Herding within Chinese Equity Markets in Response to COVID-19 'Lockdown-Style' Containment Measures

By

Sam L. Blundell

Applied Economic Dissertation

School of Economics

UNIVERSITY OF BRISTOL

March 2023

Contents

1 Introduction

Herding within financial markets, the theory of investors abandoning private information and mimicking the decisions of others, has received significant empirical attention given its theoretical ability to exacerbate market volatility and facilitate asset bubbles (Avery & Zemsky, 1998). Existing research on the interaction between government responses and herding has concluded that governments reduce multidimensional uncertainty by providing investors with quality and timely information (Kizys *et al.*, 2021; Sharif *et al.*, 2020), which in turn mitigates herding behaviour (Avery & Zemsky, 1998). However, 'lockdown-style' policies have been linked to increased cognitive stress (Dubey *et al.*, 2020; Aknin *et al.*, 2022) , which conversely is theorised to increase herding behaviour (Prechter, 2010). Amongst a backdrop of increased focus on behavioural elements in policy-making decisions (ESMA, 2017), this paper utilises methodology proposed by Chang *et al.* (2000) to investigate the effect of 'lockdown-style' policy responses to COVID-19 in China on herding within Chinese equity markets. We find that whilst stricter government responses to the pandemic can be linked to a decrease in herding behaviour, this effect reverses under the strictest 5% of lockdowns, with herd behaviour increasing during these periods. The paper is structured as follows: firstly in Section 2, we will explore the theoretical mechanisms that cause herding and the empirical literature documenting the presence of herding in financial markets. Finally, in Section 3, we will document this paper's empirical methodology, subsequently discussing our results in the context of the empirical landscape.

2 Theoretical and Empirical Review

2.1 Theory of Herding

Rational Models: Banerjee (1992) presents the foremost model for rational herding, arguing that in cases such as financial markets, where economic agents cannot fully determine what information other agents possess entirely through their actions, it is often rational for agents to disregard their private information and instead follow those before them. The paper presents a sequential game, where a population of investors of size N select an asset i to invest in from a non-finite set of assets, in order to maximise their monetary gain from the investment. An unknown asset from this group, i^*, will result in the highest monetary payoff $z(i)$ of $z(i^*) = z$ above that of all other assets $i \neq i^*$. The private information an investor possesses in this game is a so-called ‘signal’ i', which tells an investor which asset is the true highest-returning asset, $i' = i^*$. Uncertainty is introduced into the game, as investors receive such a signal with probability α, and if they do receive a signal it is correct with probability β. For example, an investor receives a misleading signal of $i' \neq i^*$ with probability $\alpha(1-\beta)$. Under these rules, investors sequentially choose which asset to invest into, with the first investor chosen at random. The next randomly chosen investor follows, with the knowledge of the actions of the previous investors, but not knowledge of their signal. Subsequently, investors have a choice of whether to follow their own signal, if they have one, or to follow the decisions of others and ‘herd’. This decision, under Bayesian logic, depends on the probability under each choice of choosing the correct asset i^*. It is important to note that the game has three foundational assumptions, with their purpose being to resolve tie-breaking circumstances whilst minimizing the possibility of herding arising from the game’s construction - maintaining the model’s external validity: Assumption A: Whenever a decision-maker has no signal and everyone else has chosen $i = 0$, they always choose $i = 0$. Assumption B: When decision-makers are indifferent between following their own signal and following someone else’s choice, they always follow their own signal. Assumption C: When a decision-maker is indifferent between following more than one of the previous decision-makers, she chooses to follow the one who has the highest value of i.

Working sequentially through the game, the first decision-maker will simply follow their own signal if they have one, and if not will choose $i = 0$ as per Assumption A. The second decision-maker will follow the prior investor if they do not have a signal. If they do, they are aware that both their own and the other investor’s signal are equally as likely, with probability β, to be correct. Under Assumption B, they will follow their own signal. The third decision-maker has multiple options. If they do not have a signal, they will follow the investor that chooses the highest i (under Assumption C). If one or both prior investors choose $i = 0$, the decision-maker will follow their own signal if they have one. If both investors choose $i \neq 0$, but they choose different i, the third investor will still follow their own signal. This is because they know that a single investor that chooses an $i \neq 0$ is just as likely to be correct as their own signal is. But what if both prior investors agree upon an $i \neq 0$?

We can consider this final third investor’s decision-making process as an illustrative case for the kth investor for the rest of the game. In this case, the two prior investors have chosen the same asset $\bar{i} \neq i^*$, with this event denoted by H. This can happen if the second investor does not have a signal and the first investor has signal $i' = \bar{i}$, or if both investors happen to have the same signal. Under this circumstance, the third investor can choose to follow these prior investors, or choose their own signal if they have one. They will assess this choice based on the probability of the previous investors' asset $\bar{i}$ being the highest paying asset i^*,

$$P[i^* = \bar{i}|H] = \frac{\alpha^3\beta^2(1-\beta)+\alpha^2\beta(1-\beta)(1-\alpha)}{P[H]},$$

against the probability that their own signal $i' = i^*$,

$$P[i^* = i'|H] = \frac{\alpha^2\beta(1-\beta)(1-\alpha)\beta}{P[H]}.$$

Upon inspection, it is clear that $P[i^* = \bar{i}|H] > P[i^* = i'|H]$, and investor three should choose to disregard their private information and herd. It is important to note that investor three does not know whether the previous investor has a signal, nor if any of the investors' signals are correct. This is why the probability of these two factors, α and β, enter into the decision making process. This logic can be extended: once an asset has been chosen by two investors, the subsequent investors should always follow that decision, unless their signal matches that of an asset that has already been chosen by one other investor - as these two assets have an equal likelihood of being correct.

The Nash-Equilibrium of this game builds upon this case. Importantly, investors will always herd if any option has been chosen by more than one person, with this asset being determined by early investors either via a few uncertain private signals that happen to match, or through the combination of a single signal and a lack of dissenting signals. Such a process is termed an 'informational cascade' as initial investor decisions can lead to rapid formation of herds mimicking these decisions. Regardless of whether previous investors have material information, in the form of a signal, or not, subsequent investors have no way of inferring such from their actions. As such, when an asset attracts multiple investors subsequent decision-makers implicitly assume that these investors possess information that they do not. When an investor abandons their own information and follow others, this imposes a negative externality on sequential investors - depriving them of information that their action would have provided if they had followed their signal. Importantly, as the probability of investors herding to the incorrect asset is strictly positive, the result of the model is inefficient, even in the long run.

Welch (1992) extends the logic of Banerjee's model of herding to explain herding by investors of initial public offerings (IPOs). As in Banerjee, assuming an environment where investors can only observe previous investor purchasing decisions, and not anyone else's private information, later investors observe the purchasing decisions of earlier investors, leading to aforementioned 'informational cascades' where investors optimally discard their private information and follow the decisions of others. Welch furthers this theory by considering the seller reactions to such informational cascades. The seller, which within the context of an IPO is the underwriter, faces reduced pricing pressure when an informational cascade is present. As such, issuers are better off when investors cannot communicate, and are in fact incentivised to prevent such communication.

Avery & Zemsky (1998) (herein AZ) claims, however, that models such as Banerjee (1992) and Welch (1992) that fix the price of taking an action ex ante are unsuitable for studying herding in the context of a financial market, given these models neglect the behaviour change variable pricing and subsequent market-clearing results in. AZ provide the antidote to this, developing a model for herding that possesses an adjustable price mechanism. The paper concludes that in the presence of a single dimension of uncertainty, e.g. an exogenous shock to an asset value, these adjustments prevent herd behaviour by eliminating the possibility of informational cascades[1]. Importantly, however, in the presence of multiple dimensions of uncertainty, ie. when there is uncertainty regarding an asset's price relative to its initial value and the average accuracy of private information, informational cascades become possible and, in some cases, even prevalent.

There also exists a wide range of literature detailing rational theories of herding that do not involve such informational cascades, with investors possessing alternative motivations to follow others' actions other than the belief that they possess higher-quality information. Scharfstein & Stein (1990), for instance, proposed that investment managers are incentivised to rationally herd in line with other managers in order to maintain their reputation in the eyes of the labour market, mirroring the classic 'principle-agent problem' in Economic theory. The model consists of two investor types: 'smart' investors, who receive signals based upon the profitability of an investment, and 'dumb' investors, who receive signals that consist purely of noise. The investor type is unknown to both the investment manager and the labour market, and each investor type is equally likely to receive a signal; therefore investor type cannot be inferred by the receipt of a signal. Therefore, one's belief regarding a manager's type is updated based on their investment behaviour: either how profitable a decision is, or whether this decision is in line with other managers' decisions. This second performance indicator is the critical mechanism resulting in herding - when unprofitable decisions are made, managers who act in accordance with each other 'share the blame', lessening the blow to each individual manager than if they made that decision alone. 'Smart' managers tend to receive correlated signals, as they are observing facts about the true value of an asset. As such, 'dumb' managers can masquerade as smart if they follow others, likewise, they

[1]This assumes that a non-zero number of investors possess private information

are more likely to be perceived as 'dumb' if they act in a contrarian manner. Therefore, even 'smart' managers are incentivised to discard accurate private information (it is unknown to them that the information is accurate) and follow public sentiment. As in Banerjee's model, this swamps the information space of the market for investments, and information that would be contained within an investment's price is lost.

It is worth noting that a distinction needs to be made regarding the 'intentional' herding described by these models, and 'spurious herding' as termed by Bikhchandani & Sharma (2000). Spurious herding is where investors, who face a similar decision-set, take similar actions independently. For instance, investors reacting to the economic effect of COVID-19 in similar ways may empirically be observed as herding. However, investors are not attempting to mimic others' decisions at all; they are purely processing this information similarly. Crucially in an empirical setting, this spurious herding results in a necessarily efficient outcome, and thus it is important to form a solidified hypothesis regarding the herding mechanism, rather than searching for signs of herding within a market indiscriminately.

AZ's hypothesis regarding the effect of multidimensional uncertainty is particularly relevant when considering the implications of government responses to COVID-19 and herding. The pandemic presents both significant geopolitical and economic risks (Sharif *et al.*, 2020), indicating an environment herding may be prevalent within. As Kizys *et al.* (2021) theorise, containment and closure policy responses signal to investors that the pandemic crisis is under control, providing timely and quality information to investors. These signals reduce said multidimensional uncertainty, which under AZ's rational model acts in mitigation of herding.

