Internet Appendix for "Retail Trading and Asset Prices: The Role of Changing Social Dynamics"

Fulin Li*

October 3, 2022

Contents

*The University of Chicago, fli3@chicagobooth.edu.

A1 Omitted derivations and proofs

A1.1 Dynamics of wealth shares

Since the risk-free asset is in zero net supply, the time-t aggregate wealth is equal to the market value of the risky asset, $P_t\bar{S}$.

Investor i's wealth share at time $t+1$ is thus

$$\begin{aligned}\alpha_{t+1}^i &\equiv \frac{A_{t+1}^i}{P_{t+1}\bar{S}} \\ &= \frac{A_t^i\left(w_t^i\frac{P_{t+1}}{P_t}+1-w_t^i\right)}{P_{t+1}\bar{S}} \\ &= \frac{\alpha_t^i P_t\bar{S}\left(w_t^i\frac{P_{t+1}}{P_t}+1-w_t^i\right)}{P_{t+1}\bar{S}} \\ &= \alpha_t^i\left(\left(1-w_t^i\right)\exp\left(p_t-p_{t+1}\right)+w_t^i\right)\end{aligned}$$

where the second line uses the budget constraint (17), and the assumption of constant risk-free rate $R_{f,t}=1$.

A1.2 Market clearing

Market clearing for the risk-free asset holds if and only if the aggregate wealth is equal to the market value of the risky asset, i.e.

$$\sum_i A_t^i = P_t\bar{S}$$

Market clearing condition for the risky asset is

$$\sum_i Q_t^i = \bar{S} \iff \sum_i \frac{w_t^i A_t^i}{P_t} = \bar{S} \iff \sum_i w_t^i A_t^i = P_t \bar{S}$$

Hence, the market clearing conditions reduce to

$$\sum_i A_t^i = \sum_i w_t^i A_t^i = P_t \bar{S}$$

This is equivalent to the following set of conditions

$$\sum_i \alpha_t^i w_t^i = 1, \alpha_t^i = \frac{A_t^i}{P_t \bar{S}}, \forall i \tag{A1}$$

The equilibrium price of the risky asset is then determined by the market clearing conditions in (A1) and investors' demand.

A1.3 Investors' preferences and approximate objective function

I derive the approximate objective function, following Campbell et al. (2002).

Investor i has power utility over next period's wealth, with constant relative risk aversion γ^i. At time t, his chooses portfolio weights w_t^i to maximize his expected utility over next period's wealth, i.e. his objective is

$$\max_{w_t^i} \mathbb{E}_t^i \left[\frac{\left(A_{t+1}^i\right)^{1-\gamma^i}}{1-\gamma^i} \right] \tag{A2}$$

His budget constraint is

$$A_{t+1}^i = A_t^i R_{p,t+1}^i$$

where $R_{p,t+1}^i$ is his one-period portfolio return from time t to $t+1$, and

$$R_{p,t+1}^i = w_t^i R_{t+1} + \left(1 - w_t^i\right) R_{f,t} = R_{f,t} + w_t^i \left(R_{t+1} - R_{f,t}\right)$$

Now I derive an approximate linear relationship between the log portfolio return and the log asset returns. Let $r^i_{p,t+1} \equiv \log R^i_{p,t+1}$ denote the log portfolio return, then

$$\begin{aligned} r^i_{p,t+1} - r_{f,t} &= \log\left(\frac{R^i_{p,t+1}}{R_{f,t}}\right) = \log\left(1 + w^i_t\left(\frac{R_{t+1}}{R_{f,t}} - 1\right)\right) \\ &= \log\left(1 + w^i_t\left(\exp\left(r_{t+1} - r_{f,t}\right) - 1\right)\right) \end{aligned}$$

Define the function

$$f^i_t\left(r_{t+1} - r_{f,t}\right) \equiv \log\left(1 + w^i_t\left(\exp\left(r_{t+1} - r_{f,t}\right) - 1\right)\right)$$

And note that

$$\begin{aligned} f^i_t(0) &= 0 \\ \left(f^i_t\right)'(0) &= w^i_t \\ \left(f^i_t\right)''(0) &= w^i_t\left(1 - w^i_t\right) \end{aligned}$$

A second-order Taylor expansion of the function $f^i_t\left(r_{t+1} - r_{f,t}\right)$ around the point $r_{t+1} - r_{f,t} = 0$ yields

$$\begin{aligned} f^i_t\left(r_{t+1} - r_{f,t}\right) &\approx w^i_t\left(r_{t+1} - r_{f,t}\right) + \frac{1}{2}w^i_t\left(1 - w^i_t\right)\left(r_{t+1} - r_{f,t}\right)^2 \\ &\approx w^i_t\left(r_{t+1} - r_{f,t}\right) + \frac{1}{2}w^i_t\left(1 - w^i_t\right)\mathrm{Var}^i_t\left(r_{t+1}\right) \end{aligned}$$

where the second line uses the approximation $\left(r_{t+1} - r_{f,t}\right)^2 \approx \mathrm{Var}^i_t\left(r_{t+1}\right)$, and $\mathrm{Var}^i_t\left(r_{t+1}\right)$ is investor i's perceived variance of the log return r_{t+1}. Hence, the log portfolio return can be approximated by

$$r^i_{p,t+1} \approx w^i_t\left(r_{t+1} - r_{f,t}\right) + \frac{1}{2}w^i_t\left(1 - w^i_t\right)\mathrm{Var}^i_t\left(r_{t+1}\right) + r_{f,t}$$

If the log return of the risky asset r_{t+1} is normally distributed, then log portfolio return

$r^i_{p,t+1}$ is also normally distributed. And we can rewrite the investor's objective as follows

$$
\begin{aligned}
& \max_{w^i_t} \mathbb{E}^i_t \left[\frac{\left(A^i_{t+1}\right)^{1-\gamma^i}}{1-\gamma^i} \right] \\
\Longrightarrow \quad & \max_{w^i_t} \log \left(\mathbb{E}^i_t \left[\left(1-\gamma^i\right) \exp\left(\log\left(A^i_{t+1} \right) \right) \right] \right) \\
\Longrightarrow \quad & \max_{w^i_t} \log \left(\mathbb{E}^i_t \left[\left(1-\gamma^i\right) \exp\left(r_{p,t+1} \right) \right] \right) \\
\Longrightarrow \quad & \max_{w^i_t} \left(1-\gamma^i\right) \left(w^i_t \left(\mathbb{E}^i_t \left[r_{t+1} \right] - r_{f,t} \right) + \frac{1}{2} w^i_t \left(1-w^i_t\right) \mathrm{Var}^i_t \left(r_{t+1}\right) + r_{f,t} \right) \\
& + \frac{1}{2} \left(1-\gamma^i\right)^2 \left(w^i_t\right)^2 \mathrm{Var}^i_t \left(r_{t+1}\right) \\
\Longrightarrow \quad & \max_{w^i_t} w^i_t \left(\mathbb{E}^i_t \left[r_{t+1} \right] - r_{f,t} \right) + \frac{1}{2} w^i_t \left(1-w^i_t\right) \mathrm{Var}^i_t \left(r_{t+1}\right) + \frac{1}{2} \left(1-\gamma^i\right) \left(w^i_t\right)^2 \mathrm{Var}^i_t \left(r_{t+1}\right)
\end{aligned}
$$

Throughout the paper, I assume that for any investor i, his objective is

$$
\max_{w^i_t} w^i_t \left(\mathbb{E}^i_t \left[r_{t+1} \right] - r_{f,t} \right) + \frac{1}{2} w^i_t \left(1-w^i_t\right) \mathrm{Var}^i_t \left(r_{t+1}\right) + \frac{1}{2} \left(1-\gamma^i\right) \left(w^i_t\right)^2 \mathrm{Var}^i_t \left(r_{t+1}\right)
$$

This is a good approximation of the utility maximization problem in (A2), if both the following hold:

- The log return of the risky asset is perceived to be normally distributed by investor i.
- The time interval is short.

A1.4 Optimal portfolio choice

A1.4.1 Retail investors

Retail investor j solves the following problem

$$
\begin{aligned}
U^j_t \left(A^j_t\right) &= \max_{w^j_t} w^j_t \left(\mathbb{E}^j_t \left[r_{t+1} \right] - r_{f,t} \right) + \frac{1}{2} w^j_t \left(1-w^j_t\right) \mathrm{Var}^j_t \left(r_{t+1}\right) \\
& \quad + \frac{1}{2} \left(1-\gamma^R\right) \left(w^j_t\right)^2 \mathrm{Var}^i_t \left(r_{t+1}\right)
\end{aligned}
$$

The F.O.C. is

$$
\begin{aligned}
& \mathbb{E}^j_t \left[r_{t+1} \right] - r_{f,t} + \frac{1}{2} \mathrm{Var}^j_t \left(r_{t+1}\right) - \gamma^R w^j_t \mathrm{Var}^j_t \left(r_{t+1}\right) = 0 \\
\Longrightarrow \quad & w^j_t = \frac{\mathbb{E}^j_t \left[r_{t+1} \right] - r_{f,t} + \frac{1}{2} \mathrm{Var}^j_t \left(r_{t+1}\right)}{\gamma^R \mathrm{Var}^j_t \left(r_{t+1}\right)} = \tau^R \frac{\mathbb{E}^j_t \left[r_{t+1} \right] - r_{f,t} + \frac{1}{2} \mathrm{Var}^j_t \left(r_{t+1}\right)}{\mathrm{Var}^j_t \left(r_{t+1}\right)}
\end{aligned}
$$

Substitute retail investor j's beliefs into the above expression, we get his time-0 and time-1 demand for the risky asset

$$w_0^j = \tau^R \left(\frac{\mathbb{E}_0[p_1] + y_0^j - p_0}{\sigma_0^2} + \frac{1}{2} \right) \tag{A3}$$

$$w_1^j = \tau^R \left(\frac{\mu_d + y_1^j - p_1}{\sigma_d^2} + \frac{1}{2} \right) \tag{A4}$$

A1.4.2 Long institution

The long institution IL solves the following problem

$$\begin{aligned} U_t^{IL}\left(A_t^{IL}\right) &= \max_{w_t^{IL}} w_t^{IL}\left(\mathbb{E}_t^{IL}[r_{t+1}] - r_{f,t}\right) + \frac{1}{2} w_t^{IL}\left(1 - w_t^{IL}\right) \mathrm{Var}_t^{IL}(r_{t+1}) \\ &\quad + \frac{1}{2}\left(1-\gamma^I\right)\left(w_t^{IL}\right)^2 \mathrm{Var}_t^{IL}(r_{t+1}) \\ \text{s.t.} &\quad w_t^{IL} \geq 0 \end{aligned}$$

The solution is

$$w_t^{IL} = \max\left\{0, \tau^I \frac{\mathbb{E}_t^{IL}[r_{t+1}] - r_{f,t} + \frac{1}{2}\mathrm{Var}_t^{IL}(r_{t+1})}{\mathrm{Var}_t^{IL}(r_{t+1})}\right\}$$

Substitute IL's beliefs into the above expression, we get his time-0 and time-1 demand for the risky asset

$$\begin{aligned} w_0^{IL} &= \max\left\{0, \tau^I \left(\frac{\mathbb{E}_0[p_1] + \delta_0^{IL} - p_0}{\sigma_0^2} + \frac{1}{2} \right)\right\} \\ w_1^{IL} &= \max\left\{0, \tau^I \left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2} \right)\right\} \end{aligned}$$

A1.4.3 Short institution

The short institution IS solves the following problem

$$\begin{aligned} U_t^{IS}\left(A_t^{IS}\right) &= \max_{w_t^{IS}} w_t^{IS}\left(\mathbb{E}_t^{IS}[r_{t+1}] - r_{f,t}\right) + \frac{1}{2} w_t^{IS}\left(1 - w_t^{IS}\right) \mathrm{Var}_t^{IS}(r_{t+1}) \\ &\quad + \frac{1}{2}\left(1-\gamma^I\right)\left(w_t^{IS}\right)^2 \mathrm{Var}_t^{IS}(r_{t+1}) \\ \text{s.t.} &\quad w_t^{IS} \geq -\frac{1}{m} \end{aligned}$$

The solution is

$$w_t^{IS} = \max\left\{-\frac{1}{m}, \tau^I \frac{\mathbb{E}_t^{IS}[r_{t+1}] - r_{f,t} + \frac{1}{2}\mathrm{Var}_t^{IS}(r_{t+1})}{\mathrm{Var}_t^{IS}(r_{t+1})}\right\}$$

Substitute IS's beliefs into the above expression, we get his time-0 and time-1 demand for the risky asset

$$w_0^{IS} = \max\left\{-\frac{1}{m}, \tau^I\left(\frac{\mathbb{E}_0[p_1] + \delta_0^{IS} - p_0}{\sigma_0^2} + \frac{1}{2}\right)\right\}$$
$$w_1^{IS} = \max\left\{-\frac{1}{m}, \tau^I\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right)\right\}$$

A1.5 Proof of Lemma 1

Proof. I first restate the timeline and the wealth share dynamics of individual retail investors. At time $t-1$ after trading, retail investor j has dollar wealth A_t^j and wealth share α_t^j. At time t before trading, retail investors first split their aggregate wealth $\sum_{j=1}^N A_t^j$ equally. In particular, they split their aggregate stock positions and aggregate bond positions equally. After that, retail investor j has wealth $\hat{A}_t^j = \frac{1}{N}\sum_{j=1}^N A_t^j$ and wealth share

$$\hat{\alpha}_t^j \equiv \frac{\hat{A}_t^j}{A_t} = \frac{\frac{1}{N}\sum_{j=1}^N A_t^j}{A_t} = \frac{1}{N}\sum_{j=1}^N \alpha_t^j \tag{A5}$$

