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

Fulin Li*

November 22, 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

$$\begin{aligned}
\alpha^i_{t+1} &\equiv \frac{A^i_{t+1}}{P_{t+1}\bar{S}} \\
&= \frac{A^i_t\left(w^i_t\frac{P_{t+1}}{P_t}+1-w^i_t\right)}{P_{t+1}\bar{S}} \\
&= \frac{\alpha^i_t P_t\bar{S}\left(w^i_t\frac{P_{t+1}}{P_t}+1-w^i_t\right)}{P_{t+1}\bar{S}} \\
&= \alpha^i_t\left(\left(1-w^i_t\right)\exp\left(p_t-p_{t+1}\right)+w^i_t\right),
\end{aligned}$$

where the second line uses the budget constraint (22) together with 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.,

$$\sum_i A^i_t = P_t\bar{S}.$$

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 two 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 condition

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

From equation (A1), we can solve for the equilibrium price.

## A1.3 Optimal portfolio choice

### A1.3.1 Retail investors

Retail investor $j$ solves the following problem

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

The F.O.C. is

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

Substitute retail investors' subjective beliefs into the above expression, we get their demands for the risky asset.

- For a type-1 retail investor $j$, his time-0 and time-1 demands for the risky asset are

$$w_0^j = \tau^R \left( \frac{\mathbb{E}_0 [p_1] + y_0^j - p_0}{\sigma_0^2} + \frac{1}{2} \right), \tag{A2}$$

$$w_1^j = \tau^R \left( \frac{\mu_d + y_1^j - p_1}{\sigma_d^2} + \frac{1}{2} \right). \tag{A3}$$

- For a type-2 retail investor $j'$, his time-0 and time-1 demands for the risky asset are

$$w_0^{j'} = \tau^R \left( \frac{\mathbb{E}_0 [p_1] - p_0}{\sigma_0^2} + \frac{1}{2} \right), \tag{A4}$$

$$w_1^{j'} = \tau^R \left( \frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2} \right). \tag{A5}$$

#### A1.3.2 Long institution

The long institution $IL$ solves the following problem

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

Since the objective function is quadratic in portfolio weight $w_t^{IL}$ and has a global maximum, the solution to this constrained problem 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 (equations (14), (15), and (17)) into the above expression, we get his time-0 and time-1 demands for the risky asset

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

$$w_1^{IL} = \max\left\{0, \tau^I \left( \frac{\mu_d - p_1}{\sigma_d^2} + \frac{1}{2} \right)\right\}.$$

### A1.3.3 Short institution

The short institution $IS$ solves the following problem

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

The solution is

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

Substitute $IS$'s beliefs (equations (14), (16), and (17)) into the above expression, we get his time-0 and time-1 demands for the risky asset

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

## A1.4 Proof of Lemma 1

*Proof.* I prove the existence result in two steps. First, I show that the aggregate demand of the $\bar{N}$ retail investors is equal to the demand of the aggregate retail investor specified in equations (39) and (40), and thus the equilibrium price can be solved from the market clearing condition (42). Then I derive the wealth share dynamics of the aggregate retail investor in equation (41).

I begin by restating 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 $\alpha_t^j$. At time $t$ before trading, the $\bar{N}$ retail investors first split their aggregate wealth $\sum_{k=1}^{\bar{N}} A_t^k$ equally. In particular, they split their aggregate stock position and aggregate bond position equally. After that, retail investor $j$ has wealth $\hat{A}_t^j = \frac{1}{\bar{N}}\sum_{k=1}^{\bar{N}} A_t^k$ and wealth share

$$\hat{\alpha}_t^j \equiv \frac{\hat{A}_t^j}{A_t} = \frac{\frac{1}{\bar{N}}\sum_{k=1}^{\bar{N}} A_t^k}{A_t} = \frac{1}{\bar{N}}\sum_{k=1}^{\bar{N}} \alpha_t^k. \tag{A6}$$

Retail investors then trade with each other. Specifically, 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. After trading, his end-of-period wealth share becomes

$$\alpha_{t+1}^j = \hat{\alpha}_t^j \left( \left(1 - w_t^j\right) \exp\left(p_t - p_{t+1}\right) + w_t^j \right). \tag{A7}$$

Next, I show that the equilibrium price of the risky asset is the same as that in an economy with three investors – an aggregate retail investor, the long institution, and the short institution. And the demand of the aggregate retail investor is the sum of the demand of the $\bar{N}$ retail investors.