Irrational Models: In contrast to these rational explanations for herding, Prechter (2010) utilises psychological models of the brain to create an irrational model of herding based on impulsive mental activity. Pretcher uses the "Triune Brain" model, created by psychologist Paul MacLean, which splits the brain into three basic parts: the brain stem, the limbic system and the neocortex. Within MacLean's model, the limbic system is viewed as responsible for the more "primitive" responses, such as the fight-or-flight impulse (Macklin, 1978). Specifically within the limbic system, Pretcher claims the basal ganglia is responsible for a deep-rooted impulse to herd - impelling desires to be part of a group and to seek acceptance. These processes, MacLean's model claims, overwhelm higher brain functions within the neocortex, which are responsible for rational thought, in emotionally charged situations. Mimicking others is viewed as a deeply ingrained process: a survival mechanism to diffuse attacks. Pretcher likens herding in financial markets to observations of a long-hidden stone age tribe copying a researcher's movements (Rubinstein & Gajdusek, 1970). As such, Pretcher suggests this immutable "primitive" process is the primary mover of financial market prices, with herding behaviour as a substitute for rigorous reasoning when knowledge is lacking. Pretcher points to Olsen (1996) as evidence of this theory, who studies 4000 corporate earnings estimates by company analysts. The study finds a link between the difficulty in forecasting earnings-per-share, which is interpreted as a source of stress (an emotional charge), the greater the herding within estimates. This stress leads to a negative feedback loop: market participants herd out of stress of failure, which results in incorrect decision-making, which induces additional stress. It is important to highlight that whilst both Pretcher's irrational model and Banerjee's rational model are driven by informational uncertainty, irrational herding is not a Bayesian logic-adhering decision that is optimal for the decision-maker - rather it is a sub-optimal decision even within the context of informational uncertainty.

MacLean's Tribiune Brain model which forms the basis of Pretcher theory, however, is severely criticised within modern neurological research. Cesario *et al.* (2020) argue that the idea of primitive, impulsive brain systems being underneath more complex, newer systems is entirely a misconception - with empirical research evidencing that areas of the brain are not added over time radially, but most often are transformed from existing parts. Likewise, "impulsive" decisions such as herding are not immutable, but rather highly moderated by context (Kidd *et al.*, 2013; Gawronski & Cesario, 2013). In other words, it is much more likely that impulsive herding is an interplay between an automatic, irrational process and a rational one. Whilst Pretcher's underlying neurological model is highly questionable, modern psychological research still gives credence to the idea of automatic imitation, or impulsive herding. As Cecilia Heyes' paper on the topic states; "... although automatic imitation is subject to input modulation by attentional processes, and output modulation by inhibitory processes, it is mediated by learned, long-term sensorimotor associations that cannot be altered directly by intentional processes" (Heyes, 2011).

Dubey *et al.* (2020)'s research on the mental health implications of the COVID-19 pandemic provides insight into how Pretcher's model applies to this paper's empirical design. Dubey details the psycho-social burden of quarantine and isolation enforced by government responses to the pandemic. Specifically, the loss of control stemming from quarantines has been found to generate severe feelings of distress and anxiety, which can be magnified by separation from family (Brooks *et al.*, 2020). These adverse mental health outcomes have been reported to be positively associated with quarantine duration (Hawryluck *et al.*, 2004), and specifically linked to policy stringency (Aknin *et al.*, 2022). In contrast to rational based models of herding, we would expect a positive relationship between herding behaviour and 'lockdown-style' policy stringency. The stricter the lockdown policy, and the longer these policies last, the greater stress financial market participants will be subject to, and the greater the likelihood the negative feedback loop detailed by Pretcher will come into effect.

2.2 Empirical Literature Review

In the 30 years following the seminal article Scharfstein & Stein (1990), 329 articles have been published on the topic of herding in financial markets with 15,900 citations (Choijil *et al.*, 2022). Since the Great Financial Crisis, there has been a significant surge in the number of papers published on the topic, with the majority of them appearing in the last seven years. In fact, the number of articles published and the corresponding citations during this period have surpassed those of the previous 24 years. This large body of empirical literature covers numerous models of herding, markets, events and herding consequences.

Measures of Herding: During his herding in financial markets literature review, Spyrou (2013) splits empirical approaches to measuring herding into two categories - approaches that rely on micro-data or proprietary data, and approaches that rely on aggregate pricing or market data.

The latter approach was pioneered by Christie & Huang (1995) (herein referred to as CH), who based their methodology on the observation that if herding is present within a market, the dispersion of returns from the average market return is expected to decrease. This is due to the fact that investors who follow each other's investment choices will drive individual assets returns towards the mean. CH defines dispersions as the cross-sectional standard deviation of returns:

$$CSSD = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n-1}},$$

where r_i is the individual stock return of stock i and $\bar{r}$ is the cross-sectional average of the n returns within the market portfolio. CH as such suggests that during periods of 'market stress', defined as abnormally high or low aggregate market returns, we expect investors are more likely to herd, and thus dispersions will decrease in herding during these periods. They define a regression specification to capture this relationship,

$$CSSD_t = \alpha + \beta_1 D_t^L + \beta_2 D_t^U + \varepsilon_t,$$

where D_t^L and D_t^U are binary indicators capturing when market returns fall within the lower and upper tails of the distribution, respectively. Within CH's regressions, they set these tails both at the 5% and 1% criteria arbitrarily. Crucially to the explanatory power of this method, this prediction is in complete contradiction to that of rational asset-pricing models. These models dictate that dispersions will increase as market returns reach the tails of the distribution, as individual assets vary in degrees of sensitivity to overall market returns. As such, significant negative values of β_1 and β_2 contradict rational asset-pricing models and indicate herding is present within a market. CH carries out an analysis of herding within ordinary common shares traded in US equity markets, breaking these shares into industry groups to determine the presence of industry-based herding. CH estimates a significantly positive β_1 and β_2 for all industries and overall, indicating herding is not present within the market. This result is consistent when using both daily and monthly data, failing to confirm CH's hypothesis that the use of daily data may restrict the type of herding that can be observed[2].

Building upon this approach, Chang *et al.* (2000) (herein CCK) asserts that CH's methodology is too restrictive, consequently constructing their own methodology. CCK argued that not only do rational asset-pricing models predict that the relationship between dispersion and returns are positive, but necessarily linear. CCK demonstrates this by constructing the expected cross-sectional deviation of stock returns ($ECSAD_t$), based on the difference between asset returns and conditional CAPM predictions, showing that the dispersion measure's second-order derivative in expected market returns, $E_t(R_{m,t})$, is equal to 0:

$$ECSAD_t = \frac{1}{N}\sum_{i=1}^{N}|\beta_i - \beta_m|E_t(R_m - \gamma_0),$$

where β_i is the systematic risk measure of security i within the market portfolio, m, of size N, γ_0 is the return of the zero-beta (risk-free) portfolio and E_t is the expectation operator at time t. Whilst this relationship holds in theory, given the presence of expectation-based variables such as $E_t(R_{m,t})$, we cannot observe nor measure this mechanism in reality, as such. To combat this issue, CCK constructs the measure cross-sectional absolute dispersion,

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|R_{i,t} - R_{m,t}|,$$

which proxies for the unobservable $E_t(R_{m,t})$, as the measure of dispersion the paper utilises empirically. This established positive and linear relationship between $E_t(R_{m,t})$ and $ECSAD_t$ is exploited to construct speci-

[2] Richards (1999) suggests that the highest-frequency data available should always be used when studying idiosyncratic variance.

fications that capture a continuous measure of herding, rather than relying on arbitrary binary indicators as per CH:

$$CSAD_t = \alpha + \gamma_1^{UP}|R_{m,t}^{UP}| + \gamma_2^{UP}(R_{m,t}^{UP})^2 + \varepsilon_t, \quad CSAD_t = \alpha + \gamma_1^{DOWN}|R_{m,t}^{DOWN}| + \gamma_2^{DOWN}(R_{m,t}^{DOWN})^2 + \varepsilon_t,$$

where $|R_{m,t}^{UP}|$ and $|R_{m,t}^{DOWN}|$ are the absolute value of the equally-weighted market portfolio in period t when the market is up and down, respectively. This split of up and down markets allows for asymmetric herding in market conditions to be captured. The inclusion of a squared returns term, crucially, allows observation of a non-linear relationship between dispersion and market returns - as captured in a negative and statistically significant γ_2. Importantly, CH's specification requires a significantly greater magnitude of non-linearity in return dispersion than is suggested by conditional CAPM, and that is captured within CCK's specification. CCK employs both of these specifications and a dummy-distribution approach akin to CH for US, Hong-Kong, Japanese, South Korean and Taiwanese equity markets for sample periods varying from 1963 to 1997. The dummy regression results corroborate CH's findings for the US, and all models fail to find results consistent with herding for the US, Hong Kong and Japan. CCK suggests that the insignificant γ_2 in these samples confirms the validity of the linear dispersion-returns relationship. However, the two developing markets within the sample, South Korea and Taiwan, generate significantly negative γ_2^{UP} and γ_2^{DOWN}, suggesting a breakdown of this linear relationship and indicating herding is present in these markets. Specifically, CCK observes a γ_2 ranging between -4.03 and -5.63 for these two countries. CCK suggests this relationship stems from the increased role macroeconomic information has in these economies, given their size. Interestingly, the adjusted R^2 for these two samples are, on average, significantly higher than those for developed markets. This suggests that unsystematic risk has a greater impact on the variation in dispersion within the model than systematic risk, which supports the validity of the herding relationship that CH and CCK base their methodology on.

The approach put forth by these papers is not free from criticism, however. Hirshleifer & Teoh (2003) comment that CH's assertion that herding will be more likely to materialise during periods of 'market stress' is not an obvious one - a fair observation given CH's admittedly arbitrary threshold for what constitutes such stress. Furthermore, Richards (1999) criticises dispersion measures, such as CSSD, as a measure of idiosyncratic variance for their susceptibility to omitted variable bias. Likewise, Richards also comments that equally weighted measures such as CH's $\bar{r}$ will yield dispersions that are dominated by high-variance smaller stocks -

potentially failing to capture herding contributed by larger, more economically significant stocks. Studies such as Chiang *et al.* (2010) that calculate both equally-weighted and value-weighted market portfolios, however, report very similar results in practice. Chiang *et al.* (2010) do contend that the use of OLS within these papers is likely inefficient, given that the error distribution does not conform to Gaussian requirements - the presence of extreme outliers can significantly affect the tails of the distribution and produce bias variance estimators. Even after correcting for such a bias, OLS is inefficient. As such, Chiang suggests utilising a quantile regression, which produces estimates of the γth conditional-quantile by minimizing weighted deviations from the conditional quantile function. This not only produces a more efficient estimator in the presence of autocorrelative errors, but also allows observation of the herding relationship at different points within the returns dispersion distribution.

There are two papers that pioneered the use of an alternative measure of herding, utilising investor-level micro-data: Lakonishok *et al.* (1992) and Sias (2004). Lakonishok (herein LSV) proposed that herding could be detected in a market if there is a degree of correlation across investors in buying and selling a given stock, or in other words ending up on the same side of the market. The paper assessed end-of-quarter holdings for 769 U.S. all equity tax-exempt funds, primarily pension funds, run by 341 institutional money managers between 1985 and 1989. For each stock i held within a given quarter t, LSV calculates the measure of herding $H(i)$,

$$H(i) = |\frac{B(i)}{B(i)+S(i)} - p(t)| - AF(i),$$

where $B(i)$ is the number of money managers who increase their holdings in the stock in a quarter (net buyers), $S(i)$ is the net sellers, $p(t)$ is the expected proportion of money managers buying in that quarter relative to the number active $(\frac{1}{N}\sum_{i=1}^{N}\frac{B(i)}{B(i)+S(i)})$, and $AF(i)$ is an adjustment factor - the expected value of the first equation term under the null hypothesis of no herding. The construction of this measure aims to capture the magnitude of the difference in the ratio of buyers to active money managers (the first term) and the aggregate, or expected, ratio across the whole market in a given period as defined by $p(t)$. This is adjusted with $AF(i) = E[(|\frac{B(i)}{S(i)+B(i)} - p(t)|)|H_0]$, declining in the number of money managers active in a given stock. This measure is computed for each i and averaged across sub groups. If herding is present within a sub group, we would expect a significantly different number of investors to end up on the same side of the market, i.e. a significantly positive value of $H(i)$. Crucially, LSV claims that the sample of money managers does not constitute a random sample, an as such we may expect these investors who compete for the same customers to follow others' signals. LSV finds an average of $H(i) = 0.027$, implying that if $p(t) = 0.5$ then 52.7% of investors would end up on the same side of the market of an average stock. This is a notably small effect, unsurprising given 95% of the sample's trading volume was concentrated in the largest stocks where herding rarely occurred.

