defined by $\epsilon_0 = v_0 \sqrt{h_0}$ where v is white noise.

For all simulations, the default GARCH parameter specification used is $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.1, 0.8)$ as this characterizes the conditional heteroscedasticity com-

monly observed in financial data, which is also what I find in the empirical analysis in Section 5. Further, I use a baseline sample of $T = 300$ as this is a representative sample size used in recent empirical papers studying the effect of crisis periods on herding behavior, see Appendix E for a detailed description. To investigate the effect of sample size on efficiency and test size, I also consider $T = 500$ and $T = 750$ as such samples occur in the empirical literature, see Appendix E.

When generating data, I use a burn-in of 5000 observations to remove the effect of initial values. For all simulations, 5000 MC replications are used on the same seed.

The coming sections will investigate the gains in terms of efficiency and size control associated with estimating μ with a MLE on (B1)-(B3) over an OLS with HAC standard errors on (B1) alone. Before moving on, recall the definition of HAC variance:

$$\hat{\sigma}_{HAC}^2 = \frac{1}{T} \left(\hat{\gamma}_0 + 2 \sum_{\ell=1}^{T-1} w(\ell) \hat{\gamma}_\ell \right), \quad (B4)$$

where γ_0 is the unconditional variance, γ_ℓ the ℓ^{th} autocovariance and $w(\ell)$ are kernel weights. For this exercise, I use the popular Bartlett kernel defined as:

$$\begin{cases} w_B(\ell) = 1 - \frac{\ell}{M+1} \text{ for } \ell \leq M \\ w_B(\ell) = 0 \text{ otherwise.} \end{cases} \quad (B5)$$

As is apparent from (B5), only lags up until (and including) M are included when computing the HAC variance, i.e. M is a truncating parameter in the Bartlett kernel. A common choice for M is based on the following rule of thumb: $M = 0.75T^{1/3}$ (Stock and Watson, 2011), which I will use when not otherwise stated.

D.2 Efficiency

It is well known that the Maximum Likelihood estimator achieves the Cramer-Rao lower bound, i.e. a MLE on both the mean and variance equations is expected to be more efficient compared to an OLS on the mean equation alone. This can be illustrated by estimating μ with both approaches and comparing the variance of the point estimate. More concretely, by comparing $\mathbb{V}(\hat{\mu}^{OLS})$ with $\mathbb{V}(\hat{\mu}^{MLE})$ where $\mathbb{V}(\cdot)$

denotes the sample variance and is defined as

$$\mathbb{V}(\hat{\mu}^j) = \frac{1}{N-1} \sum_{n=1}^N (\hat{\mu}_n^j - \bar{\mu}^j)^2, j \in \{OLS, MLE\}. \quad (B6)$$

Here, $\hat{\mu}_n^j$ refer the mean from the n^{th} MC replication estimated with $j \in \{OLS, MLE\}$ while $\bar{\mu}^j$ is the average of those estimates of the mean over all N MC replications. I will compare these two quantities in levels, and also report the relative efficiency of MLE, the latter which is defined as

$$RE_{MLE} = \frac{\mathbb{V}(\hat{\mu}^{OLS})}{\mathbb{V}(\hat{\mu}^{MLE})}, \quad (B7)$$

where $RE_{MLE} < 1$ when OLS is more efficient than MLE and $RE_{MLE} > 1$ when MLE is more efficient. $RE_{MLE} = 1$ means the estimators are equally efficient.

When assessing the efficiency of the two estimators, I will consider two cases. First, for a grid of autoregressive parameters, I will generate data with the default GARCH specification for different sample sizes. Second, for the same grid of autoregressive parameters, I will consider alternative specifications on the GARCH parameters, for the default sample size. These two cases are discussed in turn in the next sections.

D.2.1 Efficiency when varying T

This section investigates how the efficiency of OLS and MLE depends on the sample size and the degree of autocorrelation. I consider the cases when $T \in \{300, 500, 750\}$, $\gamma \in [0.1, 0.9]$ and $\mu = 0$. Once data is generated for each pair (T, γ) , I estimate μ with OLS and HAC robust standard errors on (B1), and with a MLE on (B1)-(B3). After each such estimation, I store the point estimates. The n^{th} MC replication thus gives me $\hat{\mu}_n^{OLS}$ and $\hat{\mu}_n^{MLE}$ which I use to compute the variance of the point estimate.

In Figure 10, I plot $\mathbb{V}(\hat{\mu}^{OLS})$ and $\mathbb{V}(\hat{\mu}^{MLE})$ in the left column, while the right column report the relative efficiency of the MLE. The general pattern in the left column stay the same for the different sample sizes. However, as the sample size increases, the variance in the point estimate obtained from both estimators decreases, as expected. The relative difference does, on the other hand, increase. This visualizes the preference