At time $t$, market clearing for the risky asset implies that

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

where the fourth line uses the definition of $\hat{\alpha}_t^j$ in equation (A6), and the fifth line uses the optimal portfolio weights of retail investors in equations (29)-(32).

Define

$$A_t^R \equiv \sum_{j=1}^{\bar{N}} A_t^j, \alpha_t^R \equiv \sum_{j=1}^{\bar{N}} \alpha_t^j, \tag{A8}$$

$$\delta_t^R \equiv \theta\left(N_t\right) \frac{1}{N_t} \sum_{j=1}^{N_t} y_t^j, \tag{A9}$$

$$w_t^R \equiv \tau^R \left( \frac{\mathbb{E}_t\left[p_{t+1}\right] + \delta_t^R - p_t}{\sigma_t^2} + \frac{1}{2} \right) = \frac{1}{\bar{N}} \sum_{j=1}^{\bar{N}} w_t^j. \tag{A10}$$

Then the market clearing condition can be written as

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

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

Hence, the equilibrium price of the risky asset is the same as that in an economy with three investors – an aggregate 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 an aggregate retail investor whose demand for the risky asset is given by equation (A10). The aggregate retail investor has constant relative risk tolerance $\tau^R = \frac{1}{\gamma^R}$ and subjective beliefs

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

Finally, I derive the wealth share dynamics of the aggregate retail investor. From the definition of $\alpha_{t+1}^R$ in (A8),

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

where the second line uses investor $j$'s wealth share dynamics in equation (A7), and the last line uses the aggregate retail investor's demand in equation (A10). □

## A1.5 Proof of Proposition 1

*Proof.* I focus on monotone equilibrium of Definition 1, with sentiment cutoffs $\delta_1^m, \delta_1^h$ 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 (42).

- 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, (40), (34), and (36) into the market clearing condition (42), I get

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

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 $\delta_1^m$ 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, (40), (34), and (36) into the market clearing condition (42), I get

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

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 $\delta_1^h$ 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, (40), (34), and (36) into the market clearing condition (42), I get

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

Define the function

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

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

□

### A1.6 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 $\delta_1^R$: $\frac{\partial G\left(p_1,\delta_1^R\right)}{\partial\delta_1} = 1 > 0$.*
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 $\delta_1^R$: $\frac{\partial^2 G\left(p_1,\delta_1^R\right)}{\partial p_1 \partial\delta_1^R} = 0$.*
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

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

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

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

From the wealth share dynamics, we get

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

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

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

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 $\delta_1^R$, 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. $\square$

## A1.7 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, $\forall p_1 > p_1^h$, $\forall \delta_1^R \in \left[\delta_1^h, \bar{\delta}_1\right]$,

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

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 $\delta_1^R$. Furthermore, $p_1\left(\delta_1^R\right)$ is continuous in $\delta_1^R$ 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 $x_1\left(\delta_1^R\right) \leq p_1^h < x_2\left(\delta_1^R\right)$. 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

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

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

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

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 $\delta_1^R$. Furthermore, $p_1\left(\delta_1^R\right)$ is continuous in $\delta_1^R$ on $\delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right]$, and is discontinuous at $\delta_1^R = \delta_1^h$.

□

## A1.8 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, (40), (34), (36), and the market clearing condition (42), 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, (40), (34), (36), and the market clearing condition (42), 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,

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

$$\implies \frac{dp_1}{d\delta_1^R} = \frac{\frac{\alpha_1^R(p_1)\tau^R}{\hat{\tau}_1(p_1)}}{1 - \frac{1}{\hat{\tau}_1(p_1)}\left(\frac{d\left(\alpha_1^R(p_1)\right)}{dp_1}\tau^R\delta_1^R + \frac{d\hat{\tau}_1(p_1)}{dp_1}\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right)\right)}$$

- High sentiment $\delta_1^R \in \left(\delta_1^h, \bar{\delta}_1\right)$: from the optimal portfolio choices of the three investors, (40), (34), (36), and the market clearing condition (42), we get