LSV has some significant drawbacks in comparison to CCK and CH's measures, however. As observed by Bikhchandani & Sharma (2000), the nature of the sample means that too few stocks are actually changing hands within each quarter, which contributes to the low herding magnitude. additionally, volume is not considered within the model at all. More significantly, the lengthy time periods mean that intertemporal herding, any herding occurring within a time frame smaller than a quarter, cannot be captured.

Herding within Markets: CH and CCK's methodology emerged as the dominant measure of herding within the empirical literature given its replicability. The first paper to apply this measure to Chinese equity markets was Demirer & Kutan (2006) (herein DK). DK hypothesised that the weak rule of law and high levels of government involvement would result in additional volatility (Su & Fleisher, 1998), and lead investors to rely more on well-informed "government insiders". As such, DK suggested herding was likely in Chinese markets. Employing CH's methodology and specification with a sample of daily returns from 375 equities listed on the Shanghai and Shenzhen Stock Exchange from 1999 to 2002, however, DK found no evidence that herding was present within the market.

These results would be contradicted by Tan *et al.* (2008), this time utilising CCK's specification. Chinese A-share markets are dominated by domestic individual investors, whereas B-shares are primarily for foreign investors. Tan, suggesting that individual A-share investors lacked significant knowledge of investments, exploited this differential to determine whether these two groups differed in herding behaviour. Using a sample period from 1997 to 2003 of daily returns for dual-listed Shanghai A and B and Shenzhen A and B stocks, Tan found significantly negative $R^2_{m,t}$ coefficients for both A and B share markets. Not only does this contradict DK's findings, it also fails to confirm CCKs finding that investors in developed markets do not display herding, given herding is present in B-share markets where international investors dominate. Tan suggests that developed market participants may exhibit different behavioural tendencies in their own markets compared to international markets, with international investors relying more on public information in markets outside of their own. This theory is substantiated by Kim & Wei (2002), which using a variation on LSV's herding measure found non-resident investors herd more than resident investors in Korean equity markets. One could suggest that the relative inaccessibility and reliability of fundamental information on Chinese equities would bolster this uncertainty. However, Tan uses a much smaller sample of 87 dual-listed firms compared to DK, with both samples covering similar time periods. Therefore it is possible that variation in the results of the two papers stems from differing sample characteristics rather than differing methodologies.

Likewise, Chiang & Zheng (2010) found significant herding within Chinese equity markets between 1988 and 2009, alongside significant herding within almost all of the 18 country sample excluding the U.S. and Latin America. The study suggested that this exception may be due to Wall Street's role as the primary disseminator of trading strategies and information, with markets outside of the Americas following the US' lead. Interestingly, Chiang and Zheng further this hypothesis by suggesting CCK's specification is inappropriate for a global financial system that is likely to be highly integrated with the US. As such, the paper expands CCK's specification by adding US $R^2_{m,t}$ as an argument within the right-hand side of the specification, in effect controlling for US market conditions. Interestingly for this paper's purposes, herding in China becomes insignificant following this addition, suggesting that CCK's original specification suffers from omitted-variable bias, overstating herding effects. The US $R^2_{m,t}$ coefficient is also observed to be significantly negative within the Chinese sample, as with almost all other non-US markets, suggesting these markets form herding behaviour in-line with the US.

Given the association between herding behaviour and financial instability (Avery & Zemsky, 1998), many papers have found herd behaviour is present during asset price bubbles (Bekiros *et al.*, 2017) and financial events such as the 2008 Global Financial Crisis (Chang *et al.*, 2020; Chiang & Zheng, 2010) and more recently the COVID-19 pandemic (Wu *et al.*, 2020; Yarovaya *et al.*, 2021).

Of particular note to this paper, Kizys *et al.* (2021) investigate the effect of government responses to COVID-19 in 72 stock markets from the 1st of January 2020 to the 31 of March 2020, utilising CCKs methodology and policy stringency data from the Oxford Covid-19 Government Response Tracker. The paper also implements the VIX within the regression as a control for global uncertainty, although this does not alter the results. The paper finds a statistically positive relationship between policy stringency and herding behaviour globally, associating a 10-point increase in the index with an increase in CSAD of 0.174 percentage points. Interestingly, policy stringency impacted herding behaviour in the APAC region significantly less than others, with greater baseline herding.

3 Empirical Research

3.1 Econometric Theory

Data: We view Chinese A-share markets as the ideal environment to study the effect of government responses on herd behaviour, as China imposed significantly harsher COVID-19 lockdown policies than the global average (Hale *et al.*, 2021) and A-share markets are dominated by domestic investors, with foreign investors accounting for only 7.3$ of share ownership (UBS, 2020). Financial data used in this paper was obtained from TuShare, a China-based community-run API that provides Chinese market data. From this API, we retrieve the daily closing price of each individual stock comprising the CSI 300 index, the largest market capitalisation equities from the Shanghai Stock Exchange and Shenzhen Stock Exchange. We use daily data following Christie & Huang (1995)'s observation that "herd behavior is a very short-lived phenomenon", which has been confirmed within the literature (Tan *et al.*, 2008). We use the CSI 300 composition as of the beginning of our sample, 3 January 2020, in order to mitigate the effect of survivorship bias occurring from stocks dropping out of the index. Studying the same mix of equities for the entirety of the sample period maintains the power of our analysis. Data is collected from 22nd January 2020 to 30th December 2022, which results in 714 days of observations after eliminating non-trading days. This sample period was chosen to focus on the effects of variation in severity of 'lockdown-style' containment measures rather than the introduction of such measures.

In addition to market data, we collect data within the same period on the Chinese government response to COVID-19 from the Oxford Covid-19 Government Response Tracker (OxCGRT), as produced by the Blavatnik School of Government at the University of Oxford (Hale *et al.*, 2021). In particular, we utilise the Stringency Index within our main specification, which seeks to record the severity of 'lockdown-style' policies that primarily restricts people's behaviour and movement. We also gather COVID-19 daily case data for China from the Our World in Data COVID-19 database.

Methodology: We adopt the approach proposed by Christie & Huang (1995) (CH), utilising the relationship between dispersion and market returns as our measure of herding, due to its replicability and ability to capture herding within short time periods. However, we have chosen to utilise Chang *et al.* (2000)'s (CCK) alternative specification, namely that our specification's measure of dispersion is the cross-sectional absolute deviation (CSAD), expressed as

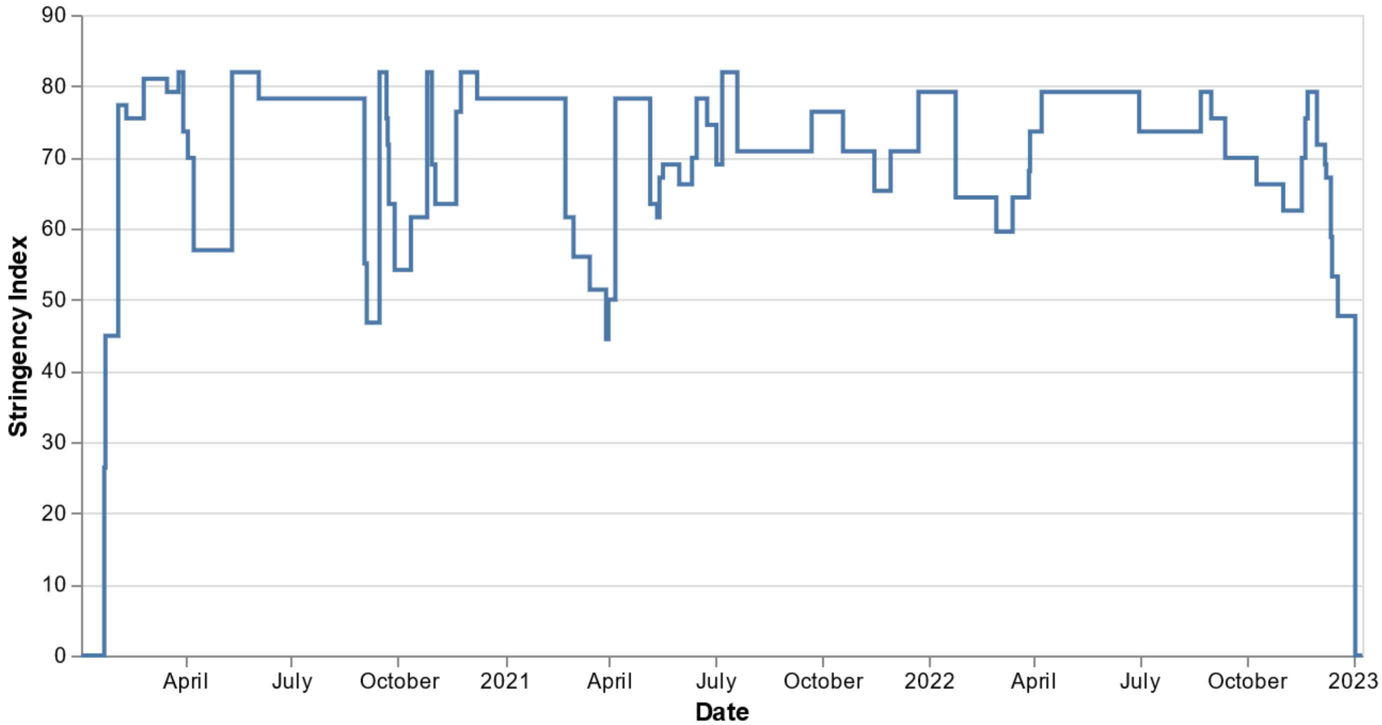

Figure 1: Stringency Index value between the 1st January 2020 and 1st January 2023

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|R_{i,t} - R_{m,t}|,$$

where N is the number of equities within our portfolio (being the CSI 300 as of 03/01/2020) $R_{i,t}$ is the observed stock return of equity i at time t and $R_{m,t}$ is the equally-weighted cross-sectional average of the N returns within the portfolio. Following CCK, the foundation of our specification is

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

where γ denotes the model's coefficients and ε_t the error term at time t. This model's rationale is as follows: according to rational asset-pricing models, as individual assets have varying degrees of sensitivity to overall market return, the relationship between dispersions and market returns is linear and increasing. However, if herding is present within the market CCK suggests that this relationship will no longer hold, as investors' conformity will result in a non-linear, decreasing or even a negative relationship between the two variables. As such, if herding is present we would expect a negative γ_2, indicating a non-linear relationship. Conversely, in the absence of herding we would expect γ_1 to be positive and γ_2 to be insignificant.