Then trade opens at time t, and retail investor j allocates his wealth $\hat{A}_t^j$ into the risky asset and the risk-free asset. His demand for the risky asset (in terms of the number of shares) is $Q_t^j = \frac{w_t^j \hat{A}_t^j}{P_t}$, where w_t^j is his optimal portfolio weight in equations (A3) and (A4). After trading, his wealth share becomes

$$\alpha_{t+1}^j = \hat{\alpha}_t^j\left(\left(1 - w_t^j\right)\exp\left(p_t - p_{t+1}\right) + w_t^j\right) \tag{A6}$$

Next, I show that the equilibrium price of the risky asset is the same as that in an economy with three investors – a representative retail investor, the long institution, and the short institution. This proves the existence of a representative retail investor. At time t, market clearing for the risky asset is

$$
\begin{aligned}
& \sum_{j=1}^{N} Q_t^j + Q_t^{IL} + Q_t^{IS} = \bar{S} \\
\Longrightarrow \quad & \sum_{j=1}^{N} \frac{w_t^j \hat{A}_t^j}{P_t} + \frac{w_t^{IL} A_t^{IL}}{P_t} + \frac{w_t^{IS} A_t^{IS}}{P_t} = \bar{S} \\
\Longrightarrow \quad & \sum_{j=1}^{N} w_t^j \left(\frac{1}{N} \sum_{k=1}^{N} \alpha_t^k \right) + w_t^{IL} \alpha_t^{IL} + w_t^{IS} \alpha_t^{IS} = 1 \\
\Longrightarrow \quad & \left(\sum_{k=1}^{N} \alpha_t^k \right) \frac{1}{N} \sum_{j=1}^{N} \tau^R \left(\frac{\mathbb{E}_t \left[p_{t+1} \right] + y_t^j - p_t}{\sigma_t^2} + \frac{1}{2} \right) + w_t^{IL} \alpha_t^{IL} + w_t^{IS} \alpha_t^{IS} = 1 \\
\Longrightarrow \quad & \left(\sum_{k=1}^{N} \alpha_t^k \right) \tau^R \left(\frac{\mathbb{E}_t \left[p_{t+1} \right] + \frac{1}{N} \sum_{j=1}^{N} y_t^j - p_t}{\sigma_t^2} + \frac{1}{2} \right) + w_t^{IL} \alpha_t^{IL} + w_t^{IS} \alpha_t^{IS} = 1
\end{aligned}
$$

where the third line uses the definition of $\hat{\alpha}_t^j$ in equation (A5), and the fourth line uses the optimal portfolio weight of retail investor j in equations (A3) and (A4).

Define

$$
A_t^R \equiv \sum_{j=1}^{N} A_t^j, \alpha_t^R \equiv \sum_{j=1}^{N} \alpha_t^j \tag{A7}
$$

$$
\delta_t^R \equiv \frac{1}{N} \sum_{j=1}^{N} y_t^j \tag{A8}
$$

$$
w_t^R \equiv \tau^R \left(\frac{\mathbb{E}_t \left[p_{t+1} \right] + \delta_t^R - p_t}{\sigma_t^2} + \frac{1}{2} \right) = \frac{1}{N} \sum_{j=1}^{N} w_t^j \tag{A9}
$$

and substitute into the market clearing condition to get

$$
w_t^R \alpha_t^R + w_t^{IL} \alpha_t^{IL} + w_t^{IS} \alpha_t^{IS} = 1
$$

with $\alpha_t^R + \alpha_t^{IL} + \alpha_t^{IS} = \sum_{j=1}^{N} \alpha_t^j + \alpha_t^{IL} + \alpha_t^{IS} = 1$.

Hence, the equilibrium price of the risky asset is the same as that in an economy with three investors – a representative retail investor R, the long institution IL, and the short institution IS, where the three investors have demand $\left(w_t^R, w_t^{IL}, w_t^{IS} \right)$, and wealth shares $\left(\alpha_t^R, \alpha_t^{IL}, \alpha_t^{IS} \right)$. In other words, there exists a representative retail investor whose demand for the risky asset is given by

$$
\begin{aligned}
w_0^R &= \tau^R \left(\frac{\mathbb{E}_0\left[p_1\right] + \delta_0^R - p_0}{\sigma_0^2} + \frac{1}{2} \right) \\
w_1^R &= \tau^R \left(\frac{\mu_d + \delta_1^R - p_1}{\sigma_d^2} + \frac{1}{2} \right)
\end{aligned}
$$

The representative retail investor has constant relative risk tolerance τ^R and beliefs

$$
\begin{aligned}
&\mathbb{E}_0^R\left[p_1\right] = \mathbb{E}_0\left[p_1\right] + \delta_0^R, \mathrm{Var}_0^R\left(p_1\right) = \sigma_0^2 \\
&\mathbb{E}_1^R\left[d\right] = \mu_d + \delta_1^R, \mathrm{Var}_1^R\left(d\right) = \sigma_d^2
\end{aligned}
$$

where $\delta_0^R \equiv \frac{1}{N}\sum_{j=1}^N y_0^j$ and $\delta_1^R \equiv \frac{1}{N}\sum_{j=1}^N y_1^j$.

Finally, I derive the wealth share dynamics of the representative retail investor. Start from the definition of α_{t+1}^R

$$
\begin{aligned}
\alpha_{t+1}^R &\equiv \sum_{j=1}^N \alpha_{t+1}^j \\
&= \sum_{j=1}^N \hat{\alpha}_t^j \left(\left(1 - w_t^j\right) \exp\left(p_t - p_{t+1}\right) + w_t^j\right) \\
&= \left(\frac{1}{N}\sum_{k=1}^N \alpha_t^k\right) \sum_{j=1}^N \left(\left(1 - w_t^j\right) \exp\left(p_t - p_{t+1}\right) + w_t^j\right) \\
&= \alpha_t^R \left(\left(1 - \frac{1}{N}\sum_{j=1}^N w_t^j\right) \exp\left(p_t - p_{t+1}\right) + \frac{1}{N}\sum_{j=1}^N w_t^j\right) \\
&= \alpha_t^R \left(\left(1 - w_t^R\right) \exp\left(p_t - p_{t+1}\right) + w_t^R\right)
\end{aligned}
$$

where the second equality uses investor j's wealth share dynamics in equation (A6), and the last equality uses the representative retail investor's demand in equation (A9). □

A1.6 Proof of Proposition 1

Proof. I focus on monotone equilibrium of Definition 1, with sentiment cutoffs δ_1^m, δ_1^h satisfying $\underline{\delta}_1 < \delta_1^m < \delta_1^h < \bar{\delta}_1$. Hence, $\forall \delta_1^R \in [\underline{\delta}_1, \delta_1^m)$, the equilibrium price $p_1\left(\delta_1^R\right) < p_1^m$. Similarly, $\forall \delta_1^R \in \left[\delta_1^m, \delta_1^h\right)$, the price $p_1\left(\delta_1^R\right) \in \left[p_1^m, p_1^h\right)$. And $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$, the price $p_1\left(\delta_1^R\right) \geq p_1^h$.

Next, I solve the equilibrium price from the market clearing condition in equation (35).

- For $\delta_1^R \in [\underline{\delta}_1, \delta_1^m)$, I look for an equilibrium price $p_1 < p_1^m$. Substitute the optimal portfolio choices of the three investors, (34), (28), and (30) into the market clearing condition (35), I get

$$
\begin{aligned}
& \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \sum_i \alpha_1^i(p_1)\tau^i\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\
\Longrightarrow \quad & p_1 = \mu_d + \left(\frac{\frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R - 1}{\sum_i \alpha_1^i(p_1)\tau^i} + \frac{1}{2}\right)\sigma_d^2 \\
\Longrightarrow \quad & p_1 = \mu_d + \left(\frac{1}{2} + \frac{\frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R - 1}{\tau_1(p_1)}\right)\sigma_d^2
\end{aligned}
$$

where

$$
\tau_1(p_1) \equiv \sum_i \alpha_1^i(p_1)\tau^i = \alpha_1^R(p_1)\tau^R + \left(1 - \alpha_1^R(p_1)\right)\tau^I
$$

Define the function

$$
J\left(p_1, \delta_1^R\right) \equiv \mu_d + \left(\frac{1}{2}\sigma_d^2 + \frac{\alpha_1^R(p_1)\tau^R\delta_1^R - \sigma_d^2}{\tau_1(p_1)}\right) - p_1
$$

Then the equilibrium price p_1 solves $J\left(p_1, \delta_1^R\right) = 0$.

The cutoff sentiment shock δ_1^m solves $J\left(p_1^m, \delta_1^m\right) = 0$, which yields

$$
\delta_1^m = \frac{\left(p_1^m - \mu_d - \frac{1}{2}\sigma_d^2\right)\tau_1(p_1^m) + \sigma_d^2}{\alpha_1^R(p_1^m)\tau^R} = \frac{\sigma_d^2}{\alpha_1^R(p_1^m)\tau^R}
$$

- For $\delta_1^R \in \left[\delta_1^m, \delta_1^h\right)$, I look for an equilibrium price $p_1 \in \left[p_1^m, p_1^h\right)$. Substitute the optimal portfolio choices of the three investors, (34), (28), and (30) into the market clearing condition (35), I get

$$
\begin{aligned}
& \alpha_1^R(p_1)\tau^R\left(\frac{\mu_d + \delta_1^R - p_1}{\sigma_d^2} + \frac{1}{2}\right) + \alpha_1^{IS}(p_1)\tau^I\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\
\Longrightarrow \quad & \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \left(\alpha_1^R(p_1)\tau^R + \alpha_1^{IS}(p_1)\tau^I\right)\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\
\Longrightarrow \quad & p_1 = \mu_d + \left(\frac{1}{2} + \frac{\frac{1}{\sigma_d^2}\alpha_1^R(p_1)\tau^R\delta_1^R - 1}{\alpha_1^R(p_1)\tau^R + \alpha_1^{IS}(p_1)\tau^I}\right)\sigma_d^2 \\
\Longrightarrow \quad & p_1 = \mu_d + \left(\frac{1}{2} + \frac{\frac{1}{\sigma_d^2}\alpha_1^R(p_1)\tau^R\delta_1^R - 1}{\hat{\tau}_1(p_1)}\right)\sigma_d^2
\end{aligned}
$$

where

$$\hat{\tau}_1(p_1) \equiv \alpha_1^R(p_1)\tau^R + \alpha_1^{IS}(p_1)\tau^I$$

Define the function

$$H\left(p_1, \delta_1^R\right) \equiv \mu_d + \left(\frac{1}{2}\sigma_d^2 + \frac{\alpha_1^R(p_1)\tau^R\delta_1^R - \sigma_d^2}{\hat{\tau}_1(p_1)}\right) - p_1$$

Then the equilibrium price p_1 solves $H\left(p_1, \delta_1^R\right) = 0$.

The cutoff sentiment shock δ_1^h solves $H\left(p_1^h, \delta_1^h\right) = 0$, which yields

$$\delta_1^h = \frac{\left(p_1^h - \mu_d - \frac{1}{2}\sigma_d^2\right)\hat{\tau}_1\left(p_1^h\right) + \sigma_d^2}{\alpha_1^R\left(p_1^h\right)\tau^R} = \frac{\frac{1}{m\tau^I}\hat{\tau}_1\left(p_1^h\right) + 1}{\alpha_1^R\left(p_1^h\right)\tau^R}\sigma_d^2$$

- For $\delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$, I look for an equilibrium price $p_1 \geq p_1^h$. Substitute the optimal portfolio choices of the three investors, (34), (28), and (30) into the market clearing condition (35), I get

$$\begin{aligned} & \alpha_1^R(p_1)\tau^R\left(\frac{\mu_d + \delta_1^R - p_1}{\sigma_d^2} + \frac{1}{2}\right) - \alpha_1^{IS}(p_1)\frac{1}{m} = 1 \\ \Longrightarrow \quad & p_1 = \mu_d + \delta_1^R + \left(\frac{1}{2} - \frac{1 + \alpha_1^{IS}(p_1)\frac{1}{m}}{\alpha_1^R(p_1)\tau^R}\right)\sigma_d^2 \end{aligned}$$

Define the function

$$G\left(p_1, \delta_1^R\right) = \mu_d + \delta_1^R + \left(\frac{1}{2} - \frac{1 + \alpha_1^{IS}(p_1)\frac{1}{m}}{\alpha_1^R(p_1)\tau^R}\right)\sigma_d^2 - p_1$$

Then the equilibrium price p_1 solves $G\left(p_1, \delta_1^R\right) = 0$.

□

A1.7 Lemma A1 and proof

Lemma A1 (Properties of the implicit function $G\left(p_1, \delta_1^R\right)$). *Consider a monotone equilibrium of Definition 1, where the time-0 portfolios satisfy $w_0^R > 1$, $w_0^{IS} < 0$, $w_0^R > w_0^{IL} > w_0^{IS}$, and investors always have strictly positive wealth $\forall \delta_1 \in \left(\underline{\delta}_1, \bar{\delta}_1\right)$. Let p_1^R denote* *the price at which the retail investor's time-*1 *wealth is zero,*

$$p_1^R \equiv p_0 + \log\left(1 - \frac{1}{w_0^R}\right)$$

Then the implicit function $G\left(p_1, \delta_1^R\right)$ *has the following properties on* $p_1 \in \left(p_1^R, +\infty\right)$*:*

1. $G\left(p_1, \delta_1^R\right)$ *is continuous and strictly increasing in* δ_1^R*:* $\frac{\partial G\left(p_1, \delta_1^R\right)}{\partial \delta_1} = 1 > 0$.