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

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

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

Using the implicit function theorem,

$$\alpha_1^R(p_1)\tau^R + \frac{d\left(\alpha_1^R(p_1)\tau^R\delta_1^R\right)}{dp_1}\frac{dp_1}{d\delta_1^R} + \frac{d\alpha_1^R(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\tau^R\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) - \alpha_1^R(p_1)\tau^R\frac{dp_1}{d\delta_1^R}$$
$$-\frac{d\alpha_1^{IS}(p_1)}{dp_1}\frac{dp_1}{d\delta_1^R}\frac{1}{m}\sigma_d^2 = 0$$

$$\implies \frac{dp_1}{d\delta_1^R} = \frac{1}{1 - \frac{1}{\alpha_1^R(p_1)\tau^R}\left(\frac{d\alpha_1^R(p_1)}{dp_1}\tau^R\delta_1^R + \frac{d\alpha_1^R(p_1)}{dp_1}\tau^R\left(\mu_d + \frac{1}{2}\sigma_d^2 - p_1\right) - \frac{d\alpha_1^{IS}(p_1)}{dp_1}\frac{1}{m}\sigma_d^2\right)}$$

□

## A1.9 Proof of Proposition 4

*Proof.* To derive the time-0 equilibrium price, substitute the optimal portfolio choices of the three investors, (39), (33), and (35) into the market clearing condition (42),

$$\left(\alpha_0^R(p_0)\tau^R + \left(1 - \alpha_0^R(p_0)\right)\tau^I\right)\left(\frac{\mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right] - p_0}{\sigma_0^2} + \frac{1}{2}\right) + \sum_i \frac{\alpha_0^i(p_0)\tau^i\delta_0^i}{\sigma_0^2} = 1$$

$$\implies \tau_0(p_0)\left(\mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right] - p_0 + \frac{1}{2}\sigma_0^2\right) + \sum_i \alpha_0^i(p_0)\tau^i\delta_0^i = \sigma_0^2$$

$$\implies p_0 = \mathbb{E}_0\left[p_1\left(\delta_1^R\right)\right] + \left(\frac{1}{2}\sigma_0^2 + \frac{\sum_i \alpha_0^i(p_0)\tau^i\delta_0^i - \sigma_0^2}{\tau_0(p_0)}\right)$$

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. □

## A1.10 Implicit price at time $-1$

I assume that, at time $-1$, investors do not anticipate future sentiment shocks. They believe that the prices at time 0 and 1 will reflect the present value of the terminal dividend, and the prices are deterministic. Hence, from time $-1$ to 0 and from time 0 to 1, the risky asset should have the same one-period return as the risky-free asset. This implies that $p_{-1} = \tilde{p}_0 = \tilde{p}_1$, where $\tilde{p}_0$ and $\tilde{p}_1$ denote investors' beliefs about time-0 and time-1 prices, respectively.

The implicit price $p_{-1}$ is such that investors do not want to trade at time $-1$. Since $p_{-1} = \tilde{p}_0 = \tilde{p}_1$, investors believe that they will not have incentives to trade at time 0 and 1, and thus they believe their asset positions and wealth shares remain constant from time $-1$ to time 1. In this case, the aggregate risk tolerance remains constant from time $-1$ to time 1, and is equal to

$$\tau_{-1} = \alpha_{-1}^R\tau^R + \left(1-\alpha_{-1}^R\right)\tau^I.$$

Impose the market clearing condition in equation (42), we can solve for the implicit price

$$p_{-1} = \tilde{p}_0 = \tilde{p}_1 = \mu_d + \left(\frac{1}{2} - \frac{1}{\tau_{-1}}\right)\sigma_d^2.$$

Note that at time $-1$, investors do not want to trade, because they believe that the risky asset has the same return as the risk-free asset.

## A1.11 Proof of Lemma 2

*Proof.* I first compute the $m$-th moment of $d_{j,t}^{in}$ in the cross section of retail investors, using the PDF specified in equation (58) with support $[d_{\min}, d_{\max}(N_t)]$.