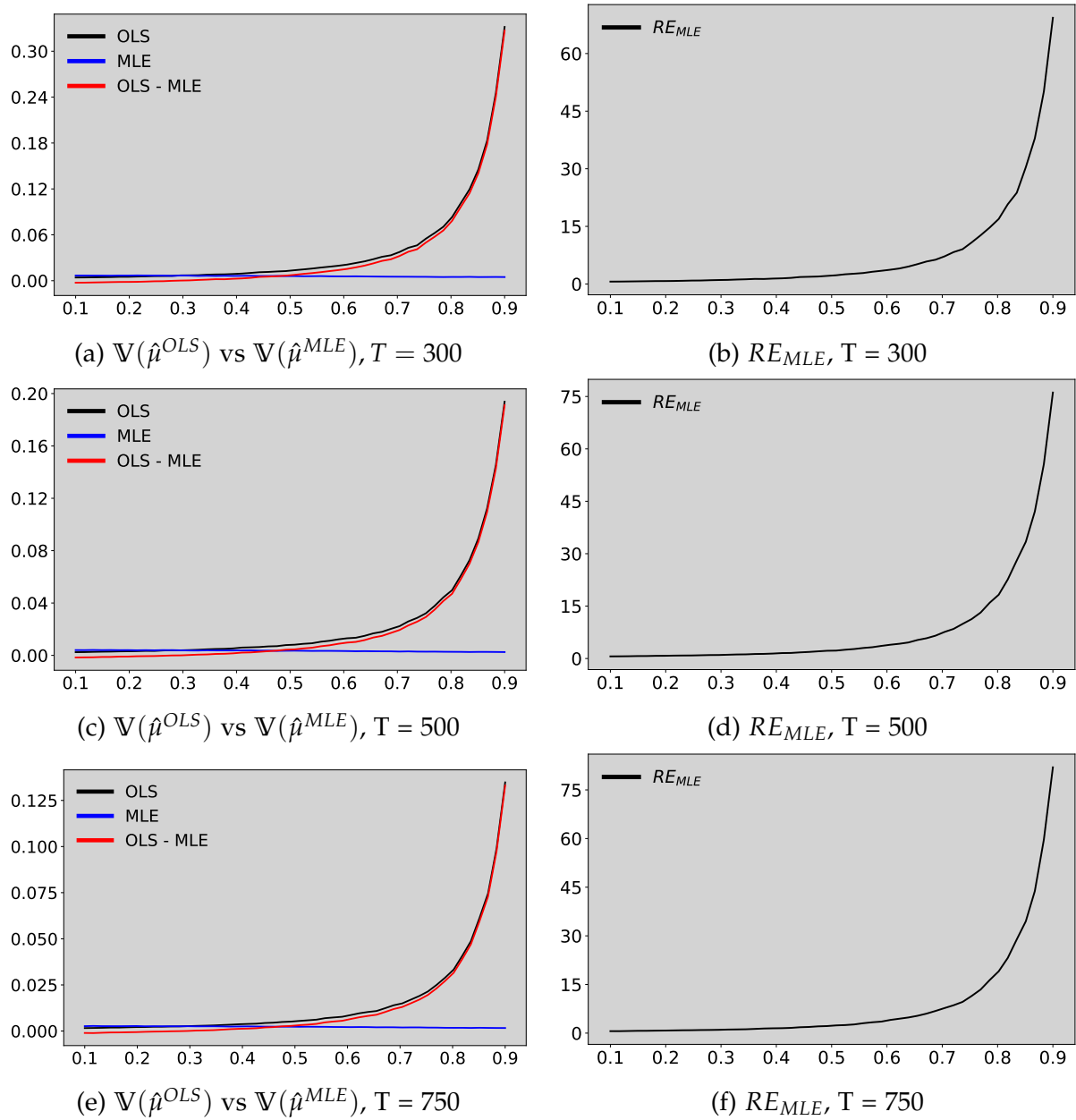


Figure 9: Efficiency for Different Sample Sizes, Default GARCH Specification.

for an estimator that accounts for dynamics in both the mean and variance equation, MLE in this case. For comparison, when $\gamma = 0.4$, RE_{MLE} increases from 1.41 to 1.48 as the sample increases from 300 to 750. Similarly, RE_{MLE} increase from 2.27 to 2.38 when $\gamma = 0.5$, from 3.72 to 4.07 when $\gamma = 0.6$ and from 7.19 to 7.72 when $\gamma = 0.7$.

For the most persistent cases, $RE_{MLE} \geq 65$ for all sample sizes. When $\gamma = 0.3$ the OLS and MLE are as efficient while OLS has a slight advantage when $\gamma < 0.3$.

D.2.2 Efficiency under alternative GARCH specifications

While the above section investigates how efficiency depends on the sample size, this section explore the impact of the GARCH specification. For $T = 300$, I consider $(\alpha_1, \beta_1) = \{(0, 0), (0.1, 0.8), (0.4, 0.4)\}$ as these alternative specifications capture a range of possible combinations (α_1, β_1) observed in financial data. Where $(\alpha_1, \beta_1) = (0, 0)$, I set $\alpha_0 = 0$ to have an IID case. For the remaining specifications, I set $\alpha_0 = 0.1$.

The left column of Figure 10 report $\mathbb{V}(\hat{\mu}^{OLS})$ and $\mathbb{V}(\hat{\mu}^{MLE})$ while RE_{MLE} is plotted in the right column. Each row correspond to a GARCH specification for $T = 300$. Here, RE_{MLE} increase with the autoregressive parameter for all specifications, as in Section D.2.1. However, it is also clear that the relative efficiency of MLE also depends on α_1 as RE_{MLE} is larger in Figure 10(f) compared to 10(b) and 10(d) for larger values γ and lower in Figure 10(f) for lower values γ .

This illustrates that the relative efficiency of MLE depends on both conditional heteroscedasticity and autocorrelation. As the empirical estimates of (α_1, β_1) are close to $(0.1, 0.8)$, MLE is expected to be more efficient than OLS at a wide range of γ values.