2. $G\left(p_1, \delta_1^R\right)$ *is continuous and strictly concave in* p_1*:* $\frac{\partial^2 G\left(p_1, \delta_1^R\right)}{\partial p_1^2} < 0$.

3. $\frac{\partial G\left(p_1, \delta_1^R\right)}{\partial p_1}$ *does not depend on* δ_1^R*:* $\frac{\partial^2 G\left(p_1, \delta_1^R\right)}{\partial p_1 \partial \delta_1^R} = 0$.

4. $G\left(p_1, \delta_1^R\right)$*, as a function of* p_1*, has at most two distinct roots on* $p_1 \in \left(p_1^R, +\infty\right)$.

Proof. First, I derive p_1^R from

$$\begin{aligned}
& \alpha_1^R\left(p_1^R\right) = 0 \\
\implies & 0 = \alpha_0^R\left(\left(1 - w_0^R\right)\exp\left(p_0 - p_1^R\right) + w_0^R\right) \\
\implies & p_1^R = p_0 + \log\left(1 - \frac{1}{w_0^R}\right)
\end{aligned}$$

Then $\forall p_1 > p_1^R$, $\alpha_1\left(p_1\right) > 0$. And thus $G\left(p_1, \delta_1^R\right)$ is continuous and twice differentiable, $\forall p_1 > p_1^R$, $\forall \delta_1^R$.

To show Properties 1-3, compute the following derivatives

$$
\begin{aligned}
\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial \delta_1^R} &= 1\\
\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1} &= -\left(\alpha_1^R\left(p_1\right)\tau^R\right)^{-2}\\
&\quad\cdot\left(\frac{d\alpha_1^{IS}\left(p_1\right)}{dp_1}\frac{1}{m}\alpha_1^R\left(p_1\right)\tau^R-\frac{d\alpha_1^R\left(p_1\right)}{dp_1}\tau^R\left(1+\alpha_1^{IS}\left(p_1\right)\frac{1}{m}\right)\right)\sigma_d^2-1\\
&= \left(\alpha_1^R\left(p_1\right)\tau^R\right)^{-2}\exp\left(p_0-p_1\right)\\
&\quad\cdot\tau^R\left(\alpha_0^{IS}\left(1-w_0^{IS}\right)\frac{1}{m}\alpha_1^R\left(p_1\right)-\alpha_0^R\left(1-w_0^R\right)\left(1+\alpha_1^{IS}\left(p_1\right)\frac{1}{m}\right)\right)\sigma_d^2\\
&\quad -1\\
&= \left(\alpha_1^R\left(p_1\right)\tau^R\right)^{-2}\exp\left(p_0-p_1\right)\\
&\quad\cdot\alpha_0^R\tau^R\left(w_0^R-1+\frac{1}{m}\alpha_0^{IS}\left(w_0^R-w_0^{IS}\right)\right)\sigma_d^2-1\\
\frac{\partial^2 G\left(p_1,\delta_1^R\right)}{\partial p_1\partial \delta_1^R} &= 0\\
\frac{\partial^2 G\left(p_1,\delta_1^R\right)}{\partial p_1^2} &= -\left(\alpha_1^R\left(p_1\right)\tau^R\right)^{-2}\sigma_d^2\\
&\quad\cdot\left(\frac{d\alpha_1^R\left(p_1\right)}{dp_1}\tau^R\left(1+\alpha_1^{IS}\left(p_1\right)\frac{1}{m}\right)-\frac{d\alpha_1^{IS}\left(p_1\right)}{dp_1}\frac{1}{m}\alpha_1^R\left(p_1\right)\tau^R\right)\\
&\quad\cdot\left(1+\frac{2}{\alpha_1^R\left(p_1\right)}\frac{d\alpha_1^R\left(p_1\right)}{dp_1}\right)
\end{aligned}
$$

From the wealth share dynamics, we get

$$
\begin{aligned}
&\alpha_{t+1}^i\left(p_{t+1}\right)=\alpha_t^i\left(\left(1-w_t^i\right)\left(p_t-p_{t+1}\right)+w_t^i\right)\\
\implies\quad &\frac{d\alpha_{t+1}^i\left(p_{t+1}\right)}{dp_{t+1}}=-\alpha_t^i\left(1-w_t^i\right)\exp\left(p_t-p_{t+1}\right)
\end{aligned}
$$

Since $w_0^R>1$ and $w_0^{IS}<0$, we have

$$
\frac{d\alpha_1^R\left(p_1\right)}{dp_1}>0,\frac{d\alpha_1^{IS}\left(p_1\right)}{dp_1}<0
$$

Hence, $\frac{\partial^2 G\left(p_1,\delta_1^R\right)}{\partial p_1^2}<0$, i.e. $G\left(p_1,\delta_1^R\right)$ is strictly concave in p_1, $\forall p_1\in\left(\delta_1^R,+\infty\right)$.

Next, I show property 4. For a given δ_1^R, suppose $G\left(p_1,\delta_1^R\right)$ has more than two roots. Let x_1,x_2,x_3 denote three of the roots, with $x_1<x_2<x_3$. Then $\exists\lambda\in(0,1)$, such that

$x_2 = \lambda x_1 + (1-\lambda) x_3$. Since $G\left(p_1, \delta_1^R\right)$ is continuous and strictly concave in p_1,

$$0 = \lambda G\left(x_1, \delta_1^R\right) + (1-\lambda) G\left(x_3, \delta_1^R\right) = G\left(\lambda x_1 + (1-\lambda) x_3, \delta_1^R\right) < G\left(x_2, \delta_1^R\right) = 0$$

A contradiction. Hence, $\forall p_1 \in \left(p_1^R, +\infty\right)$, $G\left(p_1, \delta_1^R\right)$ (as a function of p_1) has at most two distinct roots. $\square$

A1.8 Proof of Proposition 2

Proof. I first show that $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, $G\left(p_1, \delta_1^R\right) = 0$ has exactly one root that satisfies $p_1 > p_1^h$. Suppose otherwise, then from Lemma A1, there are two roots x_1 and x_2 which satisfy $p_1^h < x_1 < x_2$, and $G\left(x_1, \delta_1^R\right) = G\left(x_2, \delta_1^R\right) = 0$. Since $G\left(p_1^h, \delta_1^h\right) = 0$ and $\frac{\partial G\left(p_1, \delta_1^R\right)}{\partial \delta_1^R} = 1 > 0$, then $G\left(p_1^h, \delta_1^R\right) > G\left(p_1^h, \delta_1^h\right) = 0$, $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$. $p_1^h < x_1 < x_2 \rightarrow \exists \lambda \in (0,1)$ such that $x_1 = \lambda p_1^h + (1-\lambda) x_2$. And since $G\left(p_1, \delta_1^R\right)$ is strictly concave in p_1, we have

$$0 < \lambda G\left(p_1^h, \delta_1^R\right) + (1-\lambda) G\left(x_2, \delta_1^R\right) < G\left(\lambda p_1^h + (1-\lambda) x_2, \delta_1^R\right) = G\left(x_1, \delta_1^R\right) = 0$$

A contradiction. Hence, $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, $G\left(p_1, \delta_1^R\right)$ has exactly one root that satisfies $p_1 > p_1^h$. In a monotone equilibrium of Definition 1, this is the unique equilibrium price in the high sentiment region $\delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$.

Next, I derive conditions for discontinuity in price. Consider the following two cases:

- Case 1: $\left.\frac{\partial G\left(p_1, \delta_1^h\right)}{\partial p_1}\right|_{p_1 = p_1^h} \leq 0$.

  From the strict concavity of $G\left(p_1, \delta_1^R\right)$ in Lemma A1, $\forall p_1 > p_1^h$, $\frac{\partial G\left(p_1, \delta_1^h\right)}{\partial p_1} < \left.\frac{\partial G\left(p_1, \delta_1^h\right)}{\partial p_1}\right|_{p_1 = p_1^h} \leq 0$. This implies that $G\left(p_1, \delta_1^h\right) < G\left(p_1^h, \delta_1^h\right) = 0, \forall p_1 > p_1^h$. Hence, p_1^h is the largest root of $G\left(p_1, \delta_1^h\right) = 0$.

  From Lemma A1, $\frac{\partial G\left(p_1, \delta_1^R\right)}{\partial p_1 \partial \delta_1^R} = 0$ and $\frac{\partial^2 G\left(p_1, \delta_1^R\right)}{\partial p_1^2} < 0$. Then

$$\begin{aligned} & \left.\frac{\partial G\left(p_1, \delta_1^h\right)}{\partial p_1}\right|_{p_1 = p_1^h} \leq 0 \\ \Longrightarrow \quad & \left.\frac{\partial G\left(p_1, \delta_1^R\right)}{\partial p_1}\right|_{p_1 = p_1^h} \leq 0, \forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right] \\ \Longrightarrow \quad & \frac{\partial G\left(p_1, \delta_1^R\right)}{\partial p_1} < 0, \forall p_1 > p_1^h, \forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right] \end{aligned}$$

Moreover, if $\left.\frac{\partial G\left(p_1,\delta_1^h\right)}{\partial p_1}\right|_{p_1=p_1^h} = 0$, then $\left.\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\right|_{p_1=p_1^h} = 0$, $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$. Otherwise, $\left.\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\right|_{p_1=p_1^h} < 0$, $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$.

Using the implicit function theorem, $\forall p_1 > p_1^h$, $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$,

$$
\begin{aligned}
& \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} + \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial \delta_1^R} = 0 \\
\implies\quad & \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} + 1 = 0 \\
\implies\quad & \frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} = -\frac{1}{\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}} > 0
\end{aligned}
$$

Hence, $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$, the equilibrium price $p_1\left(\delta_1^R\right)$ is strictly increasing in δ_1^R. Furthermore, $p_1\left(\delta_1^R\right)$ is continuous in δ_1^R on $\delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, and is right-continuous at $\delta_1^R = \delta_1^h$.

- Case 2: $\left.\frac{\partial G\left(p_1,\delta_1^h\right)}{\partial p_1}\right|_{p_1=p_1^h} > 0$.

  First, I prove that $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$, $G\left(p_1,\delta_1^R\right) = 0$ has two distinct roots, denoted as $x_1\left(\delta_1^R\right)$ and $x_2\left(\delta_1^R\right)$, with $x_1\left(\delta_1^R\right) \le p_1^h < x_2\left(\delta_1^R\right)$. And $x_1\left(\delta_1^R\right) = p_1^h$ if and only if $\delta_1^R = \delta_1^h$.

  - $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, we have $G\left(p_1^h,\delta_1^R\right) > G\left(p_1^h,\delta_1^h\right) = 0$, and $G\left(+\infty,\delta_1^R\right) = -\infty$. Let p_1^R denote the price at which the retail investor's time-1 wealth share is exactly zero, then p_1^R satisfies

    $$
    \begin{aligned}
    & \alpha_1^R\left(p_1^R\right) = 0 \\
    \implies\quad & 0 = \alpha_0^R\left(\left(1-w_0^R\right)\exp\left(p_0 - p_1^R\right) + w_0^R\right) \\
    \implies\quad & p_1^R = p_0 + \log\left(1 - \frac{1}{w_0^R}\right)
    \end{aligned}
    $$

    And we have $G\left(p_1^R,\delta_1^R\right) = -\infty$. Then $G\left(p_1^R,\delta_1^R\right) = G\left(+\infty,\delta_1^R\right) = -\infty < 0 < G\left(p_1^h,\delta_1^R\right)$. By the intermediate value theorem, $G\left(p_1,\delta_1^R\right) = 0$ has two distinct roots $x_1\left(\delta_1^R\right), x_2\left(\delta_1^R\right)$ such that $p_1^R < x_1\left(\delta_1^R\right) < p_1^h < x_2\left(\delta_1^R\right)$, $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$. In a monotone equilibrium of Definition 1, $x_2\left(\delta_1^R\right)$ is the unique equilibrium price.

    Next, I show that $\forall \delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, $\left.\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\right|_{p_1=x_2\left(\delta_1^R\right)} < 0$. Suppose otherwise, then $\left.\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\right|_{p_1=x_2\left(\delta_1^R\right)} \geq 0 \implies \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1} > 0, \forall p_1 < x_2\left(\delta_1^R\right)$. This implies $0 = G\left(p_1^h,\delta_1^h\right) < G\left(p_1^h,\delta_1^R\right) < G\left(x_2\left(\delta_1^R\right),\delta_1^R\right) = 0$, a contradiction.

– At the cutoff $\delta_1^R = \delta_1^h$, $\left.\frac{\partial G\left(p_1,\delta_1^h\right)}{\partial p_1}\right|_{p_1=p_1^h} > 0$ implies that, $\exists \varepsilon > 0$ and small, $G\left(p_1^h+\varepsilon,\delta_1^h\right) > G\left(p_1^h,\delta_1^h\right) = 0$. Together with $G\left(+\infty,\delta_1^h\right) = -\infty < 0$, this implies that $G\left(p_1,\delta_1^h\right)$ has two distinct roots $x_1\left(\delta_1^h\right), x_2\left(\delta_1^h\right)$ such that $x_1\left(\delta_1^h\right) = p_1^h < x_2\left(\delta_1^h\right)$.