CSAD benefits from its ability to capture non-linear herding effects, in combination with the inclusion of $R^2_{m,t}$, that *CSSD* is unable to capture, which CCK notes results in CH's approach being overly strict. It is notable that, additionally, *CSAD* provides a better fit for our data compared to *CSSD*, which has been observed in prior literature (Gleason *et al.*, 2004). Whilst observers state concerns regarding *CSAD*'s construction, notably it's reliance on proxying for unobservable variables within conditional CAPM (Yao *et al.*, 2014; Tan *et al.*, 2008), the two measures share a correlation coefficient of $\rho = 0.89$ within our sample and as such are expected to produce similar results. Additionally, as highlighted by Yao *et al.* (2014), the model has high levels of structural multicollinearity regarding $R_{m,t}$ and $R^2_{m,t}$, which will reduce the efficiency of the model's standard errors. Therefore it is important we standardize the $R^2_{m,t}$ variable by subtracting its arithmetic mean, $\bar{R_m}$, removing much of the variable's multicollinearity. The approach of purely subtracting the mean from our variable was chosen over creating a Z-statistic to preserve normal interpretation of the estimated coefficient. Therefore, Model-1 specification is expressed as

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 (R_{m,t} - \bar{R_m})^2 + \varepsilon_t. \tag{1}$$

Following empirical consensus (Yao *et al.*, 2014; Tan *et al.*, 2008; Chiang *et al.*, 2010) we utilise logarithmic daily returns within our calculations of $CSAD_t$ and $R_{m,t}$ such that $R_{i,t} = 100 \times (\ln P_{i,t} - \ln P_{i,t-1})$, where $P_{i,t}$ is the closing price of equity i at time t. This is in order to preserve time-additivity and log-normality, in addition to aiding the interpretation of our model's predicted coefficients.

Extension 1 - The Effect of 'lockdown-style' Policies on Herding: In order to distil China's containment policy response over time, we utilise OxCGRT's 'Stringency Index'. This index calculates a daily average of 9 'lockdown-style' policy indicators that the database records: closing of schools and universities, closing of workplaces, cancelling of public events, limits on gatherings, closing of public transport, shelter-in-place orders, restrictions on internal city/region movements, restrictions on international travel and presence of COVID-19 public information campaigns. Importantly to China, policies that only apply to a particular region rather than nationally are weighted less within the final index. We chose a variation of this Stringency Index that is split based upon how these policies depend on vaccination status, with weightings applied according to the vaccination rate of the population:

$$STRINGENCY_t = [(stringency_{v,t} * W_{v,t}) + (stringency_{nv,t} * W_{nv,t}]/100,$$

where $STRINGENCY_t$ is our specification's measure, $stringency_{v,t}$ is the index score for policies applying to vaccinated individuals at t, $stringency_{nv,t}$ the index score for non-vaccinated based policies, and $W_{v,t}$ and $W_{nv,t}$ are the percentage of the population vaccinated and not-vaccinated respectively. We chose this to more accurately reflect how containment policies apply to China's population at a given time. We have re-scaled the stringency index from 0 to 1, compared to its original form of 0 to 100, in order to aid the interpretation of our estimated coefficients.

As highlighted by Aknin *et al.* (2022), many previous studies utilising 'lockdown-style' policy regressors fail to control for confounding variables such as local infection and mortality rate. Not including these variables within our regression could result in an understatement within our $STRINGENCY_t$ coefficient given previous studies have suggested the pandemic results in lower herding behaviour within China (Wu *et al.*, 2020). As such, our second model takes the following form:

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 (R_{m,t} - \bar{R_m})^2 + \gamma_3 STRINGENCY_t + \Gamma A_t + \varepsilon_t, \tag{2}$$

where A is a matrix of controls with the vector Γ of coefficients. Formally, within our controls we include a 7 day moving average of the number of deaths, and the daily single-vaccination rate. A moving average has been chosen for our mortality rate measure in order to smooth measurement errors present in daily data. This approach has limited power to capture the possible non-linear relationship between containment policies and mental health in herding. It may be that cognitive stress strong enough to affect herding only occurs during demonstrably limiting lockdown policies. We have included multiple alterations to Model 2 that seek to capture this non-linear relationship. Firstly, we have replaced $STRINGENCY_t$ with dummy indicators $S^{j\%}$, taking the value 1 if the Stringency Index value falls within the upper jth percentile of the index's distribution. As such, our Model 2 is expressed as

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 (R_{m,t} - \bar{R_m})^2 + \gamma_3 (R_{m,t} - \bar{R_m})^2 * S_t^{j\%} + \Gamma A_t + \varepsilon_t. \tag{3}$$

We run regressions with an $S^{j\%}$ for the upper 25th, 10th and 5th percentile. This addition is constructed as an interaction term to capture the effect severe lockdowns have on the relationship between $(R_{m,t} - \bar{R_m})^2$ and $CSAD_t$.

Secondly, following Aknin *et al.* (2022) we also evaluate the possible cumulative effect of lockdown-style policies by including the variable $DAYS_t$ within Model 3, taking on the number of days the population were under high-stringency policies:

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 (R_{m,t} - \bar{R_m})^2 + \gamma_3 LOCKDOWN_t + \gamma_4 DAYS_t + \Gamma A_t + \varepsilon_t, \tag{4}$$

where $LOCKDOWN_t$ is the OxCGRT's indicator of stay-at-home policies, (which takes the value 0 if there is no lockdown policy, 1 if it is recommend to not leave the house, 2 if it is a requirement to not leave the house, with exceptions for daily exercise, grocery shopping, and 'essential' trips and 3 if it is a requirement to not leave the house, with minimal exceptions) and $DAYS_t$ is the cumulative number of days that $LOCKDOWN_t = 3$, which resets every time the indicator drops below 3. By using this approach, it becomes possible to monitor the impact of quarantine length-related stress (Brooks *et al.*, 2020) on herding behaviour.

Extension 2 - Time Series Analysis: Our static model is likely subject to issues causing failure of the Gauss-Markov assumptions necessary to maintain efficiency of OLS as an estimator and validity of standard errors. In particular, it is likely our error term suffers from serial correlation, as market dispersion measures such as CSAD and high-frequency market data in general are frequently observed to exhibit high levels of autocorrelation (Chang *et al.*, 2000). As such it is necessary we test for correlations between our model's residuals over time in order to determine if the true disturbances are auto-correlated. It is also prudent to observe the partial autocorrelation function (PACF). The PACF partials out residual correlations of intervening lags between time periods, compared to the autocorrelation function, as such allowing us to observe the level of partial autocorrelation at each time period. Any significant partial-correlations cause us to reject the null that our underlying process can be represented by a moving-average process of MA(0), and instead indicate that variation $CSAD_t$ is comprised of previous time-period values $CSAD$. We also investigate the possibility of our model's estimated auto-regressive component actually being a unit root process, or a random-walk:

$$CSSD_t = \rho CSSD_{t-1} + \ldots + \varepsilon_t,$$

with $\rho = 1$. Such a highly persistent process fails the requirement of stationarity and we will not be able to utilise the Ergodic Theorem nor the Central Limit Theorem necessary for coefficient estimation. As such it is prudent to run an Augmented Dickey & Fuller (1976) test, fitting the model above with an intercept and no drift (after visual inspection of our data). The test is augmented with lags of $\Delta CSAD_t$ in order to control for serially-correlated errors within the test's model. Under the null the parameter for $CSAD_{t-1}$ is equal to 0, equivalent to $\rho = 1$ above. The t values used to test this null do not have standard distribution, however p values can be approximated via Monte Carlo simulations (MacKinnon, 1994) to a sufficient degree of accuracy. As a complementary assumption to stationarity, we will also need to observe the autocorrelation function (ACF) to verify that the series can be described as weakly dependent, i.e. $Corr(CSAD_t, CSAD_{t+h}) \to 0$ with sufficient speed so that, in combination with stationarity, the Ergodic Theorem holds. We confirm the results of the PACF and Augmented Dickey-Fuller test by running a Breusch (1978) Godfrey (1978) test. This is necessary as, given the test utilises the Lagrange Multiplier (LM) statistic, we can conduct a joint test of higher autocorrelation up to a specified level of lags. The test, derived from constrained optimisation, is equivalent to running an OLS regression on a 'restricted' version of our model, or our model specification excluding any auto-correlated errors, to obtain predicted residuals $\hat{\varepsilon}_t$ for all $t \in [0, N]$. $\hat{\varepsilon}_t$ is then regressed upon the remaining predicted residuals and independent variables. It is worth noting that this test is beneficial above the Augmented Dickey-Fuller as the presence of our independent variables within the auxiliary regression means the strict-exogeneity assumption is no longer needed. The LM statistic is constructed such that $LM = (n-p)R^2_{\hat{\varepsilon}} \sim \chi^2_p$, where n is the number of observations, p is the number of autocorrelations and $R^2_{\hat{\varepsilon}}$ the R-squared of the auxiliary regression. This statistic, following a Chi-Squared distribution, tests the joint null that the coefficient of each level of autocorrelation is equal to 0. We also verify our model's errors are normally distributed, necessary for efficiency. We do this visually, observing a histogram of our CSAD variable.

In addition to autocorrelation, heteroskedastic errors within our model pose a threat to the efficiency of OLS as an estimator in addition to the validity of our standard errors. To correct for these factors within our standard errors and ensure the validity of inference tests, we chose to utilise the Newey & West (1987) estimator for our regressions in order to compute heteroskedastic and autocorrelative consistent (HAC) errors. The Newey-West estimator extends the White (1980) formulation, which builds upon the standard OLS variance estimator and produces heteroskedastic consistent errors by estimating the variance at each value of t within our sample via $\hat{\varepsilon}_t^2 x_t x_t'$. The White estimator can be expressed as $Var(\hat{\beta}) = (X'X)^{-1} X' \hat{\Omega}_0 X (X'X)^{-1}$, with:

$$X' \hat{\Omega}_0 X = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon_t}^2 x_t' x_t,$$

where X is a $t \times 1$ matrix of the vectors x_t containing our independent variables at t, $\hat{\beta}$ is a matrix of the model's predicted coefficients, $\hat{\varepsilon}_t$ is the predicted residuals at t and Ω_0 is the variance-covariance matrix for our residuals under no serial correlation. When introducing the possibility of autocorrelation, the White formulation is no longer valid as the covariance values within the variance-covariance matrix can no longer be assumed to be 0. Likewise, we cannot extend this White formulation to be $X'\hat{\Omega}X$, as given $\hat{\Omega} = \hat{\varepsilon}\hat{\varepsilon}'$, this estimator results in a $\widehat{Var(\hat{\beta})} = 0$. The Newey-West estimator proposes a solution to this by applying a Barlett Kernal weighting of $w_l = 1 - \frac{l}{L+1}$ to each autocorrelation within the variance-covariance matrix according to their level of autocorrelation, L. As such, the Newey-West variance estimator is expressed by

$$\widehat{Var_{nw}}(\hat{\beta}) = (X'X)^{-1}X'\hat{\Omega}_{nw}X(X'X)^{-1} = X'\hat{\Omega}_0X + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T} w_l\varepsilon_t\varepsilon_{t-l}(x_tx'_{t-l} + x_{t-l}x'_t).$$

Autocorrelations further away from t are weighted less within our estimator, which converge to 0 under the assumption of weak-dependence, maintaining the efficiency of the estimator by preventing too many estimations of coefficients. In order to calculate how many levels of autocorrelations we take into account, i.e. our truncation parameter L, we observe both the common practice outlined in Greene (2008), dictating that $L \approx T^{1/4}$, and the results of our PACF.