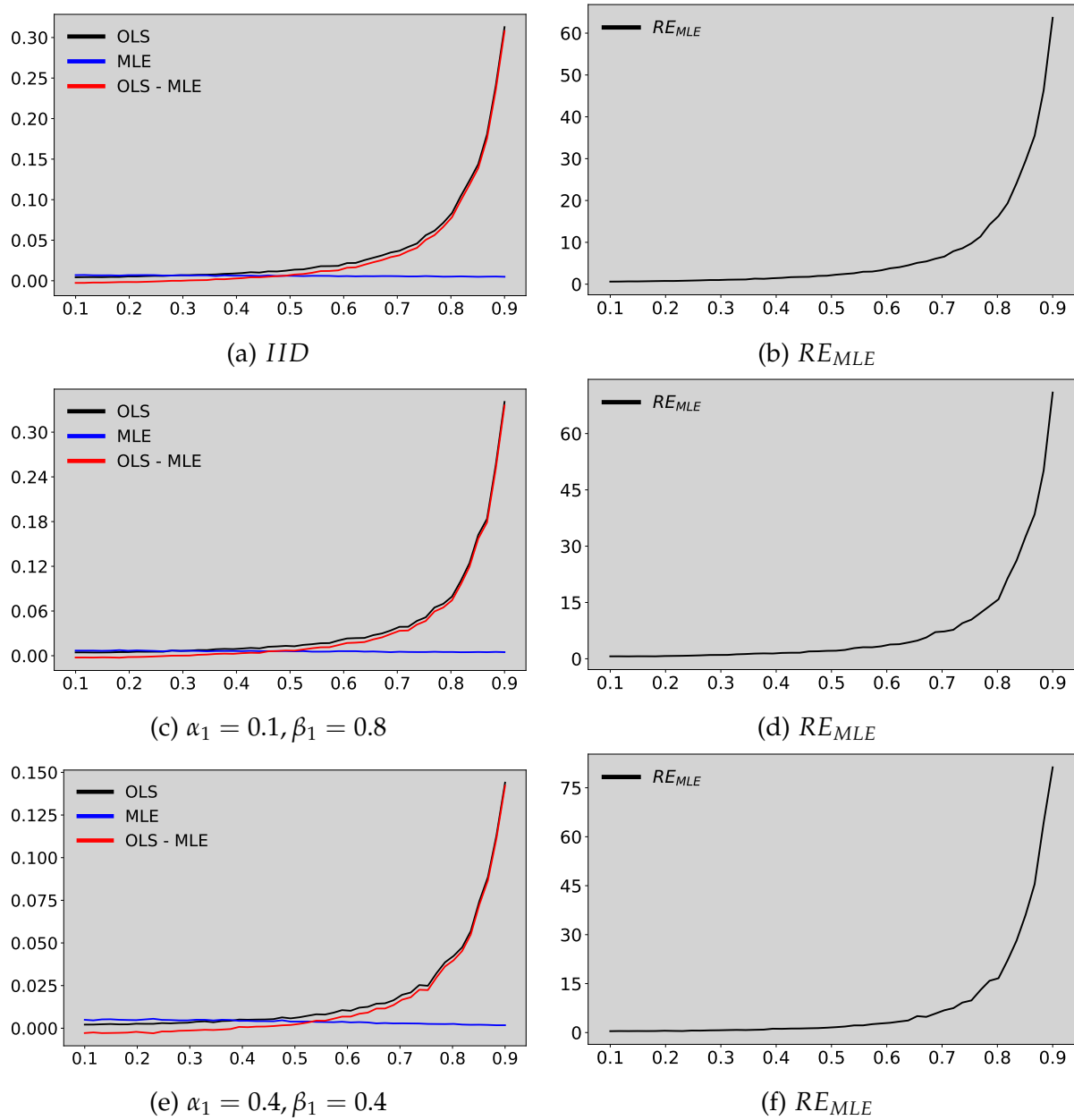


Figure 10: Efficiency for Different GARCH Specifications, $T = 300$.

D.3 Over-sized tests

D.3.1 Setup

This section considers test size, which is the probability of falsely rejecting a null hypothesis. To assess the size of the OLS-HAC robust tests, I borrow the setup from

Kiefer and Vogelsang (2005)⁵, which means generating data according to (B1)-(B3) with $\mu = 0$ and compute rejection probabilities for the null hypothesis $H_0 : \mu \leq 0$ against the alternative $H_1 : \mu > 0$ using the standard HAC robust t-statistic for μ :

$$t_{HAC} = \frac{\sqrt{T}(\hat{\mu} - \mu)}{\sqrt{\hat{\sigma}_{HAC}^2}}, \quad (B8)$$

where $\hat{\sigma}_{HAC}^2$ is computed as in (B4) with the Bartlett kernel weights. The results are reported for all possible bandwidths, i.e. $M \in \{1, 2, \dots, T\}$. When evaluating size, I use the standard normal critical value 1.645 corresponding to a one sided test at the 5% level for all simulations. Note that a two-sided test is qualitatively similar since the distribution of the test statistic is symmetrical, see Kiefer and Vogelsang (2005) for details. To assess size of the MLE tests, I use the same null and alternative as above and compute rejection probabilities using the standard t statistics for $\hat{\mu}$ from a MLE on (B1)-(B3). The t-statistic from the MLE will be denoted t_{MLE} .

In sum, the following sections presents the probability with which $H_0 : \mu \leq 0$ is rejected, for different estimators, specifications and sample sizes. As such, in each iteration of the Monte Carlo I will generate data according to (B1)-(B3), compute t_{MLE} and t_{HAC} and record whether or not the t-statistics exceed the nominal level. The simulation is run 5000 times per specification and 5000 observations are used as a burn in when generating data.

D.3.2 Test Size in Larger Samples

Here I consider the case when $\gamma = \{0.3, 0.5, 0.7, 0.9\}$ and $T = \{300, 500, 750\}$. The rejection probabilities of the OLS and HAC tests are reported in Figure 11 below while the MLE rejection rates for the different cases are found in Table 6. Each row in the Figure corresponds to a sample size and the left column report the results for $\gamma \in \{0.3, 0.5\}$ while the results for $\gamma \in \{0.7, 0.9\}$ are found in the right column. The nominal 5% level is indicated with the dashed black line in all panels. While it can be noted that the simulated rejection frequency is 0.05 ± 0.00604 , i.e. $[0.044, 0.056]$, it

⁵Kiefer and Vogelsang (2005) consider a case in which the errors in (B1) have ARMA structure. I adapt their setup to a case in which the errors are characterized by an AR(1)-GARCH(1,1) structure.