  Next, I show that $\left.\frac{\partial G\left(p_1,\delta_1^h\right)}{\partial p_1}\right|_{p_1=x_2\left(\delta_1^h\right)} < 0$. Suppose otherwise, then $\left.\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\right|_{p_1=x_2\left(\delta_1^h\right)} \geq 0 \implies \frac{\partial G\left(p_1,\delta_1^h\right)}{\partial p_1} > 0, \forall p_1 < x_2\left(\delta_1^h\right)$. This implies $0 = G\left(p_1^h,\delta_1^h\right) < G\left(x_2\left(\delta_1^h\right),\delta_1^h\right) = 0$, a contradiction.

In a monotone equilibrium of Definition 1, $\forall \delta_1^R \in \left(\delta_1^h,\bar{\delta}_1\right]$, the equilibrium price has to be greater than p_1^h. Hence, $x_2\left(\delta_1^R\right)$ is the unique equilibrium price on $\delta_1^R \in \left(\delta_1^h,\bar{\delta}_1\right]$. And since $p_1^h < x_2\left(\delta_1^h\right)$, the pricing function $p_1\left(\delta_1^R\right)$ is discontinuous at $\delta_1^R = \delta_1^h$.

Using the implicit function theorem, $\forall p_1 > x_2\left(\delta_1^h\right)$, $\forall \delta_1^R \in \left[\delta_1^h,\bar{\delta}_1\right]$,

$$
\begin{aligned}
& \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} + \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial \delta_1^R} = 0 \\
\implies & \frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}\frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} + 1 = 0 \\
\implies & \frac{dp_1\left(\delta_1^R\right)}{d\delta_1^R} = -\frac{1}{\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial p_1}} > 0
\end{aligned}
$$

Hence, $\forall \delta_1^R \in \left[\delta_1^h,\bar{\delta}_1\right]$, the equilibrium price $p_1\left(\delta_1^R\right)$ is strictly increasing in δ_1^R. Furthermore, $p_1\left(\delta_1^R\right)$ is continuous in δ_1^R on $\delta_1^R \in \left(\delta_1^h,\bar{\delta}_1\right]$, and is discontinuous at $\delta_1^R = \delta_1^h$.

□

A1.9 Proof of Proposition 3

Proof.

- Low sentiment $\delta_1^R \in [\underline{\delta}_1, \delta_1^m)$: from the optimal portfolio choices of the three investors, (34), (28), (30), and the market clearing condition (35), we get

$$\begin{aligned} & \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \sum_i \alpha_1^i(p_1)\tau^i\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\ \Longrightarrow\quad & \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \tau_1(p_1)\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\ \Longrightarrow\quad & \alpha_1^R(p_1)\tau^R\delta_1^R + \tau_1(p_1)\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) = \sigma_d^2 \end{aligned}$$

  Using the implicit function theorem,

$$\begin{aligned} & \alpha_1^R(p_1)\tau^R + \frac{d\left(\alpha_1^R(p_1)\tau^R\delta_1^R\right)}{dp_1}\frac{dp_1}{d\delta_1^R} + \frac{d\tau_1(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) - \tau_1(p_1)\frac{dp_1}{d\delta_1^R} = 0 \\ \Longrightarrow\quad & \frac{dp_1}{d\delta_1^R} = \frac{\frac{\alpha_1^R(p_1)\tau^R}{\tau_1(p_1)}}{1 - \frac{1}{\tau_1(p_1)}\left(\frac{d\alpha_1^R(p_1)}{dp_1}\tau^R\delta_1^R + \frac{d\tau_1(p_1)}{dp_1}\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right)\right)} \end{aligned}$$

- Medium sentiment $\delta_1^R \in \left(\delta_1^m, \delta_1^h\right)$: from the optimal portfolio choices of the three investors, (34), (28), (30), and the market clearing condition (35), we get

$$\begin{aligned} & \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \left(\alpha_1^R(p_1)\tau^R + \alpha_1^{IS}(p_1)\tau^I\right)\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\ \Longrightarrow\quad & \frac{\alpha_1^R(p_1)\tau^R}{\sigma_d^2}\delta_1^R + \hat{\tau}_1(p_1)\left(\frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2}\right) = 1 \\ \Longrightarrow\quad & \alpha_1^R(p_1)\tau^R\delta_1^R + \hat{\tau}_1(p_1)\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) = \sigma_d^2 \end{aligned}$$

  Using the implicit function theorem,

$$\begin{aligned} & \alpha_1^R(p_1)\tau^R + \frac{d\left(\alpha_1^R(p_1)\tau^R\delta_1^R\right)}{dp_1}\frac{dp_1}{d\delta_1^R} + \frac{d\hat{\tau}_1(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) - \hat{\tau}_1(p_1)\frac{dp_1}{d\delta_1^R} = 0 \\ \Longrightarrow\quad & \frac{dp_1}{d\delta_1^R} = \frac{\frac{\alpha_1^R(p_1)\tau^R}{\hat{\tau}_1(p_1)}}{1 - \frac{1}{\hat{\tau}_1(p_1)}\left(\frac{d\left(\alpha_1^R(p_1)\right)}{dp_1}\tau^R\delta_1^R + \frac{d\hat{\tau}_1(p_1)}{dp_1}\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right)\right)} \end{aligned}$$

- High sentiment $\delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$: from the optimal portfolio choices of the three investors, (34), (28), (30), and the market clearing condition (35), we get

$$\begin{aligned}
& \alpha_1^R(p_1)\,\tau^R\left(\frac{\mu_d+\delta_1^R-p_1}{\sigma_d^2}+\frac{1}{2}\right)-\alpha_1^{IS}(p_1)\,\frac{1}{m}=1 \\
\Longrightarrow\quad & \frac{\alpha_1^R(p_1)\,\tau^R}{\sigma_d^2}\delta_1^R+\alpha_1^R(p_1)\,\tau^R\left(\frac{\mu_d-p_1}{\sigma_d^2}+\frac{1}{2}\right)-\alpha_1^{IS}(p_1)\,\frac{1}{m}=1 \\
\Longrightarrow\quad & \alpha_1^R(p_1)\,\tau^R\delta_1^R+\alpha_1^R(p_1)\,\tau^R\left(\mu_d+\frac{1}{2}\sigma_d^2-p_1\right)-\alpha_1^{IS}(p_1)\,\frac{1}{m}\sigma_d^2=\sigma_d^2
\end{aligned}$$

Using the implicit function theorem,

$$\begin{aligned}
& \alpha_1^R(p_1)\,\tau^R+\frac{d\left(\alpha_1^R(p_1)\,\tau^R\delta_1^R\right)}{dp_1}\frac{dp_1}{d\delta_1^R}+\frac{d\alpha_1^R(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\tau^R\left(\mu_d+\frac{1}{2}\sigma_d^2-p_1\right)-\alpha_1^R(p_1)\,\tau^R\frac{dp_1}{d\delta_1^R} \\
& -\frac{d\alpha_1^{IS}(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\frac{1}{m}\sigma_d^2=0 \\
\Longrightarrow\quad & \frac{dp_1}{d\delta_1^R}=\frac{1}{1-\frac{1}{\alpha_1^R(p_1)\tau^R}\left(\frac{d\alpha_1^R(p_1)}{dp_1}\tau^R\delta_1^R+\frac{d\alpha_1^R(p_1)}{dp_1}\tau^R\left(\mu_d+\frac{1}{2}\sigma_d^2-p_1\right)-\frac{d\alpha_1^{IS}(p_1)}{dp_1}\frac{1}{m}\sigma_d^2\right)}
\end{aligned}$$

□

A1.10 Proof of Proposition 4

Proof. To derive the time-0 equilibrium price, substitute the optimal portfolio choices of the three investors, (33), (27), and (29) into the market clearing condition (35),

$$\begin{aligned}
& \left(\alpha_0^R(p_0)\,\tau^R+\left(1-\alpha_0^R(p_0)\right)\tau^I\right)\left(\frac{\mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right]-p_0}{\sigma_0^2}+\frac{1}{2}\right)+\sum_i\frac{\alpha_0^i(p_0)\,\tau^i\delta_0^i}{\sigma_0^2}=1 \\
\Longrightarrow\quad & \tau_0(p_0)\left(\mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right]-p_0+\frac{1}{2}\sigma_0^2\right)+\sum_i\alpha_0^i(p_0)\,\tau^i\delta_0^i=\sigma_0^2 \\
\Longrightarrow\quad & p_0=\mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right]+\left(\frac{1}{2}\sigma_0^2+\frac{\sum_i\alpha_0^i(p_0)\,\tau^i\delta_0^i-\sigma_0^2}{\tau_0(p_0)}\right)
\end{aligned}$$

where

$$\tau_0(p_0)\ \equiv\ \sum_i\alpha_0^i(p_0)\,\tau^i=\alpha_0^R(p_0)\,\tau^R+\left(1-\alpha_0^R(p_0)\right)\tau^I$$

The rest of the proof follows Proposition 1. □

A1.11 Proof of Lemma 2

Proof. First compute the m-th moment of d_j^{in} on the support $[d_{\min}, d_{\max}(N)]$.

$$
\begin{aligned}
\mathbb{E}\left[\left(d_j^{in}\right)^m\right] &= \int_{d_{\min}}^{d_{\max}(N)} x^m \frac{\xi-1}{d_{\min}}\left(\frac{x}{d_{\min}}\right)^{-\xi} dx \\
&= \frac{\xi-1}{d_{\min}^{1-\xi}} \int_{d_{\min}}^{d_{\max}(N)} x^{m-\xi} dx \\
&= \frac{\xi-1}{d_{\min}^{1-\xi}} \frac{1}{m+1-\xi} x^{m+1-\xi}\Big|_{d_{\min}}^{d_{\max}(N)} \\
&= \frac{\xi-1}{\xi-m-1} \frac{1}{d_{\min}^{1-\xi}} \left(d_{\min}^{m+1-\xi} - \left(d_{\max}(N)\right)^{m+1-\xi}\right)
\end{aligned}
$$

The variance of d_j^{in} is thus

$$
\begin{aligned}
\mathrm{Var}\left(d_j^{in}\right) &= \mathbb{E}\left[\left(d_j^{in}\right)^2\right] - \left(\mathbb{E}\left[d_j^{in}\right]\right)^2 \\
&= \frac{\xi-1}{\xi-3} \frac{1}{d_{\min}^{1-\xi}} \left(d_{\min}^{3-\xi} - \left(d_{\max}(N)\right)^{3-\xi}\right) - \left(\frac{\xi-1}{\xi-2}\right)^2 \frac{1}{d_{\min}^{2-2\xi}} \left(d_{\min}^{2-\xi} - \left(d_{\max}(N)\right)^{2-\xi}\right)^2 \\
&= \frac{\xi-1}{3-\xi} \frac{1}{d_{\min}^{1-\xi}} \left(\left(d_{\max}(N)\right)^{3-\xi} - d_{\min}^{3-\xi}\right) - \left(\frac{\xi-1}{\xi-2}\right)^2 \frac{1}{d_{\min}^{2-2\xi}} \left(d_{\min}^{2-\xi} - \left(d_{\max}(N)\right)^{2-\xi}\right)^2
\end{aligned}
$$

□

A1.12 Proof of Proposition 5

Proof. The proof follows [Acemogrlu].

Define

$$
\hat{P}_N(x) = \frac{1}{N}\left|\left\{j \in \mathcal{I}_N : d_j^{in} > x\right\}\right| = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}\left\{d_j^{in} > x\right\}
$$

Let $\boldsymbol{B} = \{b_1, b_2, \cdots, b_m\}$ denote the set of values d_j^{in} takes, with $b_1 < b_2 < \cdots < b_m$, and the convention that $b_0 = 0$.