3.2 Empirical Analysis and Results

Table 3.1: Summary statistics

	Sum	Mean	SD	Min	Max	N
CSAD	1,088	1.524	0.360	0.732	2.923	714
CSSD	91	0.128	0.029	0.059	0.251	714
Market Return	-17	-0.024	1.242	-8.977	5.395	714
Market Return Squared	1,100	1.540	4.200	0.000	80.157	714
Stringency Index	511	0.716	0.089	0.264	0.819	714
C6: Shelter in Place Indicator	1,771	2.480	0.891	0.000	3.000	714
Days	23,886	33.454	37.574	0.000	135.000	714
Vaccination Rate	28,221	39.525	42.218	0.000	89.350	714
COVID-19 Deaths 7-Day Rolling Average	27,945	39.139	99.653	0.000	1294.571	714

Table 3.1 shows the summary statistics for the variables within our regressions. CSAD by definition has a lower bound of zero, indicating all individual returns are moving in perfect unison with the market. From the table, we can also observe that China had an average Stringency Index value of 71.6, staying significantly higher than the global average in almost all time periods (Hale *et al.*, 2021). This confirms that China is an ideal market to analyse the effects of highly restrictive 'lockdown-style' policies.

Regarding the statistical tests detailed in Section 3.1, the results from the Bruesch-Godfrey test indicate that there are significant autocorrelations past 20 lags, meaning we reject the null that the data series can be described as an $AR(p)$ or $MA(p)$ process for $0 < p \leq 20$. As such it is appropriate for our regression's standard errors to be corrected utilising Newey-West HAC errors. Observing the PACF, we determine that a truncation parameter of 5 lags is appropriate for our estimator, as this maintains relatively high levels of efficiency whilst accounting for the majority of serial correlation. Likewise, we conclude from the ACF that our series is characterised as weakly-dependent. We also conclude via visual inspection of CSAD's distribution that our model's error terms are normally distributed. Finally, we calculate Augmented Dickey-Fuller t statistic of -12.478, which leads us to reject the null at the 1% significance level that our series is described by a random walk without a drift, and as such we conclude that the series is stationary.

Table 3.2 contains our regression results. Our herding coefficient, $(R_{m,t} - \bar{R_m})^2$, is negative and significant at the 5% level in all panels, indicating that the linear relationship between $CSAD_t$ and $R_{m,t}$ does not hold and herding is present within the CSI 300 for our sample period. We can observe this graphically in Figure 2. This suggests that the herding relationship within Chinese A-share markets found by Tan *et al.* (2008); Chiang & Zheng (2010) holds during the COVID-19 period, which corroborates the findings of Wu *et al.* (2020).

Panel B reports the effect of containment policy stringency, as measured by the Stringency Index, on herding behaviour. The statistically significant coefficient indicates that cross-sectional average dispersion, $CSAD$, increases with stringency of government response. Specifically, an increase of 10 points within the Stringency Index (which ranges from 0 to 100 and has a mean of 71.6) results in a 4.2265 percentage point increase in dispersion. This indicates that a more strict 'lockdown-style' policy response mitigates herding behaviour. We can also see that the inclusion of the Stringency Index, in addition to pandemic controls, improves the explanatory power of the model significantly, with the adjusted R^2 increasing by 51.5% compared to Panel A.

Next, Panels C-E contain the results of the models containing dummy coefficients capturing the top 25, 10 and 5 percentile strictest containment policy observations, respectively. Interestingly, whilst the top 25th percentile indicator follows the results of Panel B, with a statistically positive relationship between $STRINGENCY$ and $(R_{m,t} - \bar{R_m})^2$, Panel E suggests a differing relationship. The coefficient for $S_t^{0.05} \times (R_{m,t} - \bar{R_m})^2$ indicates that a top 5% value of the Stringency Index is associated with a reduction of linearity within the relationship between market return and return dispersion. This relationship can be observed within Figure 3, and as such we conclude that the most stringent 'lockdown-style' government responses actually increase herding, contrary

Table 3.2: Regression Results

	(A)	(B)	(C)	(D)	(E)	(F)
$\|R_{m,t}\|$	0.18341***	0.18214***	0.16315***	0.18059***	0.19307***	0.18281***
	(0.02499)	(0.02437)	(0.02413)	(0.02464)	(0.02424)	(0.02444)
$(R_{m,t} - \bar{R_m})^2$	-0.00936**	-0.00998**	-0.01077***	-0.00954**	-0.01095**	-0.01009**
	(0.00462)	(0.00470)	(0.00321)	(0.00462)	(0.00460)	(0.00476)
$STRINGENCY_t$		0.42265**				
		(0.21290)				
$S_t^{0.25} \times (R_{m,t} - \bar{R_m})^2$			0.01377***			
			(0.00518)			
$S_t^{0.10} \times (R_{m,t} - \bar{R_m})^2$				0.00186		
				(0.00776)		
$S_t^{0.05} \times (R_{m,t} - \bar{R_m})^2$					-0.10850***	
					(0.03308)	
$LOCKDOWN_t$						0.04907*
						(0.02613)
$DAYS_t$						-0.00055
						(0.00065)
Vaccination Rate		-0.00149***	-0.00139***	-0.00144***	-0.00155***	-0.00129**
		(0.00054)	(0.00054)	(0.00054)	(0.00054)	(0.00063)
COVID-19 Deaths		-0.00035***	-0.00044***	-0.00045***	-0.00046***	-0.00036***
		(0.00013)	(0.00012)	(0.00012)	(0.00012)	(0.00012)
Constant	1.37724***	1.14927***	1.45928***	1.45403***	1.45549***	1.34074***
	(0.02616)	(0.15211)	(0.03365)	(0.03340)	(0.03411)	(0.06409)
Adjusted R^2	0.1277	0.1935	0.1892	0.1833	0.1966	0.1916

Standard errors in parentheses and based on Newey and West (1987)'s heteroscedasticity and autocorrelation consistent standard errors. This table reports the estimated coefficients and adjusted R^2 for the models defined in Section 3.1. Panel A contains results for the base model defined in Equation 1: $CSAD_t = \gamma_0 + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t} - \bar{R_m})^2 + \varepsilon_t$. Panel B contains results for the lockdown-policy extension model defined in Equation 2: $CSAD_t = \gamma_0 + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t} - \bar{R_m})^2 + \gamma_3 STRINGENCY_t + \Gamma A_t + \varepsilon_t$. Panels C D & E contain results for the model defined in equation 3: $CSAD_t = \gamma_0 + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t} - \bar{R_m})^2 + \gamma_3(R_{m,t} - \bar{R_m})^2 * S_t^{j\%} + \Gamma A_t + \varepsilon_t$ for the top 25th, 10th and 5th percentile respectively. Panel F contains the results for the model defined in Equation 4: $CSAD_t = \gamma_0 + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t} - \bar{R_m})^2 + \gamma_3 LOCKDOWN_t + \gamma_4 DAYS_t + \Gamma A_t + \varepsilon_t$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

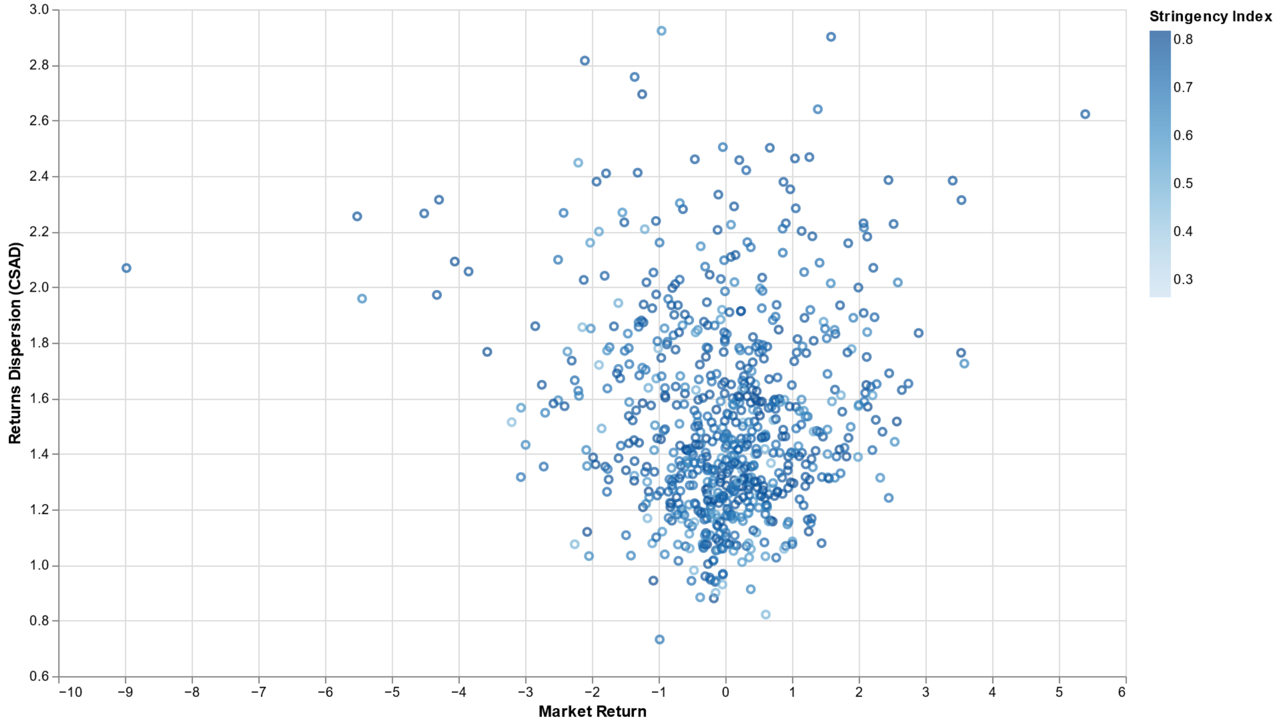

Figure 2: Plot of $CSAD_t$ against $R_{m,t}$ for the stocks within the CSI 300 between 22nd January 2020 and 30th December 2022, with observation colour reflecting Stringency Index value.