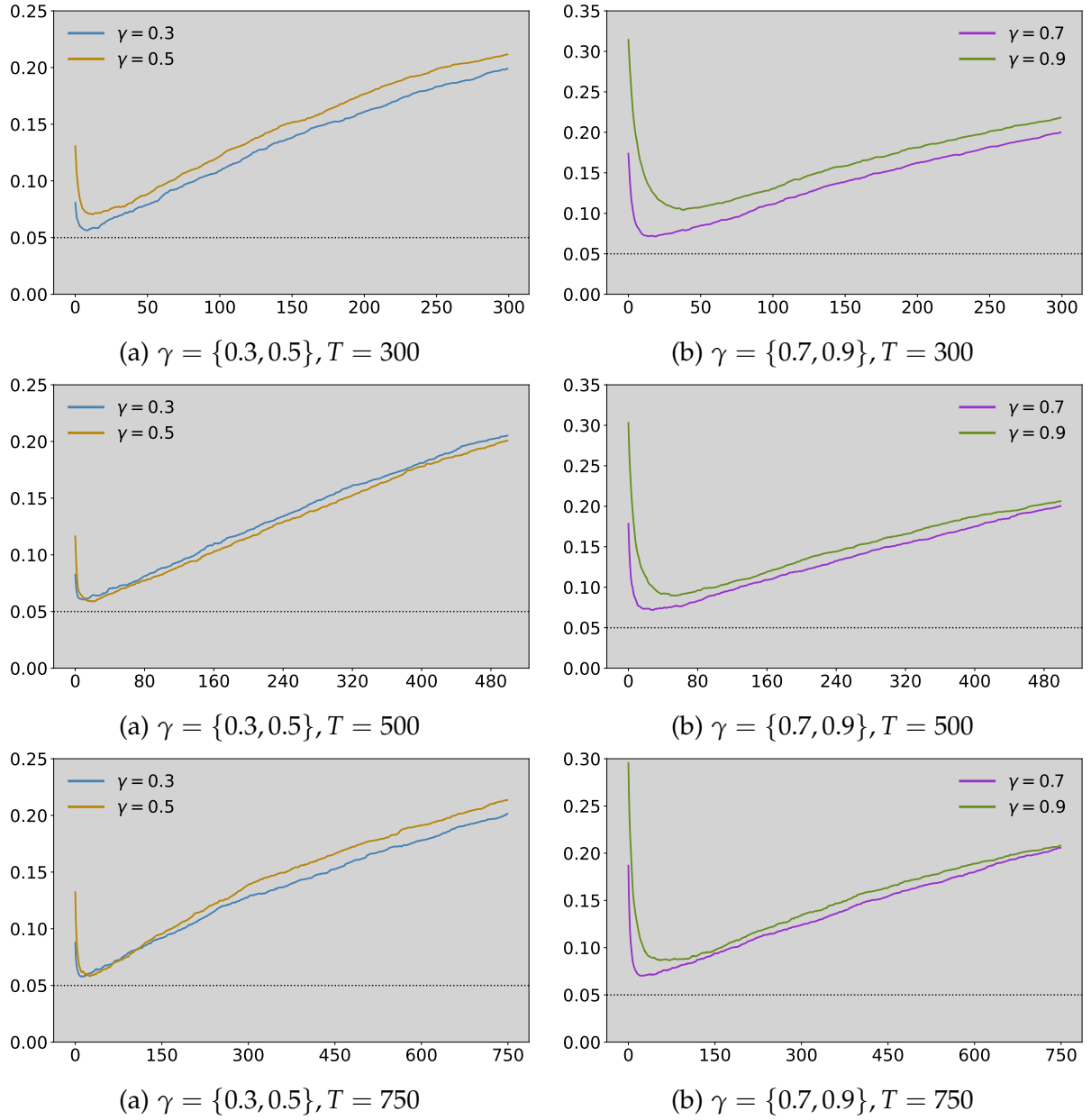


Figure 11: Rejection Rates over the Bandwidth for Different γ and T

is clear that the degree with which the HAC robust tests overreject the null depend both on the autoregressive parameter and the bandwidth. Further, for all γ values, the overrejection rates vary nonlinearly with the bandwidth. For a larger autoregressive parameter, the more the rejection rates are characterized by a 'L' shape. The reason for this pattern is a trade-off between bias and variance in $\hat{\sigma}_{HAC}^2$ that depends on M (Kiefer and Vogelsang, 2005). When M is small, $\hat{\sigma}_{HAC}^2$ is downwards biased but has small variance, hence leading to overrejection. For larger M , the bias initially falls and

the variance increase which is seen in the steep drop on rejection rates in all panels in Figure 11. However, as M continues to increase, bias eventually picks up and variance falls once again. This happens because higher weights are placed on higher-order autocovariances which, in the extreme case means that positive weight is placed on all possible autocovariances for the Bartlett kernel, see Appendix D.1 for a definition. In this situation, it is well known that $\hat{\sigma}_{HAC}^2 = 0$ because the quantity is computed using residuals whose sample average is zero (Kiefer and Vogelsang, 2005). Thus, the limiting case is $\hat{\sigma}_{HAC}^2 = 0$, an estimator with no variance and potentially large bias. This is seen in the upwards sloping region in all panels in Figure 11.

Additionally, the larger the autoregressive parameter the more is the graph shifted northeast, note the scale in the y-axis in Figure 11(a)-11(f). That is, the performance of HAC robust tests deteriorates with increases in γ , for all choices of the bandwidth parameter. For example, when $\gamma = 0.9$ and $T = 300$, HAC rejection rates vary between 10.42% – 31.42%, see Table 6. At this persistency and sample size, there is at minimum a 10.42% probability of rejecting the true null, far exceeding both the nominal and simulated rejection frequency. Selecting the HAC bandwidth with the rule of thumb does in this situation yield rejection rates of 21.04%.

Note that even for the less persistent cases, the minimum rejection rates achieved by the HAC robust tests are all strictly larger than the nominal and simulated levels. This can be seen in Figure 11 but also in Table 6 where $OLS_{Min} > 5.6\% > 5\% \forall T \in \{300, 500, 750\}, \gamma \in \{0.3, 0.5, 0.7, 0.9\}$. Here, OLS_{Min} refer to the minimal rejection rate achieved by the HAC robust test. Moreover, from Table 6 it is also clear that choosing the HAC bandwidth with the rule of thumb results in size distortions relative to that minimum achievable level. The reason is that too few autocorrelation are included when computing the HAC variance, a point made by e.g. Stock and Watson (2015).