First compute

$$
\begin{aligned}
\sum_{j=1}^{N} \theta_j^2 &= \sum_{j=1}^{N} \left(d_j^{in}\right)^2 = N \sum_{k=1}^{m} \left(b_k\right)^2 \left(\hat{P}_N\left(b_{k-1}\right) - \hat{P}_N\left(b_k\right)\right) \\
&= N\left(b_1^2\left(\hat{P}_N\left(b_0\right) - \hat{P}_N\left(b_1\right)\right) + b_2^2\left(\hat{P}_N\left(b_1\right) - \hat{P}_N\left(b_2\right)\right) + \cdots + b_m^2\left(\hat{P}_N\left(b_{m-1}\right) - \hat{P}_N\left(b_m\right)\right)\right) \\
&= N\left(\left(b_1^2 - b_0^2\right)\hat{P}_N\left(b_0\right) + \left(b_2^2 - b_1^2\right)\hat{P}_N\left(b_1\right) + \cdots + \left(b_m^2 - b_{m-1}^2\right)\hat{P}_N\left(b_{m-1}\right) - b_m^2 \hat{P}_N\left(b_m\right)\right) \\
&= N \sum_{k=0}^{m-1} \left(b_{k+1}^2 - b_k^2\right)\hat{P}_N\left(b_k\right) \\
&= N \sum_{k=0}^{m-1} \left(b_{k+1} + b_k\right)\left(b_{k+1} - b_k\right)\hat{P}_N\left(b_k\right) \\
&= 2N \sum_{k=0}^{m-1} \left(\frac{b_k + b_{k+1}}{2}\right)\left(b_{k+1} - b_k\right)\hat{P}_N\left(b_k\right)
\end{aligned}
$$

Then I use the continuous distribution to approximate the empirical distribution

$$
\begin{aligned}
\sum_{j=1}^{N} \theta_j^2 &= 2N \int_{d_{\min}}^{d_{\max}(N)} x \left(\frac{x}{d_{\min}}\right)^{1-\xi} dx \\
&= 2N \int_{d_{\min}}^{d_{\max}(N)} x \frac{d_{\min}}{2-\xi} d\left(\frac{x}{d_{\min}}\right)^{2-\xi} \\
&= 2N \frac{d_{\min}}{2-\xi}\left(x\left(\frac{x}{d_{\min}}\right)^{2-\xi}\Bigg|_{d_{\min}}^{d_{\max}(N)} - \int_{d_{\min}}^{d_{\max}(N)} \left(\frac{x}{d_{\min}}\right)^{2-\xi} dx \right) \\
&= 2N \frac{d_{\min}}{2-\xi}\left(x\left(\frac{x}{d_{\min}}\right)^{2-\xi}\Bigg|_{d_{\min}}^{d_{\max}(N)} - \frac{d_{\min}}{3-\xi}\left(\frac{x}{d_{\min}}\right)^{3-\xi}\Bigg|_{d_{\min}}^{d_{\max}(N)} \right) \\
&= 2N \frac{d_{\min}}{2-\xi}\left(d_{\max}(N)\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{2-\xi} - d_{\min} - \frac{d_{\min}}{3-\xi}\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{3-\xi} + \frac{d_{\min}}{3-\xi} \right) \\
&= 2N \frac{d_{\min}}{2-\xi}\left(\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{2-\xi}\left(d_{\max}(N) - \frac{1}{3-\xi} d_{\max}(N)\right) - \left(d_{\min} - \frac{d_{\min}}{3-\xi}\right) \right) \\
&= 2N \frac{d_{\min}}{2-\xi}\left(\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{2-\xi} \frac{2-\xi}{3-\xi} d_{\max}(N) - \frac{2-\xi}{3-\xi} d_{\min} \right)
\end{aligned}
$$

The time-0 conditional variance of δ_1^R is

$$\begin{aligned}
\mathrm{Var}_0\left(\delta_1^R\right) &= \frac{1}{N^2}\sum_{j=1}^{N}\theta_j^2\sigma_\varepsilon^2 \\
&= \frac{1}{N}\frac{2d_{\min}}{2-\xi}\sigma_\varepsilon^2\left(\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{2-\xi}\frac{2-\xi}{3-\xi}d_{\max}(N)-\frac{2-\xi}{3-\xi}d_{\min}\right) \\
&= \frac{2d_{\min}}{N}\frac{1}{3-\xi}\left(\left(\frac{d_{\max}(N)}{d_{\min}}\right)^{2-\xi}d_{\max}(N)-d_{\min}\right)\sigma_\varepsilon^2 \\
&= \frac{2d_{\min}^{\xi-1}}{N}\frac{1}{3-\xi}\left((d_{\max}(N))^{3-\xi}-d_{\min}^{3-\xi}\right)\sigma_\varepsilon^2 \\
&= O\left(N^{\frac{4-2\xi}{\xi-1}}\right)
\end{aligned}$$

where the last equality uses $d_{\max}(N) = O\left(N^{\frac{1}{\xi-1}}\right)$. Hence,

$$\sqrt{\mathrm{Var}_0\left(\delta_1^R\right)} = O\left(N^{\frac{2-\xi}{\xi-1}}\right)$$

□

A1.13 Distribution of time-1 aggregate retail sentiment shock

Define $c_j \equiv \frac{1}{N}d_j^{in}$, and the random variable $X_j = \mu + \varepsilon_1^j$, $\mu = \delta_0^R$. Let σ^2 denote the pre-truncation variance of ε_1^j, then X_j follows a truncated normal distribution on $[-\bar{\varepsilon}, \bar{\varepsilon}]$ with pre-truncation mean μ and variance σ^2, and X_j is i.i.d. in the cross section. Further define $\rho \equiv \frac{\bar{\varepsilon}}{\sigma}$, $a = \mu - \rho\sigma$, $b = \mu + \rho\sigma$. Then the PDF of X_j is

$$f_{X_j}(x) = \frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)} = \frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\rho)-1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of a standard normal random variable, respectively.

The time-1 aggregate retail sentiment shock δ_1^R can be written as

$$\delta_1^R = \sum_{j=1}^{N} c_j X_j$$

Hence, the characteristic function of δ_1^R is

$$\begin{aligned}
\varphi_{\delta_1^R}(t) &= \varphi_{X_1}(c_1 t)\,\varphi_{X_2}(c_2 t)\cdots\varphi_{X_N}(c_N t)\\
&= \prod_{j=1}^{N}\varphi_{X_j}(c_j t) = \prod_{j=1}^{N}\mathbb{E}\left[e^{itc_jX_j}\right]\\
&= \prod_{j=1}^{N}\left[\int_a^b e^{itc_jx}\frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\rho)-1}dx\right]
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_a^b e^{itc_jx}\frac{1}{\sigma_j}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\rho)-1}dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\exp\left(itc_jx-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2-2\mu x+\mu^2-2itc_jx\sigma^2}{2\sigma^2}\right)dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\exp\left(\frac{\left(\mu+itc_j\sigma^2\right)^2-\mu^2}{2\sigma^2}\right)\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(x-\left(\mu+itc_j\sigma^2\right)\right)^2}{2\sigma^2}\right)dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(x-\left(\mu+c_j\sigma^2it\right)\right)^2}{2\sigma^2}\right)dx
\end{aligned}$$

Define $y\equiv\frac{x-\left(\mu+c_j\sigma^2it\right)}{\sigma}\implies x=\sigma y+\left(\mu+c_j\sigma^2it\right)\implies dx=\sigma dy$. And note that $\frac{a-\left(\mu+c_j\sigma^2it\right)}{\sigma}=-\rho-c_j\sigma it$, $\frac{b-\left(\mu+c_j\sigma^2it\right)}{\sigma}=\rho-c_j\sigma it$. Then

$$\begin{aligned}
&\int_a^b e^{itc_jx}\frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\rho)-1}dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(x-\left(\mu+c_j\sigma^2it\right)\right)^2}{2\sigma^2}\right)dx\\
=&\ \frac{1}{2\Phi(\rho)-1}\exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\int_{\frac{a-\left(\mu+c_j\sigma^2it\right)}{\sigma}}^{\frac{b-\left(\mu+c_j\sigma_j^2it\right)}{\sigma}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy\\
=&\ \frac{1}{2\Phi(\rho)-1}\exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\int_{-\rho-c_j\sigma it}^{\rho-c_j\sigma it}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy\\
=&\ \exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\frac{\Phi(\rho-c_j\sigma it)-\Phi(-\rho-c_j\sigma it)}{2\Phi(\rho)-1}\\
=&\ \exp\left(c_j\mu it-\frac{1}{2}c_j^2\sigma^2t^2\right)\frac{\Phi(\rho-c_j\sigma it)+\Phi(\rho+c_j\sigma it)-1}{2\Phi(\rho)-1}
\end{aligned}$$

Hence,

$$\begin{aligned}
\varphi_{S_n}(t) &= \prod_{j=1}^{n}\left[\int_a^b e^{itc_j x}\frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\rho)-1}dx\right] \\
&= \exp\left(\left(\sum_{j=1}^{n} c_j\mu\right)it-\frac{1}{2}\left(\sum_{j=1}^{n} c_j^2\sigma^2\right)t^2\right)\prod_{j=1}^{n}\frac{\Phi(\rho-c_j\sigma it)+\Phi(\rho+c_j\sigma_j it)-1}{2\Phi(\rho)-1}
\end{aligned}$$

The characteristic function of δ_1^R is

$$\varphi_{S_n}(t)=\exp\left(\left(\sum_{j=1}^{N} c_j\mu\right)it-\frac{1}{2}\left(\sum_{j=1}^{N} c_j^2\sigma^2\right)t^2\right)\prod_{j=1}^{N}\frac{\Phi(\rho-c_j\sigma it)+\Phi(\rho+c_j\sigma it)-1}{2\Phi(\rho)-1}$$

$$\Longrightarrow \quad \varphi_{\delta_1}(t)=\exp\left(\left(\sum_{j=1}^{N} c_j\right)\mu it-\frac{1}{2}\left(\sum_{j=1}^{N} c_j^2\right)\sigma_\varepsilon^2 t^2\right)\prod_{j=1}^{N}\frac{\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}-c_j\sigma_\varepsilon it\right)+\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}+c_j\sigma_\varepsilon it\right)-1}{2\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}\right)-1}$$

$$\Longrightarrow \quad \varphi_{\delta_1}(t)=\exp\left(\mu it-\frac{1}{2}\left(\sum_{j=1}^{N} c_j^2\right)\sigma_\varepsilon^2 t^2\right)\prod_{j=1}^{N}\frac{\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}-c_j\sigma_\varepsilon it\right)+\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}+c_j\sigma_\varepsilon it\right)-1}{2\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}\right)-1}$$

Compare the characteristic function of δ_1^R with another random variable $\tilde{\delta}_1$, which follows a truncated normal distribution on $[\mu-\bar{\varepsilon},\mu+\bar{\varepsilon}]$, with mean $\sum_{j=1}^{N} c_j\mu=\mu$ and variance $\sum_{j=1}^{N} c_j^2\sigma_\varepsilon^2$.

$$\begin{aligned}
\varphi_{\tilde{\delta}_1}(t) &= \exp\left(\mu it-\frac{1}{2}\left(\sum_{j=1}^{N} c_j^2\right)\sigma_\varepsilon^2 t^2\right) \\
&\cdot\frac{\Phi\left(\frac{\bar{\varepsilon}}{\sqrt{\sum_{j=1}^{N} c_j^2}\sigma_\varepsilon}-\sqrt{\sum_{j=1}^{N} c_j^2}\sigma_\varepsilon it\right)+\Phi\left(\frac{\bar{\varepsilon}}{\sqrt{\sum_{j=1}^{N} c_j^2}\sigma_\varepsilon}+\sqrt{\sum_{j=1}^{N} c_j^2}\sigma_\varepsilon it\right)-1}{2\Phi\left(\frac{\bar{\varepsilon}}{\sqrt{\sum_{j=1}^{N} c_j^2}\sigma_\varepsilon}\right)-1}
\end{aligned}$$

Hence, the distribution of δ_1^R can be approximated by a truncated normal distribution, if the cross sectional distribution of c_j is skewed.

A2 Reddit data

A2.1 Variable definitions

I construct two data frames following the steps in Section 2.1.1 – one includes all the submissions, and the other includes all the comments.

In the data frame of submissions, each row is a unique submission. And it has the following fields:

- `id`: the unique id of the submission, e.g., "eifjq5". I add the prefix "t3_" to the submission `id` to facilitate the mapping between the submission and its associated comments.
- `author`: the name of the author of the submission, e.g., "Ituglobal".
- `author_fullname`: the unique user id of the author of the submission, prefixed by "t2_", e.g., "t2_6rjw5".
- `created_utc`: the UTC date and time at which the submission was created.
- `title`: the textual content of the title of the submission.
- `selftext`: the textual content of the body text of the submission.

In the data frame of comments, each row is a unique comment. And it has the following fields:

- `id`: the unique id of the comment, e.g., "fctzgly". I add the prefix "t1_" to the `id` to facilitate the mapping between the comment in question and its parent comment.
- `link_id`: the unique id of the submission that the comment in question replies to, e.g., "t3_eiwx9h".
- `parent_id`: the unique id of the parent comment (or submission) of the comment in question. If the comment is a reply to another comment, then the is prefixed by "t1_". Otherwise, it is a reply to a submission, and it's prefixed by "t3_".
- `created_utc`: the UTC date and time at which the comment was created.
- `author`: the name of the author of the comment, e.g., "urfriendosvendo".
- `author_fullname`: the unique user id of the author of the comment, prefixed by "t2_", e.g., "t2_12ol3k".
- `body`: the textual content of the comment.

A2.2 Constructing the sample of submissions and comments

I first run the following algorithm to tag submissions and comments with stock tickers, and then select samples of submissions and comments.

1. Retrieve the list of tickers of CRSP common stocks.

2. Search for stock tickers in the text of the submission.[1]

    (a) First pass search: search for CRSP stock tickers in the augmented body text[2].

    i. Preprocess the augmented body text in the following order:
        - Replace ‘’ / - with space.
        - Replace & with space if it appears between words.
        - Replace . with space.
        - Remove all other punctuation marks.
        - Tokenize augmented body text and only keep non-empty tokens.

    ii. Search for CRSP stock tickers in the augmented body text in a case-insensitive way. A submission is tagged with a ticker if the ticker can be found in the list of tokens.

    (b) Manually go over the matched tickers, add $ sign in front of those tickers that are common words, and use this updated list of tickers in the second pass search.

    (c) Second pass search: repeat the procedures in the first pass search, but using the updated list of tickers from the previous step.

3. Drop submissions where `author_fullname` is empty, or "[deleted]", or "[removed]". I also drop those where `id` is empty, or "[deleted]", or "[removed]".

4. Drop submissions where `author` is one of the bots in Table A1.

5. Only keep submissions tagged with at least one CRSP common stock ticker, and only keep the comments associated with these selected submissions (see Appendix A2.3 below for the procedure of matching submissions with comments).

    If a submission is tagged with a ticker, then the associated comments are also tagged with the same ticker. A submission or comment can be tagged with multiple stock tickers.

[1] For GameStop, I search for both its ticker "GME" and the company name "GameStop".

[2] A submission has its title and body text. I obtain the augmented body text by appending the body text to the title, separated by a white space.

Finally, I construct the following two samples of submissions and comments:

- Sample of submissions and comments for CRSP common stocks, by performing steps 1-5 above.
- Sample of all submissions and comments, by performing steps 1-4 above.