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

The cross-sectional variance of $d_{j,t}^{in}$ is thus

$$\begin{aligned}
\mathrm{Var}^{CS}\left(d_{j,t}^{in}\right) &= \mathbb{E}\left[\left(d_{j,t}^{in}\right)^2\right] - \left(\mathbb{E}\left[d_{j,t}^{in}\right]\right)^2 \\
&= \frac{\xi-1}{3-\xi}\frac{1}{d_{\min}^{1-\xi}}\left(\left(d_{\max}(N_t)\right)^{3-\xi} - d_{\min}^{3-\xi}\right) - \left(\frac{\xi-1}{\xi-2}\right)^2 \frac{1}{d_{\min}^{2-2\xi}}\left(d_{\min}^{2-\xi} - \left(d_{\max}(N_t)\right)^{2-\xi}\right)^2.
\end{aligned}$$

□

## A1.12 Proof of Proposition 5

*Proof.* The proof follows from Acemoglu et al. (2012).

Using the PDF of $d_{j,t}^{in}$ in equation (58), I first derive the counter-CDF

$$P_{N_t}(x) \equiv \Pr\left(d_{j,t}^{in} > x\right) = \int_x^{+\infty} \frac{\xi-1}{d_{\min}}\left(\frac{y}{d_{\min}}\right)^{-\xi} dy = \left(\frac{x}{d_{\min}}\right)^{1-\xi} \tag{A11}$$

Define the empirical counterpart as

$$\hat{P}_{N_t}(x) = \frac{1}{N_t}\left|\left\{j \in \mathcal{I}_{N_t} : d_{j,t}^{in} > x\right\}\right| = \frac{1}{N_t}\sum_{j=1}^{N_t} \mathbf{1}\left\{d_{j,t}^{in} > x\right\}$$

Let $\boldsymbol{B}_t = \{b_{1,t}, b_{2,t}, \cdots, b_{m_t,t}\}$ denote the set of values $d_{j,t}^{in}$ takes, with $b_{1,t} < b_{2,t} < \cdots < b_{m_t,t}$, and the convention that $b_{0,t} = 0$. Then

$$\begin{aligned}
&\sum_{j=1}^{N_t} \left(d_{j,t}^{in}\right)^2 = N_t \sum_{k=1}^{m_t} \left(b_{k,t}\right)^2 \left(\hat{P}_{N_t}\left(b_{k-1,t}\right) - \hat{P}_{N_t}\left(b_{k,t}\right)\right) \\
= \;& N_t \left(b_{1,t}^2 \left(\hat{P}_{N_t}\left(b_{0,t}\right) - \hat{P}_{N_t}\left(b_{1,t}\right)\right) + \cdots + b_{m_t}^2 \left(\hat{P}_{N_t}\left(b_{m_t-1,t}\right) - \hat{P}_{N_t}\left(b_{m_t,t}\right)\right)\right) \\
= \;& N_t \left(\left(b_{1,t}^2 - b_{0,t}^2\right) \hat{P}_{N_t}\left(b_{0,t}\right) + \cdots + \left(b_{m_t,t}^2 - b_{m_t-1,t}^2\right) \hat{P}_{N_t}\left(b_{m_t-1,t}\right) - b_{m_t,t}^2 \hat{P}_{N_t}\left(b_{m_t,t}\right)\right) \\
= \;& N_t \sum_{k=0}^{m_t-1} \left(b_{k+1,t}^2 - b_{k,t}^2\right) \hat{P}_{N_t}\left(b_{k,t}\right) \\
= \;& N_t \sum_{k=0}^{m_t-1} \left(b_{k+1,t} + b_{k,t}\right)\left(b_{k+1,t} - b_{k,t}\right) \hat{P}_{N_t}\left(b_{k,t}\right) \\
= \;& 2N_t \sum_{k=0}^{m_t-1} \left(\frac{b_{k,t} + b_{k+1,t}}{2}\right)\left(b_{k+1,t} - b_{k,t}\right) \hat{P}_{N_t}\left(b_{k,t}\right)
\end{aligned}$$

Replace the empirical counter-CDF $\hat{P}_{N_t}\left(b_{k,t}\right)$ with the continuous function in (A11).