Table 6: Rejection Rates for Different Sample Sizes.

	M	$\gamma = 0.3$	M	$\gamma = 0.5$	M	$\gamma = 0.7$	M	$\gamma = 0.9$
T = 300								
<i>MLE</i>	NA	4,4%	NA	5.02%	NA	4,58%	NA	5.9%
<i>OLS_{Min}</i>	9	5.64%	13	7.04%	20	7.14%	39	10.42%
<i>OLS_{Max}</i>	300	19.9%	300	21.16%	300	20.0%	1	31.42%
<i>OLS_{Rule}</i>	5	5.92%	5	7.98%	5	9.44%	5	21.04%
T = 500								
<i>MLE</i>	NA	4.72%	NA	4,72%	NA	5.02%	NA	5.18%
<i>OLS_{Min}</i>	10	6.06%	20	5.88%	29	7.18%	57	8.96%
<i>OLS_{Max}</i>	500	20.52%	500	20.06%	500	20.02%	1	30.3%
<i>OLS_{Rule}</i>	6	6.14%	6	6.66%	6	9.9%	6	19.26%
T = 750								
<i>MLE</i>	NA	5.08%	NA	5.12%	NA	4,64%	NA	5.6%
<i>OLS_{Min}</i>	14	5.76%	25	5.82%	23	7.02%	57	8.66%
<i>OLS_{Max}</i>	750	20.14%	750	21.36%	750	20.58%	1	29.54%
<i>OLS_{Rule}</i>	7	5.98%	7	6.98%	7	9.3%	7	18.08%

Notes: The table report rejection rates for the true null hypothesis $H_0 : \mu \leq 0$ against the alternative $H_1 : \mu > 0$ for different sample sizes under the default GARCH specification, i.e. $\alpha_0 = \alpha_1 = 0.1, \beta_1 = 0.8$. *MLE* refer to the size of the standard t statistic for $\hat{\mu}$ from a MLE on (B1)-(B3). *OLS_{Min}* and *OLS_{Max}* are minimum and maximum rejection rates achieved with the HAC robust test on (B1), respectively. *OLS_{Rule}* refer to the rejection rate when the bandwidth, M , is chosen as $M = 0.75T^{1/3}$. Since the true test size is 0.05, the 95% confidence interval for the simulated rejection frequency is 0.05 ± 0.00604 , i.e. $[0.044, 0.056]$.

In general, the size distortions for the HAC robust tests increase with the autoregressive parameter for each sample size while they decrease over sample sizes. The MLE t-tests are however suggestive of size control as the rejection rates hover around the nominal 5% level for all combinations (T, γ) and only display meaningful deviations in the near unit root case. At $\gamma = 0.9$, MLE reject the true null between 5.18 and 5.9% of the times which indicate an overall good performance, as expected given that the model estimated with MLE capture the dynamic structure in the data.

D.3.3 Test Size for Different GARCH Parameters

Here, for $\gamma = \{0.3, 0.5, 0.7, 0.9\}$ and $T = 300$, I investigate the effect of the GARCH specification on test size. Similarly to Section D.2.2, I consider the cases $(\alpha_1, \beta_1) = \{(0,0), (0.1, 0.8), (0.4, 0.4)\}$, set $\alpha_0 = 0$ when $(\alpha_1, \beta_1) = (0,0)$ and $\alpha_0 = 0.1$ otherwise. The rejection rates for HAC robust tests are found in Figure 12 while those for the MLE are reported in Table 7. The nominal rejection rate is depicted in Figure 12 as the black dashed line.