For each of the sample, I keep one data frame for submissions and another data frame for comments, with the structure described in Appendix A2.1. And I construct the network using these two data frames.

A2.3 Constructing the network

As is described in Appendix A2.1, the submission data frame and the comment data frame has a common field – the field `id` in the submission data frame corresponds to the `link_id` in the comment data frame. This allows me to recover the comment tree described in.

For each of the sample described in Appendix A2.2, I merge the submission data frame and comment data frame by the common field described above, and only keep submissions with at least one comment. In the merged dataset, each row corresponds to a comment, with information on the author of the comment, and the author of the submission that the comment replies to. This allows me to construct the network of users from the commenting relationship.

A3 FactSet data

I following the procedure in Gabaix and Koijen (2022) and Koijen et al. (2022):

1. Merge the holdings data (`[own_v5].[own_inst_eq_v5].[own_inst_13f_detail]`) with the entity sub type data (`[own_v5].[own_hub_ent_v5].[own_ent_institutions]`), by `factset_entity_id`.

   Each record in this merged dataset corresponds to a filer entity (with unique id `factset_entity_id`).

2. For those filer entities with missing entity sub type (from the previous step), find the corresponding roll-up entity (from `[own_v5].[own_hub_ent_v5].[own_ent_13f_combined_inst]`), and assign the sub type of the roll-up entity to the filer entity.

- To identify the sub type of the roll-up entity: merge the roll-up entity data
([own_v5].[own_hub_ent_v5].[own_ent_13f_combined_inst])
with the entity sub type data
([own_v5].[own_hub_ent_v5].[own_ent_institutions]),
by `factset_rollup_entity_id` in the former (`factset_entity_id` in the latter).
$\Longrightarrow$ 12,276 out of the 12,295 roll-up entities have non-missing entity sub type.

3. Classify institutions into six types using `entity_sub_type`:

- Hedge Funds: `entity_sub_type` = "AR", "FH", "FF", "FU", "FS", "HF".
- Brokers: `entity_sub_type` = "BM", "IB", "ST", "MM", "BR".
- Private Banking: `entity_sub_type` = "CP", "FY", "VC", "PB".
- Investment Advisors: `entity_sub_type` = "IC", "RE", "PP", "SB", "MF", "IA".
- Long-Term Investors: `entity_sub_type` = "FO", "SV", "IN", "PF".

A4 Modified BJZZ algorithm to identify retail trades

1. Start with any trade with price not at the midpoint of bid and ask.
2. Match the NBBO to the timestamp of the trade, and then compute bid-ask spread quoted before the trade.
3. If the spread quoted before the trade is one cent, use the original BJZZ algorithm to sign the trade.
4. If the trade price is outside the bid-ask spread, use the original BJZZ algorithm to sign the trade.
5. Otherwise, if the trade is below the midpoint, label the trade as a sell. If the trade is above the midpoint, label the trade as a buy.

   I also implement the $[0.4, 0.6]$ "donut" in this step, as in the original BJZZ algorithm.

A5 Fitting power-law distribution

For each calendar day t, I fit a power-law distribution to the vector of user influence, $\left(d^{in}_{1,t}, d^{in}_{2,t}, \cdots, d^{in}_{N_t,t}\right)^\intercal$ computed in Section 2.1.3, and estimate the exponent $\hat{\xi}_t$ and the threshold value $\hat{d}^{in}_{\min,t}$. Following Rantala (2019), I use maximum likelihood method to estimate these parameters. Specifically, I use the power.law.fit function of the igraph package in R, with the "plfit" implementation.

I use bootstrap methods to compute the confidence intervals. The steps are:

1. Generate a bootstrap sample $\left\{d_{k,t}^{in}\left(b\right)\right\}_{k=1}^{N_t}$ by sampling the original data $\left(d_{1,t}^{in}, d_{2,t}^{in}, \cdots, d_{N_t,t}^{in}\right)^{\intercal}$ randomly with replacement.

2. Estimate the parameters $\xi_t\left(b\right)$ and $d_{\min,t}\left(b\right)$ for this bootstrapped sample, using the maximum likelihood method described above.

3. Repeat steps 1 and 2 for $B = 5000$ times, and obtain the vector of estimates $\left\{\xi_t\left(b\right)\right\}_{b=1}^{B}$, $\left\{d_{\min,t}\left(b\right)\right\}_{b=1}^{B}$.

4. For the $\hat{\xi}_t$ estimate, the lower (upper) bound of the 95% confidence interval is the 2.5th (97.5th) percentile of the empirical distribution $\left\{\xi_t\left(b\right)\right\}_{b=1}^{B}$. Similarly, for the $\hat{d}_{min,t}$ estimate, the lower (upper) bound of the 95% confidence interval is the 2.5th (97.5th) percentile of the empirical distribution $\left\{d_{\min,t}\left(b\right)\right\}_{b=1}^{B}$.

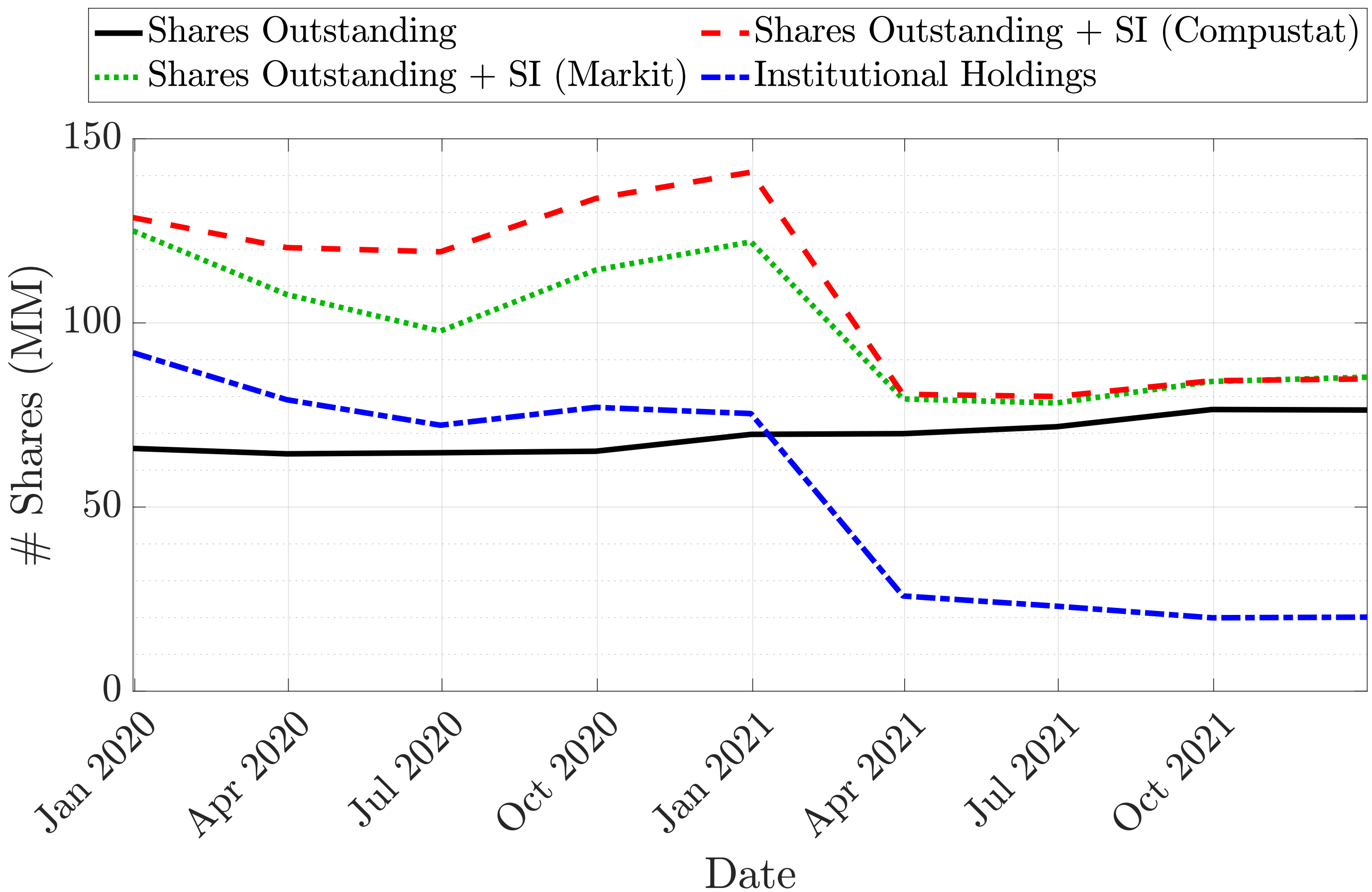


Figure A1. Shares outstanding and institutional ownership of GameStop. This figure compares the number of shares outstanding with institutional ownership of GameStop.

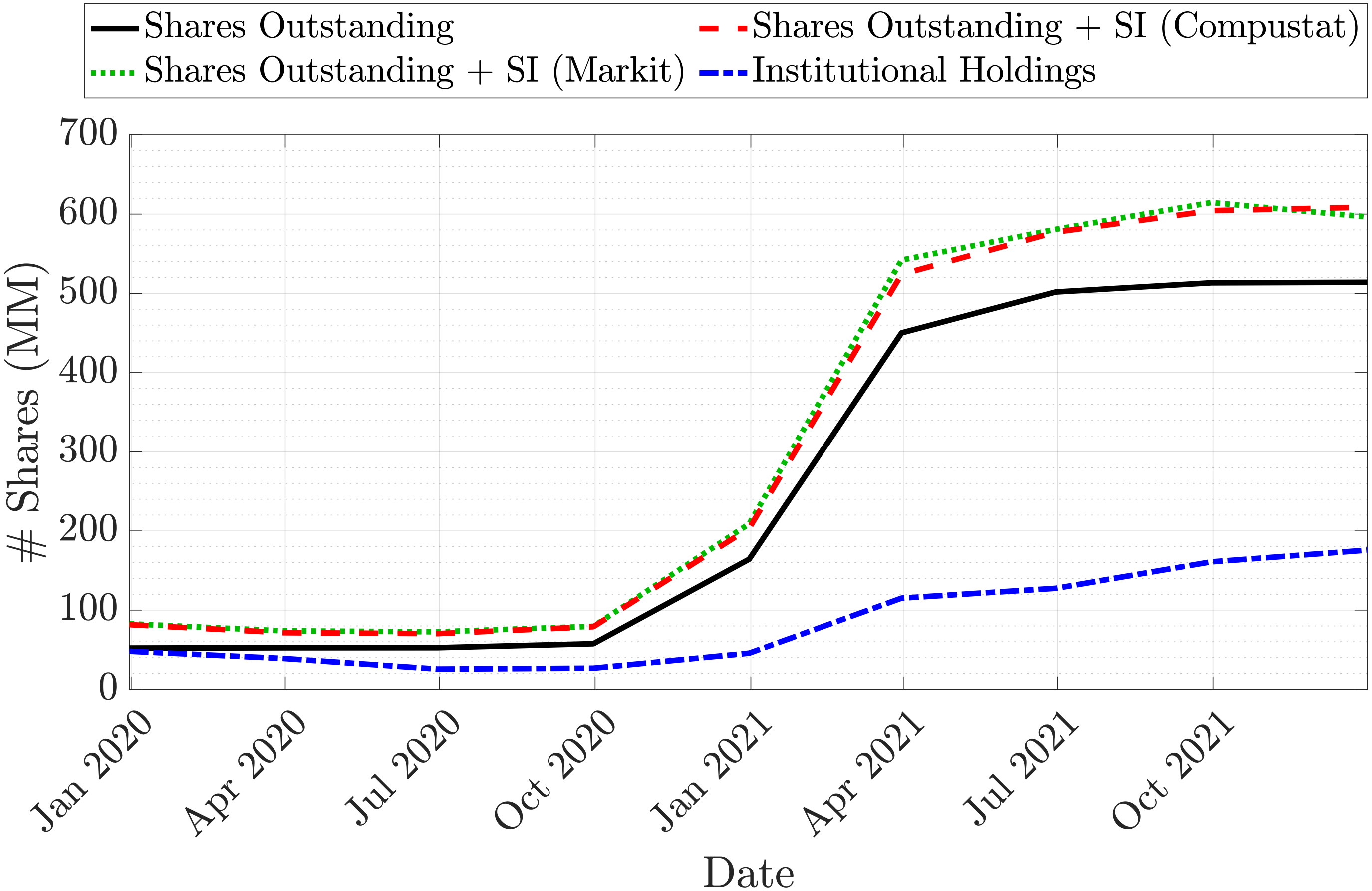


Figure A2. Shares outstanding and institutional ownership of AMC. This figure compares the number of shares outstanding with institutional ownership of AMC.