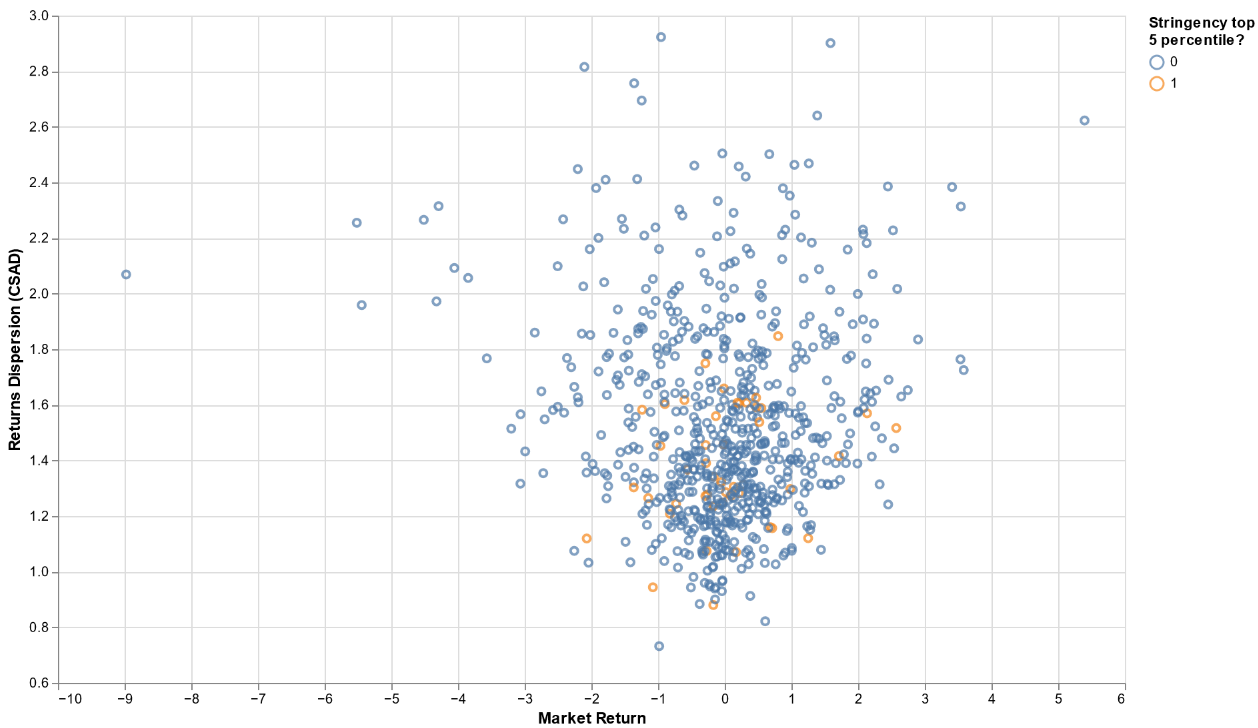

Figure 3: Plot of $CSAD_t$ against $R_{m,t}$ for the stocks within the CSI 300 between 22nd January 2020 and 30th December 2022, with observation colour reflecting whether Stringency Index value falls within top 5 percentile.

Table 3.3: Principle Component Analysis Eigenvectors

Variable	PC_1	PC_2	PC_3
Lockdowns	0.4815	0.2910	-0.2273
Workplace closures	**0.4905**	-0.1111	-0.2472
Public event cancelling	0.3405	-0.3499	0.5167
Restrictions on gatherings	0.3458	-0.2143	**0.5347**
Public transport closure	0.3087	0.2999	0.1682
Movement restrictions between cities/regions	0.4125	0.3259	-0.2459
International travel restrictions	-0.1627	**0.7338**	0.4956

Bold indicates the largest loading factor within principle component. Table contains loading factors of principle components (PC_i) with the requirement of $\lambda \geq 1$.

to less stringent policy responses. Our final panel, Panel F, indicates that the effect of lockdown duration on herding is insignificant.

These results appear to provide evidence of two antagonistic theories regarding how 'lockdown-style' policies affect herding. Firstly, as proposed by Sharif *et al.* (2020) and explored by Kizys *et al.* (2021), we find that stricter policies mitigate herding behaviour, supporting the argument that the addition of quality information from decisive government response reduces multidimensional uncertainty, which in turn we interpret as a decrease in rational herding. This can be likened to increasing the probability that a signal is correct within Banerjee (1992)'s model of herding. Simultaneously, however, the strictest 5% of lockdown-style policies lead to a less-than-proportionate mitigation of herding behaviour, which we argue is due to the countervailing effect of increased investor cognitive stress. Previous studies have found significant links between lockdowns and adverse mental-health effects (Aknin *et al.*, 2022; Dubey *et al.*, 2020), notably cognitive stress, which Prechter (2010) proposes leads to irrational herding. We interpret our results as demonstration of investor resilience to lockdown-induced cognitive stress, with an upper-ceiling of policy stringency where statistically-significant effects start to materialise.

Robustness Analysis: For our primary robustness check, following the methodology of Kizys *et al.* (2021), we utilise principle component analysis (PCA) in order to observe which factors are responsible for the majority of variation within the Stringency Index, which we can use to verify our interpretation of the relationship between policy stringency and herding. PCA is primarily a statistical method of dimensionality reduction, which arranges our random variables, in this case the indicators that compose the Stringency Index [3], into uncorrelated

[3]The construction and details of each of these indicators can be found within the OxCGRT documentation https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md

linear combinations, or PCs. Let X be a vector containing our 7 indicator variables, $X' = [X_1, X_2, ..., X_7]$. We consider an orthogonal arrangement of *uncorrelated* linear combinations of these indicator variables PC_i such that the first combination, PC_1, has maximal variance compared to the other combinations, $(i = 1, 2, ..., 7)$:

$$
\begin{aligned}
PC_1 &= a_1'X = a_{11}X_1 + a_{12}X_2 + ... + a_{17}X_7, \\
PC_2 &= a_2'X = a_{21}X_1 + a_{22}X_2 + ... + a_{27}X_7, \\
&\vdots \\
PC_7 &= a_7'X = a_{71}X_1 + a_{72}X_2 + ... + a_{77}X_7,
\end{aligned}
$$

where a_{ii} are loading factors representing the contribution of each variable to the PC, which is subject to the unit-length condition $a_1'a_1 = 1$ as a form of normalization. Given this transformation, we calculate the eigenvalue-eigenvector pairs (λ_i, e_1), such that $Var(PC_i) = e_i'\Sigma e_i = \lambda_i$, where Σ is the covariance matrix of X. In other words, PC_1 has the largest λ and as such is responsible for the most variance within the Stringency Index, with each subsequent PC possessing less variance. Following the Kaiser criterion, we maintain the top three principle components that have $\lambda \geq 1$, which are responsible for 70% of the variation within the Index. Table 3.3 contains the loading factors for our eigenvectors. As such, we can determine which indicators cluster in their variation. PC_1 can be thought of as an indicator of disruption to daily life, whereas PC_2 is more responsible for variation in restrictions to freedom of travel.

We then replace $STRINGENCY$ within our primary regression for Extension 1, defined in Equation 2, with our three primary components. Our resulting coefficients for the components are:

$$PC_1 : 0.02171^{**}(0.01044), \quad PC_2 : -0.01870(0.02315), \quad PC_3 : -0.01120(0.01941)$$

where standard errors are in brackets and $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. PC_1 is significantly positive, indicating that policies disrupting daily life are most responsible for mitigating herding. We hypothesise this is due to the reduction in multidimensional uncertainty the most visible government policies are responsible for. Our indicator for freedom of travel policies, PC_2, is negative but insignificant, suggesting these policies have little, if any, effect on herding. The same can be said for PC_3. As such, we confirm the results of Panel B within our main regression, yet we also fail to confirm the result of Panel E.

Discussion: Whilst the results of this paper are sufficiently robust, there are limitations to the empirical strategy that are worth identifying for future research. Firstly, it can be observed in Figure 2 that the results from our OLS regressions, especially the estimated coefficients for $(R_{m,t} - \bar{R_m})^2$, are likely overstated due to the existence of outliers, as contended by Chiang *et al.* (2010). As such, it may be prudent to repeat our methodology utilising a quantile regression over OLS in order to test our herding relationship conditioning for differing quantiles. Additionally, the non-persistent nature of our herding behaviour found in Panel E, regarding the strictest 5% of policies, suggests this herding relationship may be spurious, with investors facing similar a decision-set regarding an unobserved event (Bikhchandani & Sharma, 2000). In order to determine this, a study of the highest Stringency Index observations could be undertaken for a range of markets, akin to Kizys *et al.* (2021), to test the cognitive stress hypothesis further. Observing the PACF, we also detect a lone, significant autocorrelation at lag 20, indicating a structural break. The presence of such a break could be determined by utilising a Chow test (Chow, 1960). Finally, in reality 'lockdown-style' policies in China were almost always applied regionally. Whilst the Stringency Index does discount local policies compared to national, this does not reflect the proportion of investors under lockdown at a given time. Future research could be conducted focusing on Shanghai lockdown periods, a centre for financial market participants in China, as an event study.

4 Conclusion

This paper studies the effect of 'lockdown-style' policies, in response to the COVID-19 pandemic, on herding behaviour in Chinese equity markets. Utilising methodology proposed by Chang *et al.* (2000), and implementing a measure of policy response stringency developed by the Oxford COVID-19 Government Response Tracker, we found that herding was present in A-share markets during the pandemic, and containment policies mitigated this behaviour by reducing multidimensional uncertainty associated with COVID-19. However, we also found that the strictest 5% of lockdown policies increased herding behaviour, potentially due to the additional cognitive stress these policies cause to market participants. The majority of the reduction in cross-sectional average deviation (CSAD) of market returns, our herding measure, was a result of the most visible policies that disrupted daily life.

5 Appendix

References

Aknin, Lara B, Andretti, Bernardo, Goldszmidt, Rafael, Helliwell, John F, Petherick, Anna, Neve, Jan-Emmanuel De, Dunn, Elizabeth W, Fancourt, Daisy, Goldberg, Elkhonon, Jones, Sarah P, Karadag, Ozge, Karam, Elie, Layard, Richard, Saxena, Shekhar, Thornton, Emily, Whillans, Ashley, & Zaki, Jamil. 2022. Policy stringency and mental health during the COVID-19 pandemic: a longitudinal analysis of data from 15 countries. *The Lancet*, **7**(5), e417–e426.

Avery, Christopher, & Zemsky, Peter. 1998. Multidimensional uncertainty and herd behavior in financial markets. *American economic review*, 724–748.

Banerjee, Abhijit V. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, **107**(3), 797–817.

Bekiros, Stelios, Jlassi, Mouna, Lucey, Brian, Naoui, Kamel, & Uddin, Gazi Salah. 2017. Herding behavior, market sentiment and volatility: Will the bubble resume? *The North American Journal of Economics and Finance*, **42**, 107–131.

Bikhchandani, Sushil, & Sharma, Sunil. 2000. Herd behavior in financial markets. *IMF Staff papers*, **47**(3), 279–310.

Breusch, T. S. 1978. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, **17**(31), 334–355.

Brooks, Samantha K, Webster, Rebecca K, Smith, Louise E, Woodland, Lisa, Wessely, Simon, Greenberg, Neil, & Rubin, Gideon James. 2020. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The lancet*, **395**(10227), 912–920.

Cesario, Joseph, Johnson, David J., & Eisthen, Heather L. 2020. Your Brain Is Not an Onion With a Tiny Reptile Inside. *Current Directions in Psychological Science*, **29**(3), 255–260.

Chang, Chia-Lin, McAleer, Michael, & Wang, Yu-Ann. 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, Eric C., Cheng, Joseph W., & Khorana, Ajay. 2000. An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, **24**(10), 1651–1679.