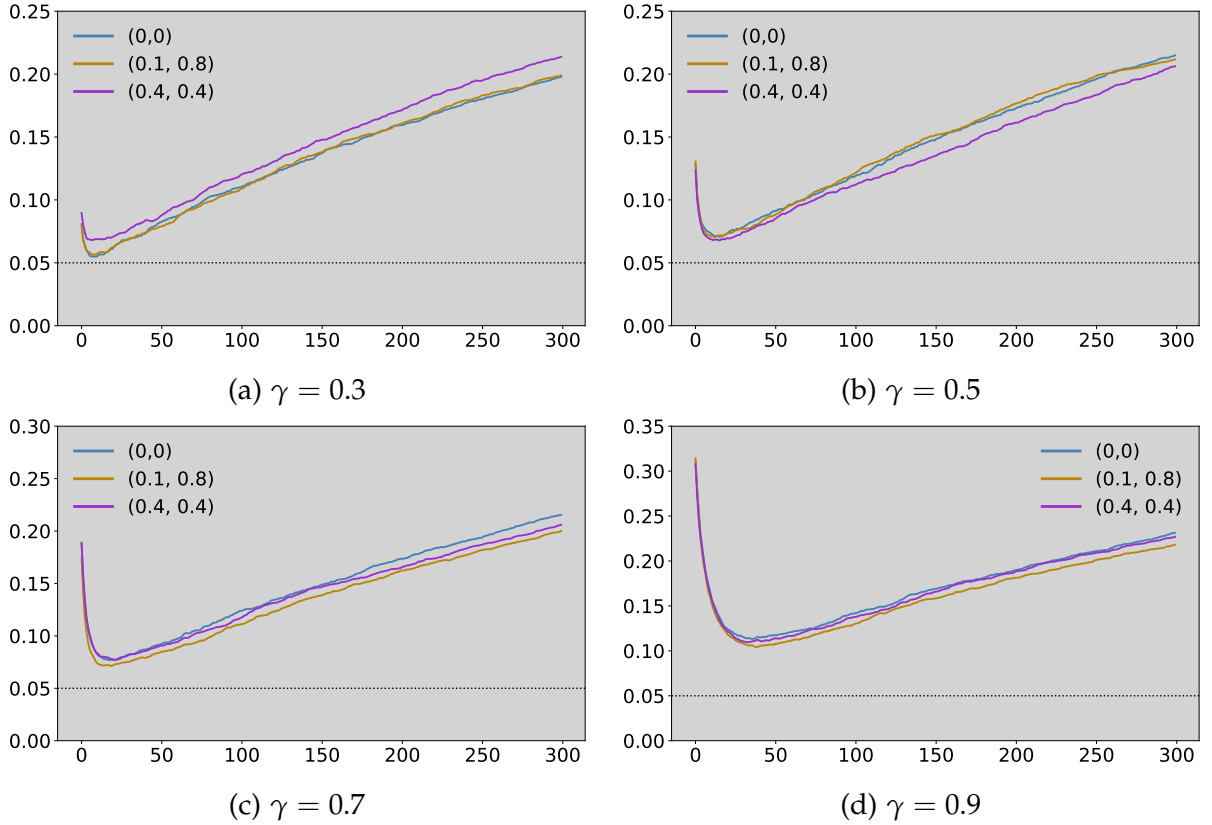


Figure 12: Rejection Rates for Different GARCH Parameters, $T = 300$.

From Figure 12 there are no meaningful differences in rejection rates related to the GARCH specification. Rather, the cases with $(\alpha_1, \beta_1) = \{(0.1, 0.8), (0.4, 0.4)\}$ are akin to adding additional noise relative to the *IID* case. This means that the general discussion from Section D.3.2 applies here as well since the size distortion for the HAC robust tests increase with the autoregressive parameter for each GARCH specification.

Similar to the above investigation over sample sizes, $OLS_{Min} > 5\%$ and $OLS_{Rule} > OLS_{Min}$ for all GARCH specifications and values γ considered. Only for the *IID* case

with $\gamma = 0.3$ does OLS_{Min} appear inside the 95% confidence interval for the simulated rejection frequency. This means that for all empirically relevant cases, the HAC robust tests overreject the null for all bandwidth values, and that choosing the bandwidth with the rule of thumb yields even larger rejection rates.

Table 7: Rejection rates for Different GARCH Specifications.

	M	$\gamma = 0.3$	M	$\gamma = 0.5$	M	$\gamma = 0.7$	M	$\gamma = 0.9$
$\alpha_0 = \alpha_1 = \beta_1 = 0$								
<i>MLE</i>	NA	4.08%	NA	5.42%	NA	4.92%	NA	6.30%
<i>OLS_{Min}</i>	10	5.48%	16	7.04%	18	7.68%	36	11.32%
<i>OLS_{Max}</i>	300	19.78%	300	21.50%	300	21.54%	1	31.04%
<i>OLS_{Rule}</i>	5	5.82%	5	8.1%	5	10.64%	5	21.74%
$\alpha_1 = 0.1, \beta_1 = 0.8$								
<i>MLE</i>	NA	4.4%	NA	5.02%	NA	4.58%	NA	5.90%
<i>OLS_{Min}</i>	9	5.64%	13	7.04%	20	7.14%	39	10.42%
<i>OLS_{Max}</i>	300	19.90%	300	21.16%	300	20.0%	1	31.42%
<i>OLS_{Rule}</i>	5	5.92%	5	7.98%	5	9.44%	5	21.04%
$\alpha_1 = 0.4, \beta_1 = 0.4$								
<i>MLE</i>	NA	5.08%	NA	4.96%	NA	5.54%	NA	5.26%
<i>OLS_{Min}</i>	8	6.80%	16	6.78%	22	7.74%	33	10.94%
<i>OLS_{Max}</i>	300	21.38%	300	20.62%	300	20.60%	1	30.76%
<i>OLS_{Rule}</i>	5	6.86%	5	7.72%	5	10.88%	5	21.60%

Notes: The table report rejection rates for the true null hypothesis $H_0 : \mu \leq 0$ against the alternative $H_1 : \mu > 0$ for the sample size $T = 300$ and different GARCH specifications. $\alpha_0 = 0.1$ except for the *IID* case. *MLE* refer to the size of the standard t statistic for $\hat{\mu}$ from a MLE on (B1)-(B3). *OLS_{Min}* and *OLS_{Max}* are minimum and maximum rejection rates achieved with the HAC robust test on (B1), respectively. *OLS_{Rule}* refer to the rejection rate when the bandwidth, M , is chosen as $M = 0.75T^{1/3}$. Since the true test size is 0.05, the 95% confidence interval for the simulated rejection frequency is 0.05 ± 0.00604 , i.e. $[0.044, 0.056]$.

The MLE t-tests show size control for all specifications with rejection rates reported between 4.08% and 6.3%. The two slight deviations reported in Table 7 occurs in the near unit root case ($\gamma = 0.9$), deviations which vanish asymptotically as seen in Table 6, which is satisfactory.

E Empirical Sample Sizes

This section documents the sample sizes used in recent empirical papers studying herding in different crisis periods. The average length of the crisis-periods in the below papers is 451 days. Considering only the COVID-19 periods, the average sample size is 425 days. Note that Espinosa-Méndez and Arias (2021) only have around 100 observations after the outbreak of the pandemic, using the reported 616 days heavily inflates the average reported here. In sum, the range of the sample sizes reported here motivates the choice of $T \in \{300, 500, 750\}$ in the Monte Carlo.

- Economou et al. (2018):
 - Full sample: January 1 2004 - July 31 2014: 2659 days
 - GFC sample: January 1 2007 - December 31 2009: 753 days
 - Structural breaks samples:
 - * US: 1531 days (January 1 2004 - February 3 2010) and 1127 days (February 4 2010 - July 31 2014)
 - * UK: 980 days (January 1 2004 - November 19 2007) and 1678 days (November 20 2007 - July 31 2014)
 - * Germany: 965 days (January 1 2004 - October 28 2007) and 1694 days (October 29 2007 - July 31 2014)
- Chang et al. (2020):
 - Full sample: March 24 2000 - May 29 2020: 5059 days
 - GFC sample: July 2 2007 - December 31 2009: 630 days
 - SARS sample: February 10 2003 - July 5 2013: 2614 days
 - COVID sample: January 30 2020 - May 29 2020: 82 days
- Espinosa-Méndez and Arias (2021):
 - Full sample: January 3 2000 - June 19 2020. France = 4940 days, Germany = 4737 days, Italy = 4679 days, UK = 4837 days and Spain = 4917 days.