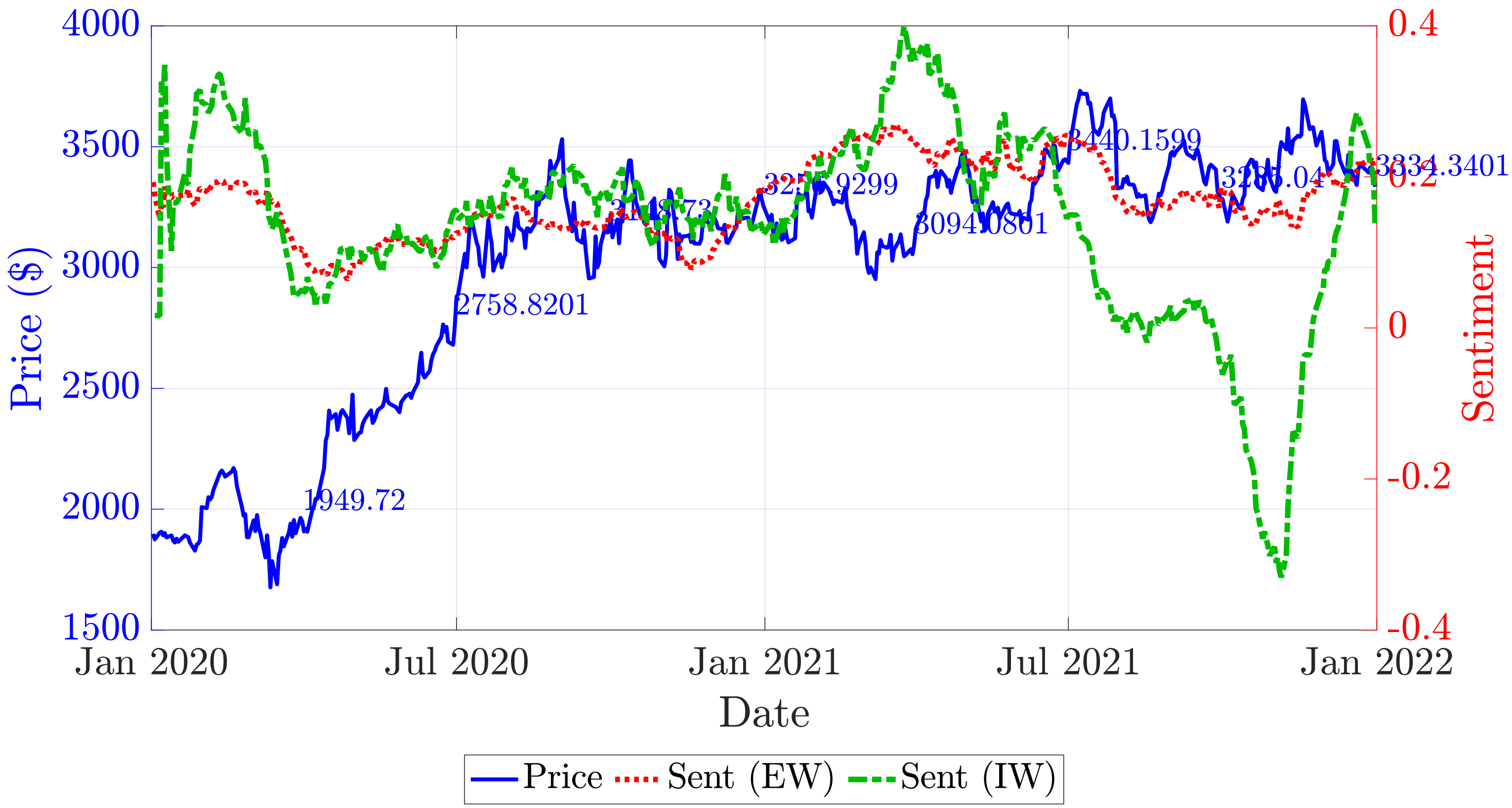


Figure A3. Price and sentiment of Amazon. This figure plots the daily close price (solid blue line), equal-weighted sentiment (dotted red line), and influence-weighted sentiment (dash-dotted green line) of Amazon. The sentiment series are 30-day moving averages.

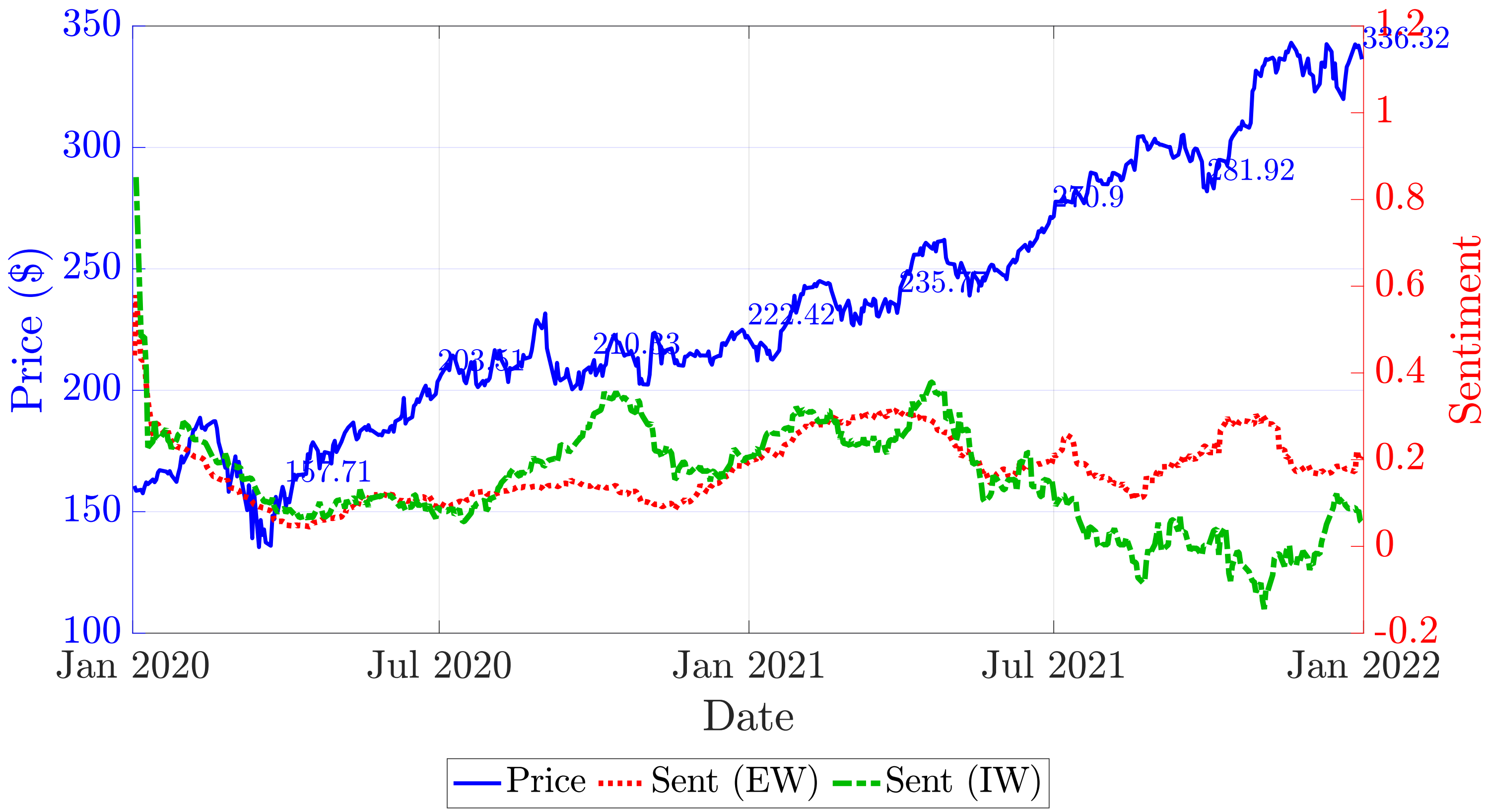


Figure A4. Price and sentiment of Microsoft. This figure plots the daily close price (solid blue line), equal-weighted sentiment (dotted red line), and influence-weighted sentiment (dash-dotted green line) of Microsoft. The sentiment series are 30-day moving averages.

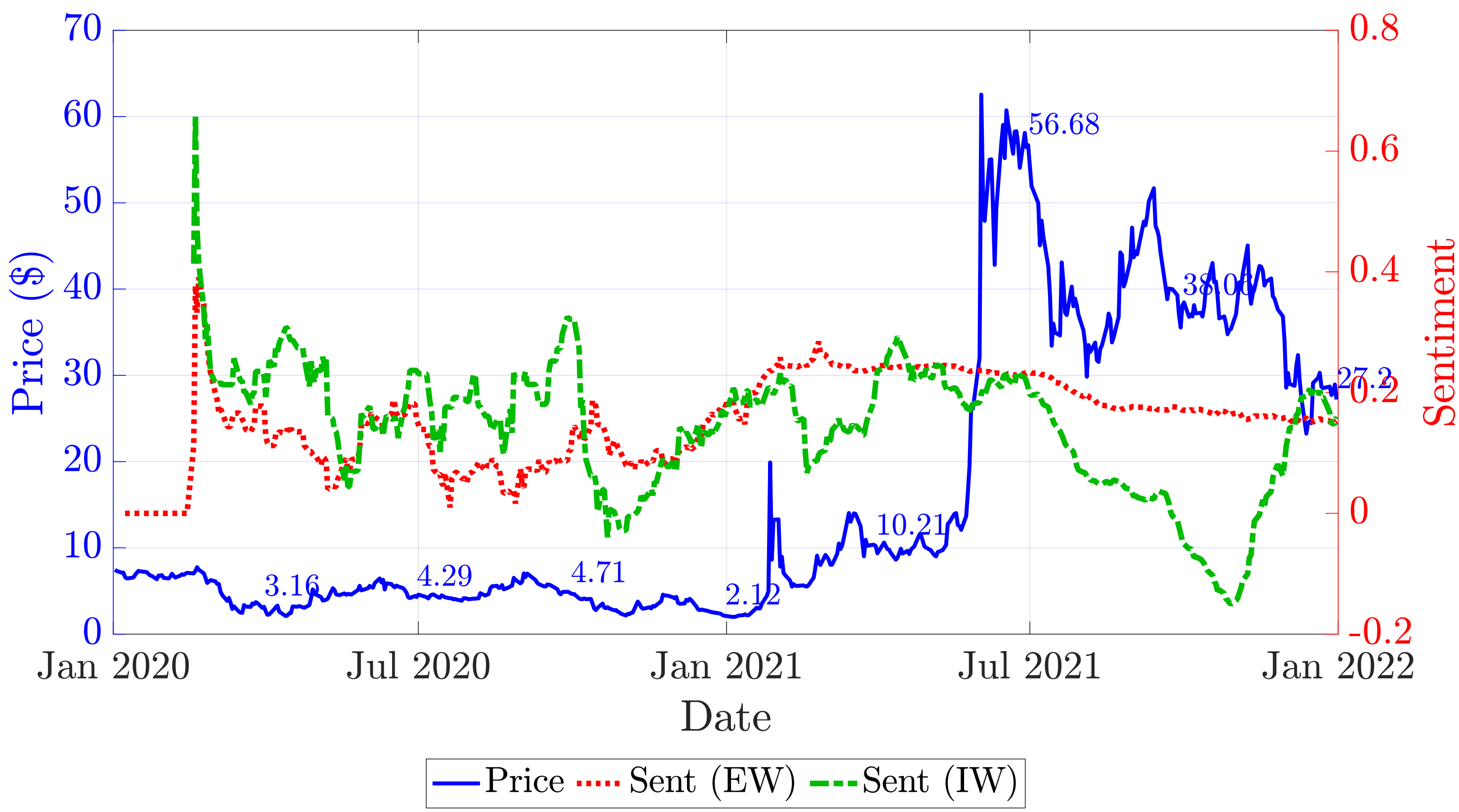


Figure A5. Price and sentiment of AMC. This figure plots the daily close price (solid blue line), equal-weighted sentiment (dotted red line), and influence-weighted sentiment (dash-dotted green line) of AMC. The sentiment series are 30-day moving averages.