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

Chiang, Thomas C., Li, Jiandong, & Tan, Lin. 2010. Empirical investigation of herding behavior in Chinese stock markets: Evidence from quantile regression analysis. *Global Finance Journal*, **21**(1), 111–124.

Choijil, Enkhbayar, Méndez, Christian, Wong, Wing-Keung, Vieito, João Paulo, & Batmunkh, Munkh-Ulzii. 2022. Thirty years of herd behavior in financial markets: A bibliometric analysis. *Research in International Business and Finance*, **59**(3), 101506.

Chow, Gregory C. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, **28**(3), 591–605.

Christie, William G., & Huang, Roger D. 1995. Following the Pied Piper: Do Individual Returns Herd around the Market? *Financial Analysts Journal*, **51**(4), 31–37.

Demirer, Rıza, & Kutan, Ali M. 2006. Does herding behavior exist in Chinese stock markets? *Journal of International Financial Markets, Institutions and Money*, **16**(2), 123–142.

Dickey, David A., & Fuller, Wayne A. 1976. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, **74**(366), 427–431.

Dubey, Souvik, Biswas, Payel, Ghosh, Ritwik, Chatterjee, Subhankar, Dubey, Mahua Jana, Chatterjee, Subham, Lahiri, Durjoy, & Lavie, Carl J. 2020. Psychosocial impact of COVID-19. *Diabetes & Metabolic Syndrome: clinical research & reviews*, **14**(5), 779–788.

ESMA. 2017. *Consultation Paper 13 July 2017*. Tech. rept.

Gawronski, Bertram, & Cesario, Joseph. 2013. Of Mice and Men: What Animal Research Can Tell Us About Context Effects on Automatic Responses in Humans. *Personality and Social Psychology Review*, **17**(2), 187–215.

Gleason, Kimberly C., Mathur, Ike, & Peterson, Mark A. 2004. Analysis of intraday herding behavior among the sector ETFs. *Journal of Empirical Finance*, **11**(5), 681–694.

Godfrey, L. G. 1978. Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, **46**(6), 1293–1301.

Greene, William H. 2008. *Econometric Analysis*. 6 edn. Pearson Practice Hall.

Hale, Thomas, Angrist, Noam, Goldszmidt, Rafael, Kira, Beatriz, Petherick, Anna, Phillips, Toby, Webster, Samuel, Cameron-Blake, Emily, Hallas, Laura, Majumdar, Saptarshi, & Tatlow, Helen. 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, **5**, 529–538.

Hawryluck, Laura, Gold, Wayne L, Robinson, Susan, Pogorski, Stephen, Galea, Sandro, & Styra, Rima. 2004. SARS control and psychological effects of quarantine, Toronto, Canada. *Emerging infectious diseases*, **10**(7), 1206.

Heyes, Cecilia. 2011. Automatic imitation. *Psychological bulletin*, **137**(3), 463.

Hirshleifer, David, & Teoh, Siew Hong. 2003. Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis. *European Financial Management*, **9**(1), 25–66.

Kidd, Celeste, Palmeri, Holly, & Aslin, Richard N. 2013. Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, **126**(1), 109–114.

Kim, Woochan, & Wei, Shang-Jin. 2002. Foreign portfolio investors before and during a crisis. *Journal of International Economic*, **56**, 77–96.

Kizys, Renatas, Tzouvanas, Panagiotis, & Donadelli, Michael. 2021. From COVID-19 herd immunity to investor herding in international stock markets: The role of government and regulatory restrictions. *International Review of Financial Analysis*, **74**, 101663.

Lakonishok, Josef, Shleifer, Andrei, & Vishny, Robert W. 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics*, **32**(1), 23–43.

MacKinnon, James G. 1994. Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests. *Journal of Business & Economic Statistics*, **12**(2), 167–176.

Macklin, Ruth. 1978. Man's ”Animal Brains” and Animal Nature: Some Implications of a Psychophysiological Theory. *Philosophy and Phenomenological Research*, **39**(2), 155–181.

Newey, Whitney K., & West, Kenneth D. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**(3), 703–708.

Olsen, Robert A. 1996. Implications of herding behavior for earnings estimation, risk assessment, and stock returns. *Financial Analysts Journal*, **52**(4), 37–41.

Prechter, Robert R. 2010. Unconscious Herding Behavior as the Psychological Basis of Financial Market Trends and Patterns. *Journal of Psychology and Financial Markets*, **2**(3), 120–125.

Richards, Anthony J. 1999. Idiosyncratic Risk: An Empirical Analysis, with Implications for the Risk of Relative-Value Trading Strategies. *IMF Working Paper*, **1999**(148).

Rubinstein, Donald H, & Gajdusek, Daniel Carleton. 1970. *A Study in nascent literacy: neo-Melanesian correspondence from a Fore, New Guinea youth*. Section of Child Growth and Development and Disease Patterns in Primitive Cultures, National Institute of Neurological Diseases and Stroke, National Institutes of Health.

Scharfstein, David, & Stein, Jeremy. 1990. Herd Behavior and Investment. *American Economic Review*, **80**(3), 465–479.

Sharif, Arshian, Aloui, Chaker, & Yarovaya, Larisa. 2020. COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *International review of financial analysis*, **70**, 101496.

Spyrou, Spyros. 2013. Herding in financial markets: a review of the literature. *Review of Behavioural Finance*, **5**(2), 175–194.

Su, Dongwei, & Fleisher, Belton M. 1998. Risk, Return and Regulation in Chinese Stock Markets. *Journal of Economics and Business*, **50**(3), 239–256.

Tan, Lin, Chiang, Thomas C., Mason, Joseph R., & Nelling, Edward. 2008. Herding behavior in Chinese stock markets: An examination of A and B shares. *Pacific-Basin Finance Journal*, **16**, 61–77.

UBS. 2020. *China A-shares: FAQs, facts, and figures*. https://www.ubs.com/global/en/assetmanagement/insights/thematic-viewpoints/apac-and-emerging/articles/stock-connect-china-a-shares-faqs-equity-investing.html [accessed: 09.03.2023].

Welch, Ivo. 1992. Sequential sales, learning, and cascades. *The Journal of finance*, **47**(2), 695–732.

White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48**(4), 817–838.

Wu, Guosong, Yang, Boxian, & Zhao, Ningru. 2020. Herding Behavior in Chinese Stock Markets during COVID-19. *Emerging Markets Finance and Trade*, **56**(15), 3578–3587.

Yao, Juan, Ma, Chuanchan, & He, William Peng. 2014. Investor herding behaviour of Chinese stock market. *International Review of Economics & Finance*, **29**(January), 12–29.

Yarovaya, Larisa, Matkovskyy, Roman, & Jalan, Akanksha. 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.

5.1 Econometric Specification Codebook

This codebooks's purpose is to create a timeseries dataset of the primary regression specification used within this Economics dissertation. All of the data files and notebooks can be found at this dissertation's GitHub page: https://github.com/SLBlundell/Economics-Dissertation/.

The following code creates the methods necessary to construct the regression specification. Namely, it calculates the regression specification's variables. These variables are placed within a time series dataset, with each row representing the specification for a given value of t.

```
import pandas as pd
import numpy as np
from typing import List, Dict


class DataProcessor:
    """
    Represents the processing of given data into the primary regression specification
    """

    def __init__(self, returns_file: pd.DataFrame, stringency_file: pd.DataFrame,
    covid_file: pd.DataFrame) -> None:
        self.df: pd.DataFrame = returns_file
        self.df_stringency: pd.DataFrame = stringency_file
        self.df_covid: pd.DataFrame = covid_file
        """
        Parameters
        ----------
        returns_file : DataFrame
            DataFrame containing market data for selected equities received from the
            TuShare API
        stringency_file : DataFrame
            DataFrame containing dates and corresponding stringency index for China
        """

    def sort_data(self) -> None:
        """Sorts data based on equity code, then ascending date"""
        self.df = self.df.sort_values(by=['ts_code', 'trade_date']).reset_index(drop=True)

    def calculate_log_returns(self) -> None:
        """Calculates daily logarithmic returns for all dates"""
        close = pd.to_numeric(self.df['close'], errors='coerce').astype('float')
        self.df['daily_log_return'] = 100 * (np.log(close) - np.log(close.shift(1)))
        self.df.loc[self.df['trade_date'] == '2020-01-02', 'daily_log_return'] = np.nan

    def calculate_return_market(self, date: int) -> float:
        """Calculates the market return, being the equally-weighted arithmetic mean of
        individual equity log-returns on a given date"""
        return np.mean(self.get_returns_on_date(date))
```

```python
def calculate_CSAD(self, date: int, R_m: float) -> float:
    """Calculates CSAD on a given date"""
    returns_on_date = self.get_returns_on_date(date)
    return (1/self.df['ts_code'].nunique()) * np.sum(np.abs([x - R_m
    for x in returns_on_date]))

def calculate_CSSD(self, date: int, R_m: float) -> float:
    """Calculates CSSD on a given date"""
    returns_on_date = self.get_returns_on_date(date)
    return (1/(self.df['ts_code'].nunique() - 1)) * np.sum([(x - R_m) ** 2
    for x in returns_on_date]) ** (1/2)

def get_returns_on_date(self, date: int) -> List[float]:
    """Returns list of all logarithmic equity returns on a given date"""
    return self.df.loc[self.df['trade_date'] == date, 'daily_log_return'].tolist()

def process_data(self):
    """Iterates through all unique dates within datasets, creating a row of all
    variables within the regression specification,
    and appending the main specification dataset with said rows"""
    self.calculate_log_returns()

    stringency_dict = self.df_stringency.set_index('Date').to_dict()
    cols_to_extract = ['StringencyIndex_WeightedAverage',
                       'C6E_Stay at home requirements',
                       'PopulationVaccinated',
                       'C2E_Workplace closing',
                       'C3E_Cancel public events',
                       'C4E_Restrictions on gatherings',
                       'C5E_Close public transport',
                       'C7E_Restrictions on internal movement',
                       'C8E_International travel controls']
    stringency_dict = {col_name: stringency_dict.get(col_name, {})
    for col_name in cols_to_extract}

    rows = []
    for date in self.df.trade_date.unique():
        R_m = self.calculate_return_market(date)
        CSAD = self.calculate_CSAD(date, R_m)
        CSSD = self.calculate_CSSD(date, R_m)
        stringency = {col_name: stringency_dict[col_name].get(date, np.nan)
        for col_name in cols_to_extract}
        covid = self.df_covid.loc[self.df_covid['date'] == date, 'rolling_deaths']
        if not covid.empty:
            covid = covid.iloc[0]
        else:
            covid = 0
```

```python
            new_row = {'date': date, 'R_m': R_m, 'CSAD': CSAD, 'CSSD': CSSD,
                       **stringency, 'covid_deaths': covid}
            rows.append(new_row)

        self.df_spec = pd.DataFrame(rows)

    def save_data(self, output_file):
        """Saves the dataset of timeseries data to a comma-deliminated file (csv)"""
        self.df_spec.to_csv(output_file)
```

Here we gather our Chinese equity market data, utilizing the China-based community run API TuShare https://tushare.pro/.

The daily_log_return column of the resultant dataset represents logarithmic returns within our variables, with *t* being the trade_date and *i* the ts_code.