- COVID-19 sample: January 2 2018 - June 19 2020: 616 days.
- Active COVID-19 periods were: 101 days for France (January 24 2020 - June 19 2020), 100 days for Germany (January 27 2020 - June 19 2020), 97 days for Italy (January 30 2020 - June 19 2020) and 96 days for the UK and Spain (January 31 2020 - June 19 2020).
- Ferreruela and Mallor (2021):
 - Full sample: January 1 2000 - May 31 2021: 5369 days.
 - Pre-crisis sample: January 1 2000 - September 30 2008: 2192 days.
 - Crisis sample: October 1 2008 - April 1 2009: 124 days.
 - Post-crisis sample: April 2 2009 - January 31 2020: 2718 days.
 - COVID-19 period: 20th February 2020 – 27th November 2021 = 448 days.
- Jia et al. (2022): No start date reported for COVID 19, I use the WHO announcement on the 20th of January as proxy here.
 - Full sample: January 1 2016 - June 30 2022: 1827 days
 - Pre-COVID sample: January 1 2016 - January 19 2020: 1015 days
 - COVID sample: January 20 2020 - June 30 2022: 616 days
- Mandaci and Cagli (2022)⁶:
 - Full sample: December 31 2018 - January 12 2021: 507 days
 - Pre-COVID sample: December 31 2018 - December 31 2019: 250 days
 - COVID sample: January 1 2020 - January 12 2021: 257 days
- Ampofo et al. (2023):
 - Full sample: December 5 2017 - February 28 2022 = 1126 days,
 - Pre-crisis sample: December 5 2017 - January 20 2020 = 593 days,
 - Crisis sample: January 20 2020 - February 28 2022 = 533 days.

⁶Note that Mandaci and Cagli (2022) study herding intensity on cryptocurrency markets using orderbook data. Therefore they have many more observations than calendar dates. The duration of their COVID period is however relevant.

References

- Adämmer, P., Prüser, J., and Schüssler, R. A. (2024). Forecasting Macroeconomic Tail Risk in Real Time: Do Textual Data Add Value? International Journal of Forecasting.
- Ampofo, R. T., Aidoo, E. N., Ntiamoah, B. O., Frimpong, O., and Sasu, D. (2023). An Empirical Investigation of COVID-19 Effects on Herding Behaviour in USA and UK Stock Markets using a Quantile Regression Approach. Journal of Economics and Finance, 47(2):517–540.
- Andres-Escayola, E., Ghirelli, C., Molina, L., Perez, J. J., and Vidal, E. (2024). Using Newspapers for Textual Indicators: Guidance Based on Spanish-and Portuguese-Speaking Countries. Computational Economics, 64(2):643–692.
- Baker, M. and Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61(4):1645–1680.
- Barber, B. M., Odean, T., and Zhu, N. (2009). Systematic Noise. Journal of Financial Markets, 12(4):547–569.
- Ben-David, I., Franzoni, F., and Moussawi, R. (2017). Exchange-traded funds. Annual Review of Financial Economics, 9(1):169–189.
- Ben-David, I., Franzoni, F., and Moussawi, R. (2018). Do etfs increase volatility? The Journal of Finance, 73(6):2471–2535.
- Bikhchandani, S. and Sharma, S. (2000). Herd Behavior in Financial Markets. IMF Staff papers, 47(3):279–310.
- Blasco, N., Corredor, P., and Ferrer, E. (2018). Analysts Herding: When does Sentiment Matter? Applied Economics, 50(51):5495–5509.
- Boubaker, S., Liu, Z., and Zhai, L. (2021). Big Data, News Diversity and Financial Market Crash. Technological Forecasting and Social Change, 168:120755.
- CDC (2023). CDC Museum COVID-19 Timeline. Available at: <https://www.cdc.gov/museum/timeline/covid19.html> (Accessed: 2024-04-08).

- Chang, C.-L., McAleer, M., and Wang, Y.-A. (2020). Herding Behaviour in Energy Stock Markets during the Global Financial Crisis, SARS, and ongoing COVID-19. Renewable and Sustainable Energy Reviews, 134:110349.
- Chang, E. C., Cheng, J. W., and Khorana, A. (2000). An Examination of Herd Behavior in Equity Markets: An International Perspective. Journal of Banking & Finance, 24(10):1651–1679.
- Chattopadhyay, M., Garg, A. K., and Mitra, S. K. (2018). Herding by Foreign Institutional Investors: An Evidential Exploration for Persistence and Predictability. Journal of Behavioral Finance, 19(1):73–88.
- Chen, T. (2013). Do investors herd in global stock markets? Journal of Behavioral Finance, 14(3):230–239.
- Choi, K.-H. and Yoon, S.-M. (2020). Investor Sentiment and Herding Behavior in the Korean Stock Market. International Journal of Financial Studies, 8(2):34.
- Christie, W. G. and Huang, R. D. (1995). Following the Pied Piper: Do Individual Returns Herd Around the Market? Financial Analysts Journal, 51(4):31–37.
- Clements, A., Hurn, S., and Shi, S. (2017). An Empirical Investigation of Herding in the US Stock Market. Economic Modelling, 67:184–192.
- Duygun, M., Tunaru, R., and Vioto, D. (2021). Herding by Corporates in the US and the Eurozone Through Different Market Conditions. Journal of International Money and Finance, 110:102311.
- Economou, F., Hassapis, C., and Philippas, N. (2018). Investors' Fear and Herding in the Stock Market. Applied Economics, 50(34-35):3654–3663.
- Elton, E. J. and Gruber, M. J. (2013). Mutual funds. In Handbook of the Economics of Finance, volume 2, pages 1011–1061. Elsevier.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica: Journal of the Econometric Society, pages 987–1007.