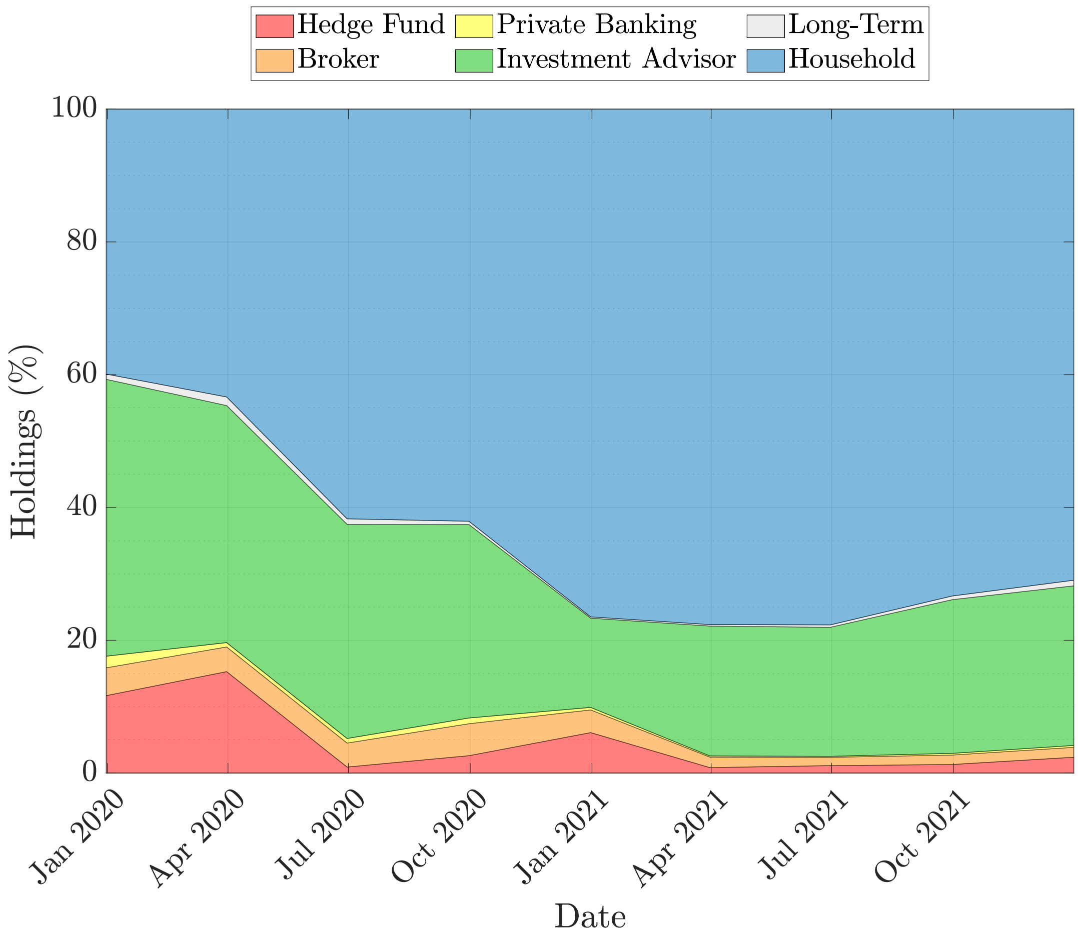
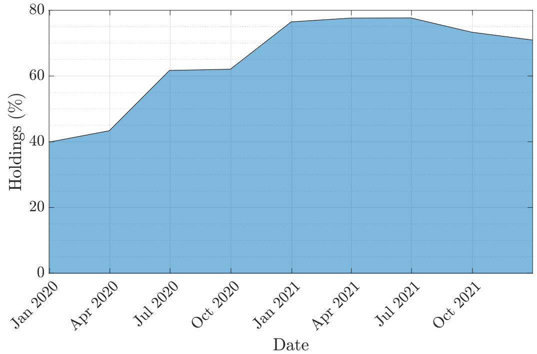
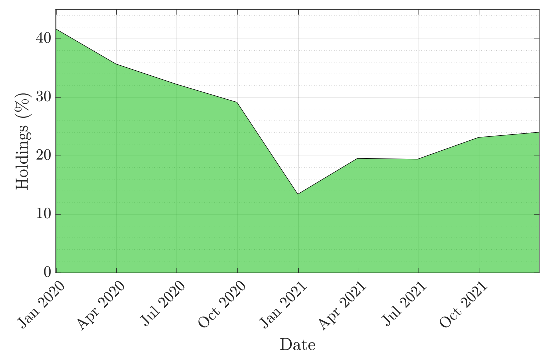


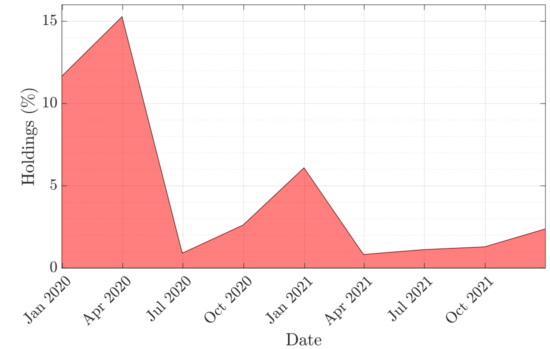
Figure A6. Holdings of long investors in AMC. This figure plots the holdings of long investors in AMC. The y axis is the number of shares held divided by number of shares outstanding plus number of shares sold short.



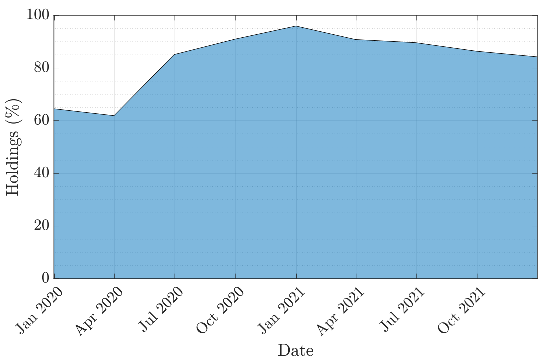
(a) Households / (SHROUT + SS)



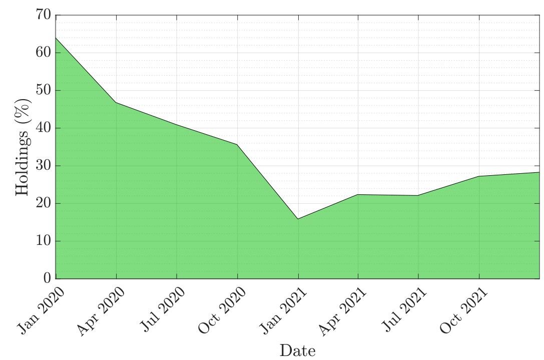
(b) Investment Advisors / (SHROUT + SS)



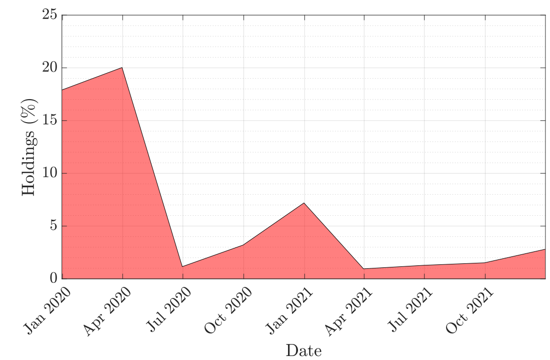
(c) Hedge Funds / (SHROUT + SS)



(d) Households / SHROUT



(e) Investment Advisors / SHROUT



(f) Hedge Funds / SHROUT

Figure A7. Holdings of AMC by investor group. This figure plots the holdings of Households, Investment Advisors and Hedge Funds in AMC. For panel (a), (b), (c), the denominator is the number of shares outstanding plus number of shares sold short. For panel (d), (e), (f), the denominator is the number of shares outstanding.

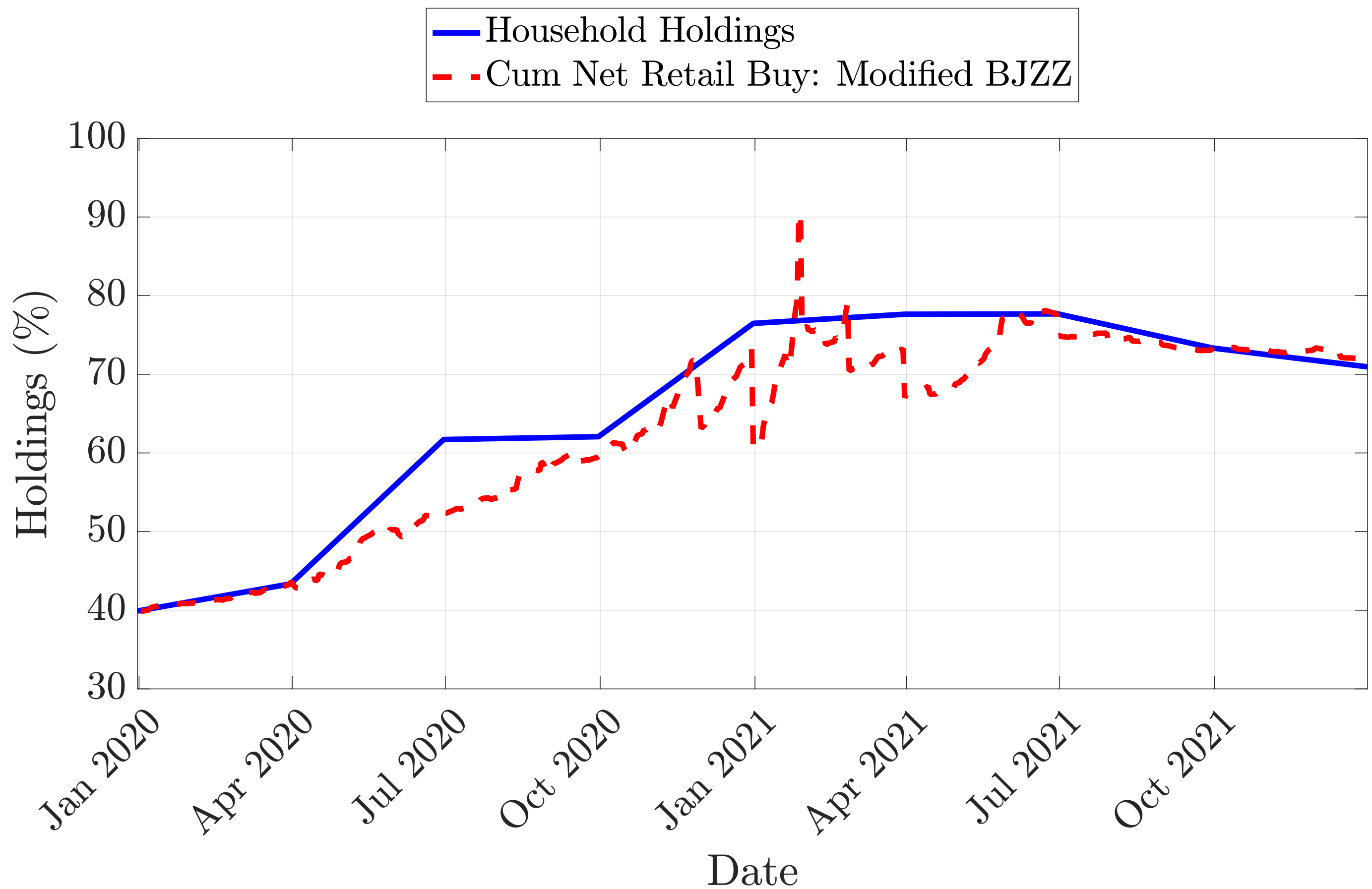


Figure A8. Household holdings versus cumulative net retail buy volume for AMC. This figure plots the quarterly household holdings of AMC (solid blue line) versus the daily cumulative net retail buy volume (dotted red line). The denominator for both series is number of shares outstanding plus number of shares sold short.

Figure A9. Price and short interest of AMC. This figure plots the daily close price of AMC (solid blue line), and the daily short interest (dotted red line). The short interest is the number of shares sold short divided by number of shares outstanding.

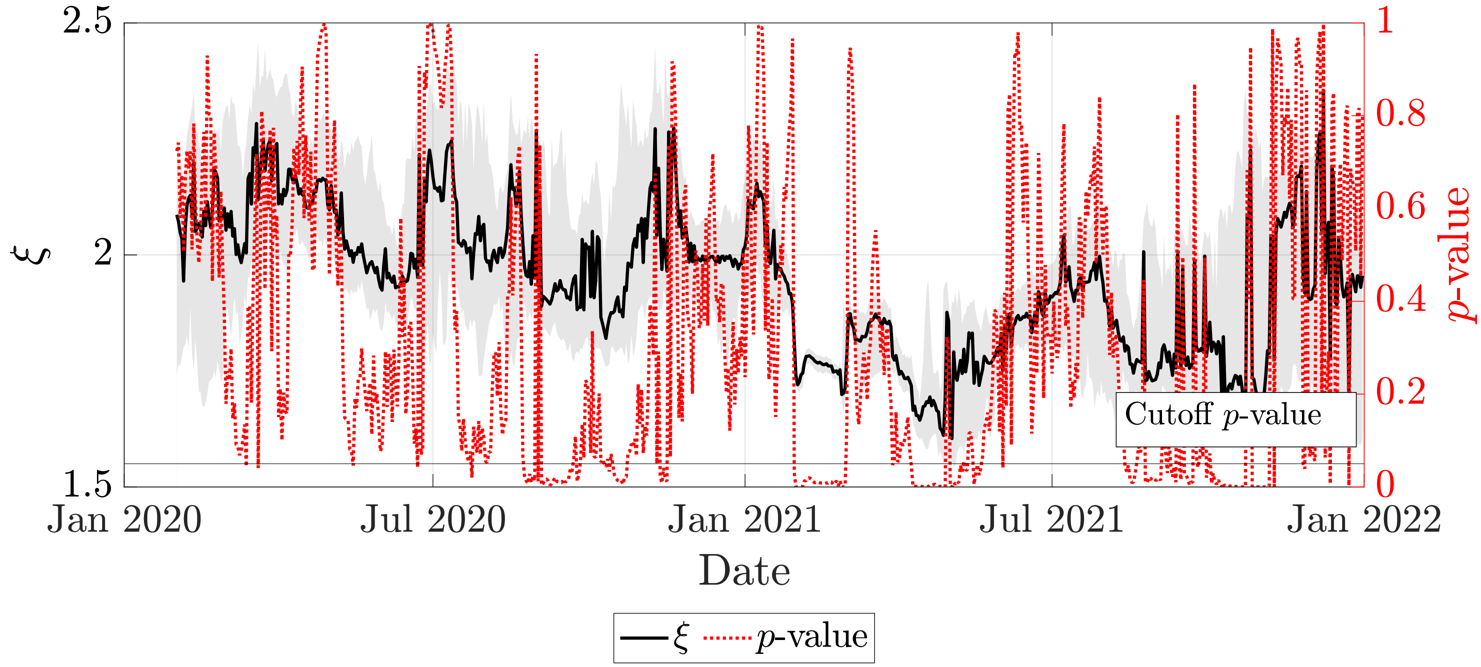


Figure A10. p**-value for power-law fitting.**

Table A1
Reddit Bots

This table shows the Reddit bots whose submissions are removed from the sample.

Bot Name
WSBVoteBot
RemindMeBot
Generic_Reddit_Bot
ReverseCaptioningBot
LimbRetrieval-Bot
NoGoogleAMPBot
RepostSleuthBot
GetVideoBot
CouldWouldShouldBot