BE WARNED: Due to the community nature of the TuShare API, only 6000 values can be requested at once for most users. As such, the method below splits the data requests into 20-day batches, which results in a considerably long runtime for the whole 3-year period.

Note: config.api_key must be replaced with your own API key, which can be obtained from TuShare by signing up to the platform using the link provided above.

```python
import tushare as ts
import re
import config

class GetChineseEquityData:
    """Represents code necessary to retrieve data from TuShare API"""

    def __init__(self, list_path, start_date, end_date):
        """
        Parameters
        ----------
        list_path : Literal
            The list of equities you are retrieving data for. Data should be csv file,
            containing the column name "code" with all your equity codes and "exchange"
            containing the equity's exchange.
        start_date : int
            The start date of the period you want to retrieve data for
        end_date : int
            The end date of the period you want to retrieve data for
        """

        self.list_path = list_path
        self.start_date = start_date
        self.end_date = end_date
        self.df_codes = pd.read_csv(list_path)
        ts.set_token(config.api_key)
        self.pro = ts.pro_api()

    def get_codes(self):
```

```
        """Returns a string of codes separated by commas."""
        return ','.join([
            f"{row['code'][0:6]}.SH" if row['exchange'] == 'Shanghai'
            else f"{row['code'][0:6]}.SZ"
            for _, row in self.df_codes.iterrows()
        ])

    def get_returns_data(self):
        """Retrieves market data for all equities in the list within the specified
        date range using batch queries."""
        codes = self.get_codes()
        t = self.start_date
        appended_data = []

        while t < self.end_date:
            data = self.pro.daily(ts_code=codes, start_date=str(t), end_date=str(t + 20))
            appended_data.append(data)
            t += 21

        appended_data = pd.concat(appended_data)
        return appended_data
```

Defining the list of equities and timeframe:

```
csi_list_path = r'data/csi_constituents.csv'
data_components = GetChineseEquityData(csi_list_path, 20200101, 20230101)
returns_data = data_components.get_returns_data()
```

Getting the COVID-19 Government Response data from the Oxford COVID Policy Tracker's GitHub https://github.com/OxCGRT/covid-policy-tracker:

```
import requests
import io
import pandas as pd
import numpy as np

class GetStringencyData:

    def __init__(self, years):
        """
        Parameters
        ----------
        years : List
            List of years you want data retrieved for, in string format
        """

        self.url_base = "https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker" +
        "/master/data/OxCGRT_nat_differentiated_withnotes_{}.csv"
        self.years = years
```

```
        self.stringency_data = pd.DataFrame(columns=['Date',
                              'StringencyIndex_WeightedAverage'])

    def get_data(self, year):
        """requests HTML content from the OxCGRT GitHub, returns content in a DataFrame"""
        url = self.url_base.format(year)
        data = requests.get(url).content
        df = pd.read_csv(io.StringIO(data.decode('utf-8')))
        return df

    def filter_data(self, df):
        """Filters data so that only dates relevant variables for China is included"""
        df = df.loc[df['CountryName'] == 'China']
        df = df[['Date', 'StringencyIndex_WeightedAverage', 'C6E_Stay at home requirements',
            'PopulationVaccinated']].replace(np.nan, 0)
        return df

    def concatenate_data(self, df):
        """Combines data for each year specified into one DataFrame"""
        self.stringency_data = pd.concat([self.stringency_data, df]).reset_index(drop=True)

    def get_and_filter(self):
        """Iterates over each year specified, retrieving and filtering data from OxCGRT"""
        for year in self.years:
            df = self.get_data(year)
            df = self.filter_data(df)
            self.concatenate_data(df)

        return self.stringency_data

sd = GetStringencyData(["2020", "2021", "2022", "2023"])
stringency_data = sd.get_and_filter()
```

Requesting and filtering world COVID-19 data provided via https://github.com/owid/covid-19-data.

```
import requests
import io
import pandas as pd
import numpy as np

class GetCovidData:

    def __init__(self):
        """
        Parameters
        ----------
        years : List
            List of years you want data retrieved for, in string format
        """
```

```
        self.url_base =
        f"https://raw.githubusercontent.com/owid/covid-19-data/
        master/public/data/cases_deaths/{}.csv"

    def get_data(self, url):
        """requests HTML content from the Our World in Data GitHub,
        returns content in a DataFrame"""
        data = requests.get(url).content
        df = pd.read_csv(io.StringIO(data.decode('utf-8')))
        df = df[['date', 'China']]
        return df

    def get_and_calculate(self):
        """Gets and calculates 7 day rolling average of deaths"""
        df_deaths = self.get_data(self.url_base.format("new_deaths"))
        df_deaths['rolling_deaths'] = df_deaths['China'].rolling(7).mean()
        df_deaths = df_deaths[['date', 'rolling_deaths']]
        """Gets cases"""
        df_cases = self.get_data(self.url_base.format("new_cases"))
        df_cases = df_cases.rename(columns={'China': 'cases'})
        df_cases = df_cases[['cases']].replace(np.nan, 0)

        df = pd.concat([df_deaths, df_cases], axis=1)
        return df

cd = GetCovidData()
covid_data = cd.get_and_calculate()
```

Finally, we can run our data processor, which returns a dataset of rows representing our primary specification at each value of t.

```
data_processor = DataProcessor(returns_data, stringency_data, covid_data)
data_processor.sort_data()
data_processor.process_data()
data_processor.save_data('data/spec_1_stringency_CSSD_CSAD_expanded.csv')
```

5.2 Stata Do-File

```
clear all
capture log close
log using "section_3-2.log", replace

import delimited "./data/spec_1.csv"

// Data Cleaning //

replace stringencyindex_weightedaverage = 0 if missing(stringencyindex_weightedaverage)

sort date
drop if stringencyindex_weightedaverage==0
gen time=_n

// Generating Regression Variables //

egen r_m_mean = mean(r_m)
gen r_m_sqr= (r_m-r_m_mean)^2
gen r_m_abs=abs(r_m)
replace stringencyindex_weightedaverage=stringencyindex_weightedaverage/100

tempvar PL75
egen `PL75' = pctile(stringencyindex_weightedaverage), p(75)
gen D25Upper = 0
label var D25Upper "S upper 25

replace D25Upper = 1 if stringencyindex_weightedaverage >= `PL75'

tempvar PL90
egen `PL90' = pctile(stringencyindex_weightedaverage), p(90)
gen D10Upper = 0
label var D10Upper "S upper 10

replace D10Upper = 1 if stringencyindex_weightedaverage >= `PL90'

tempvar PL95
egen `PL95' = pctile(stringencyindex_weightedaverage), p(95)
gen D5Upper = 0
label var D5Upper "S upper 5

replace D5Upper = 1 if stringencyindex_weightedaverage >= `PL95'

gen days = 0
label var days "Days"
replace days = days[_n-1] + 1 if c6e_stayathomerequirements == 3
replace days = 0 if c6e_stayathomerequirements < 3
```

```
// Labelling Variables //

la var r_m "Market Return"
la var stringencyindex_weightedaverage "Stringency Index"
la var c6e_stayathomerequirements "C6: Shelter in Place Indicator"
la var populationvaccinated "Vaccination Rate"
la var rolling_deaths "COVID-19 Deaths 7-Day Rolling Average"
la var cases "COVID-19 Daily Cases"
la var r_m_sqr "Market Return Squared"
la var r_m_abs "Absolute Market Return"

tsset time

outsheet csad cssd r_m r_m_abs r_m_sqr stringencyindex_weightedaverage
c6e_stayathomerequirements D25Upper D10Upper D5Upper days populationvaccinated
rolling_deaths cases using ./data/spec_1_stata.csv , comma replace

// Summary Stats //

estpost tabstat csad cssd r_m r_m_sqr stringencyindex_weightedaverage
c6e_stayathomerequirements days populationvaccinated rolling_deaths,
c(stat) stat(sum mean sd min max n)

esttab using ".\TeX_files\SummaryTable.tex", replace cells("sum(fmt(%6.0fc))
mean(fmt(%6.3fc)) sd(fmt(%6.3fc)) min(fmt(%6.3fc)) max(fmt(%6.3fc)) count")
nonumber nomtitle nonote noobs label booktabs collabels("Sum" "Mean" "SD" "Min" "Max" "N")

// Statistical tests //

pac csad
ac csad
hist csad

quietly reg csad r_m_abs r_m_sqr stringencyindex_weightedaverage populationvaccinated
rolling_deaths
estat bgodfrey, lags(1 2:20)
dfuller csad
swilk csad

// Regressions //

eststo: newey csad r_m_abs r_m_sqr, lag(5)
eststo: newey csad r_m_abs r_m_sqr stringencyindex_weightedaverage populationvaccinated
rolling_deaths, lag(5)
eststo: newey csad r_m_abs r_m_sqr c.r_m_sqr#D25Upper populationvaccinated
rolling_deaths,lag(5)
eststo: newey csad r_m_abs r_m_sqr c.r_m_sqr#D10Upper populationvaccinated
rolling_deaths,lag(5)
```

```
eststo: newey csad r_m_abs r_m_sqr c.r_m_sqr#D5Upper populationvaccinated
rolling_deaths,lag(5)
eststo: newey csad r_m_abs r_m_sqr c6e_stayathomerequirements days populationvaccinated
rolling_deaths,lag(5)


esttab, b(5) se(5) nomtitle label star(* 0.10 ** 0.05 *** 0.01)
esttab using "./TeX_files/Regressions_1.tex", replace b(5) se(5)
nomtitle label star(* 0.10 ** 0.05 *** 0.01) booktabs
title("Regression Results \label{reg1}") addnotes("First line" "Second line")
est clear

// Robustness Checks //

pca c6e_stayathomerequirements c2e_workplaceclosing c3e_cancelpublicevents
c4e_restrictionsongatherings c5e_closepublictransport c7e_restrictionsoninternalmoveme
c8e_internationaltravelcontrols, components(3)

predict pc1 pc2 pc3, score

eststo: newey csad r_m_abs r_m_sqr pc1 pc2 pc3 populationvaccinated rolling_deaths, lag(5)

eststo: newey cssd r_m_abs r_m_sqr, lag(5)
eststo: newey cssd r_m_abs r_m_sqr stringencyindex_weightedaverage populationvaccinated
rolling_deaths, lag(5)
eststo: newey cssd r_m_abs r_m_sqr c.r_m_sqr#D25Upper populationvaccinated
rolling_deaths, lag(5)
eststo: newey cssd r_m_abs r_m_sqr c.r_m_sqr#D10Upper populationvaccinated
rolling_deaths, lag(5)
eststo: newey cssd r_m_abs r_m_sqr c.r_m_sqr#D5Upper populationvaccinated
rolling_deaths, lag(5)
eststo: newey cssd r_m_abs r_m_sqr c6e_stayathomerequirements days populationvaccinated
rolling_deaths, lag(5)

esttab, b(5) se(5) nomtitle label star(* 0.10 ** 0.05 *** 0.01)
esttab using "./TeX_files/Regressions_2.tex", replace b(5) se(5)
nomtitle label star(* 0.10 ** 0.05 *** 0.01) booktabs
title("Robustness Results \label{reg2}") addnotes("First line" "Second line")
est clear

log close
```