- Espinosa-Méndez, C. and Arias, J. (2021). COVID-19 Effect on Herding Behaviour in European Capital Markets. Finance Research Letters, 38:101787.
- ETFGI (2025). ETFGI reports that assets invested in the global ETFs industry reached a new record of US\$15.50 trillion at the end of february. Press release. Available at: <https://etfgi.com/news/press-releases/2025/03/etfgi-reports-assets-invested-global-etfs-industry-reached-new-record> (Accessed: 2025-06-16).
- Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. Journal of the American Statistical Association, 93(441):359–371.
- Ferreruela, S. and Mallor, T. (2021). Herding in the Bad Times: The 2008 and COVID-19 Crises. The North American Journal of Economics and Finance, 58:101531.
- Fleissig, A. R. and Swofford, J. L. (2023). Habit Persistence in Assets Demand. Southern Economic Journal, 89(3):975–985.
- Garcia, J. (2021). Analyst Herding and Firm-Level Investor Sentiment. Financial Markets and Portfolio Management, 35(4):461–494.
- Hau, H. and Lai, S. (2017). The role of equity funds in the financial crisis propagation. Review of Finance, 21(1):77–108.
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., et al. (2020). First Case of 2019 Novel Coronavirus in the United States. New England Journal of Medicine, 382(10):929–936.
- Huang, A. H., Wang, H., and Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. Contemporary Accounting Research, 40(2):806–841.
- Hudson, Y., Yan, M., and Zhang, D. (2020). Herd Behaviour & Investor Sentiment: Evidence from UK Mutual Funds. International Review of Financial Analysis, 71:101494.

- Iglesias, E. M. and Phillips, G. D. (2001). Reconsidering the Gains in Efficiency from ML Estimation versus OLS in ARCH Models. Economics Letters, 74(1):21–24.
- Investment Company Institute (2025). Trends in mutual fund investing, april 2025. Investment Company Institute Monthly Statistical Release. Available at: https://www.ici.org/research/stats/trends_04_25 (Accessed 2025-05-30).
- Jansson, M. (2004). The Error in Rejection Probability of Simple Autocorrelation Robust Tests. Econometrica: Journal of the Econometric Society, 72(3):937–946.
- Jia, B., Shen, D., and Zhang, W. (2022). Extreme Sentiment and Herding: Evidence from the Cryptocurrency Market. Research in International Business and Finance, 63:101770.
- Kaplanski, G. and Levy, H. (2017). Analysts and Sentiment: A Causality Study. The Quarterly Review of Economics and Finance, 63:315–327.
- Kiefer, N. M. and Vogelsang, T. J. (2005). A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. Econometric Theory, 21(6):1130–1164.
- Kiely, T. (2023). List of the 16 Largest US Newspapers. Available at: <https://www.meltwater.com/en/blog/largest-us-newspapers> (Accessed: 2024-01-08).
- Kim, J. J., Dong, H., Choi, J., and Chang, S. R. (2022). Sentiment Change and Negative Herding: Evidence from Microblogging and News. Journal of Business Research, 142:364–376.
- Lakonishok, J., Shleifer, A., and Vishny, R. W. (1992). The Impact of Institutional Trading on Stock Prices. Journal of Financial Economics, 32(1):23–43.
- Leippold, M. (2023). Sentiment Spin: Attacking Financial Sentiment with GPT-3. Finance Research Letters, 55:103957.
- Lux, T. (1995). Herd Behaviour, Bubbles and Crashes. The Economic Journal, 105(431):881–896.

- Mach, K. J., Salas Reyes, R., Pentz, B., Taylor, J., Costa, C. A., Cruz, S. G., Thomas, K. E., Arnott, J. C., Donald, R., Jagannathan, K., et al. (2021). News Media Coverage of COVID-19 Public Health and Policy Information. Humanities and Social Sciences Communications, 8(1).
- Mandaci, P. E. and Cagli, E. C. (2022). Herding Intensity and Colatility in Cryptocurrency Markets during the COVID-19. Finance Research Letters, 46:102382.
- Merli, M. and Roger, T. (2013). What Drives the Herding Behavior of Individual Investors? Finance, 34(3):67–104.
- Muñoz, S. (2006). Habit Formation and Persistence in Individual Asset Portfolio Holdings: The Case of Italy. IMF Working Paper.
- Neave, H. R. (1970). An Improved Formula for the Asymptotic Variance of Spectrum Estimates. The Annals of Mathematical Statistics, 41(1):70–77.
- Oliver Wyman (2023). The renaissance of ETFs: Exchange-traded funds are fuelling market growth. Technical Report. Available at: https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2023/apr/Oliver_Wyman_Exchange_traded_funds_are_fueling_market_growth.pdf (Accessed: 2025-06-10).
- Pantula, S. G. (1988). Estimation of Autoregressive Models with ARCH Errors. Sankhyā: The Indian Journal of Statistics, Series B, pages 119–138.
- Pochea, M.-M., Filip, A.-M., and 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.
- Poterba, J. M. and Shoven, J. B. (2002). Exchange-traded funds: A new investment option for taxable investors. The American Economic Review, 92(2):422–427.
- Scharfstein, D. S. and Stein, J. C. (1990). Herd Behavior and Investment. The American Economic Review, pages 465–479.
- Stock, J. H. and Watson, M. W. (2011). Introduction to Econometrics. Pearson.

- Stock, J. H. and Watson, M. W. (2015). Time Series Econometrics. AEA Continuing Education Program. Available at: <https://www.aeaweb.org/webcasts/2015/conted> (Accessed 2025-05-23).
- Sun, Y., Phillips, P. C., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. Econometrica: Journal of the Econometric Society, 76(1):175–194.
- Welch, I. (2000). Herding Among Security Analysts. Journal of Financial Economics, 58(3):369–396.
- Wu, G., Yang, B., and Zhao, N. (2020). Herding Behavior in Chinese Stock Markets during COVID-19. Emerging Markets Finance and Trade, 56(15):3578–3587.
- Yarovaya, L., Matkovskyy, R., and 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.