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

Using the dynamics of aggregate retail sentiment $\delta_t^R$ in equation (55), we can compute the conditional mean of $\delta_t^R$

$$\mathbb{E}_{t-1}\left[\delta_t^R\right] = \frac{\theta\left(N_t\right)}{\theta\left(N_{t-1}\right)} \rho \delta_{t-1}^R,$$

and the conditional variance

$$\begin{aligned}
\mathrm{Var}_{t-1}\left(\delta_t^R\right) &= \left(\theta\left(N_t\right)\right)^2 \frac{1}{N_t^2} \sum_{j=1}^{N_t} \left(d_{j,t}^{in}\right)^2 \sigma_\varepsilon^2 \\
&= \left(\theta\left(N_t\right)\right)^2 \frac{1}{N_t} \frac{2d_{\min}}{2-\xi} \sigma_\varepsilon^2 \left(\left(\frac{d_{\max}\left(N_t\right)}{d_{\min}}\right)^{2-\xi} \frac{2-\xi}{3-\xi} d_{\max}\left(N_t\right) - \frac{2-\xi}{3-\xi} d_{\min}\right) \\
&= \left(\theta\left(N_t\right)\right)^2 \frac{2d_{\min}}{N_t} \frac{1}{3-\xi} \left(\left(\frac{d_{\max}\left(N_t\right)}{d_{\min}}\right)^{2-\xi} d_{\max}\left(N_t\right) - d_{\min}\right) \sigma_\varepsilon^2 \\
&= \left(\theta\left(N_t\right)\right)^2 \frac{2d_{\min}^{\xi-1}}{N_t} \frac{1}{3-\xi} \left(\left(d_{\max}\left(N_t\right)\right)^{3-\xi} - d_{\min}^{3-\xi}\right) \sigma_\varepsilon^2 \\
&= O\left(N_t^{\frac{4-2\xi}{\xi-1}}\right).
\end{aligned}$$

The last equality uses $d_{\max}\left(N_t\right) = O\left(N_t^{\frac{1}{\xi-1}}\right)$. Hence, the conditional volatility of aggregate retail sentiment is

$$\sqrt{\mathrm{Var}_{t-1}\left(\delta_t^R\right)} = O\left(N_t^{\frac{2-\xi}{\xi-1}}\right).$$

□

## 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 $\sigma^2$ denote the pre-truncation variance of $\varepsilon_1^j$, then $X_j$ follows a truncated normal distribution on $[-\bar{\varepsilon}, \bar{\varepsilon}]$ with pre-truncation mean $\mu$ and variance $\sigma^2$, 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

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

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

The time-1 aggregate retail sentiment shock $\delta_1^R$ can be written as

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

Hence, the characteristic function of $\delta_1^R$ 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,

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

The characteristic function of $\delta_1^R$ is

$$
\begin{aligned}
&\varphi_{S_n}(t)=\exp\left(\left(\sum_{j=1}^{N}c_j\mu\right)it-\frac{1}{2}\left(\sum_{j=1}^{N}c_j^2\sigma^2\right)t^2\right)\prod_{j=1}^{N}\frac{\Phi(\rho-c_j\sigma it)+\Phi(\rho+c_j\sigma it)-1}{2\Phi(\rho)-1} \\
\implies\quad &\varphi_{\delta_1}(t)=\exp\left(\left(\sum_{j=1}^{N}c_j\right)\mu it-\frac{1}{2}\left(\sum_{j=1}^{N}c_j^2\right)\sigma_\varepsilon^2t^2\right)\prod_{j=1}^{N}\frac{\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}-c_j\sigma_\varepsilon it\right)+\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}+c_j\sigma_\varepsilon it\right)-1}{2\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}\right)-1} \\
\implies\quad &\varphi_{\delta_1}(t)=\exp\left(\mu it-\frac{1}{2}\left(\sum_{j=1}^{N}c_j^2\right)\sigma_\varepsilon^2t^2\right)\prod_{j=1}^{N}\frac{\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}-c_j\sigma_\varepsilon it\right)+\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}+c_j\sigma_\varepsilon it\right)-1}{2\Phi\left(\frac{\bar{\varepsilon}}{\sigma_\varepsilon}\right)-1}
\end{aligned}
$$

Compare the characteristic function of $\delta_1^R$ 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 $\sum_{j=1}^{N}c_j^2\sigma_\varepsilon^2$.

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

Hence, the distribution of $\delta_1^R$ 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 five 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 $\left\{d_{k,t}^{in}(b)\right\}_{k=1}^{N_t}$ 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 $\xi_t(b)$ and $d_{\min,t}(b)$ 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 $\{\xi_t(b)\}_{b=1}^{B}$, $\{d_{\min,t}(b)\}_{b=1}^{B}$.

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 $\{\xi_t(b)\}_{b=1}^{B}$. 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 $\{d_{\min,t}(b)\}_{b=1}^{B}$.

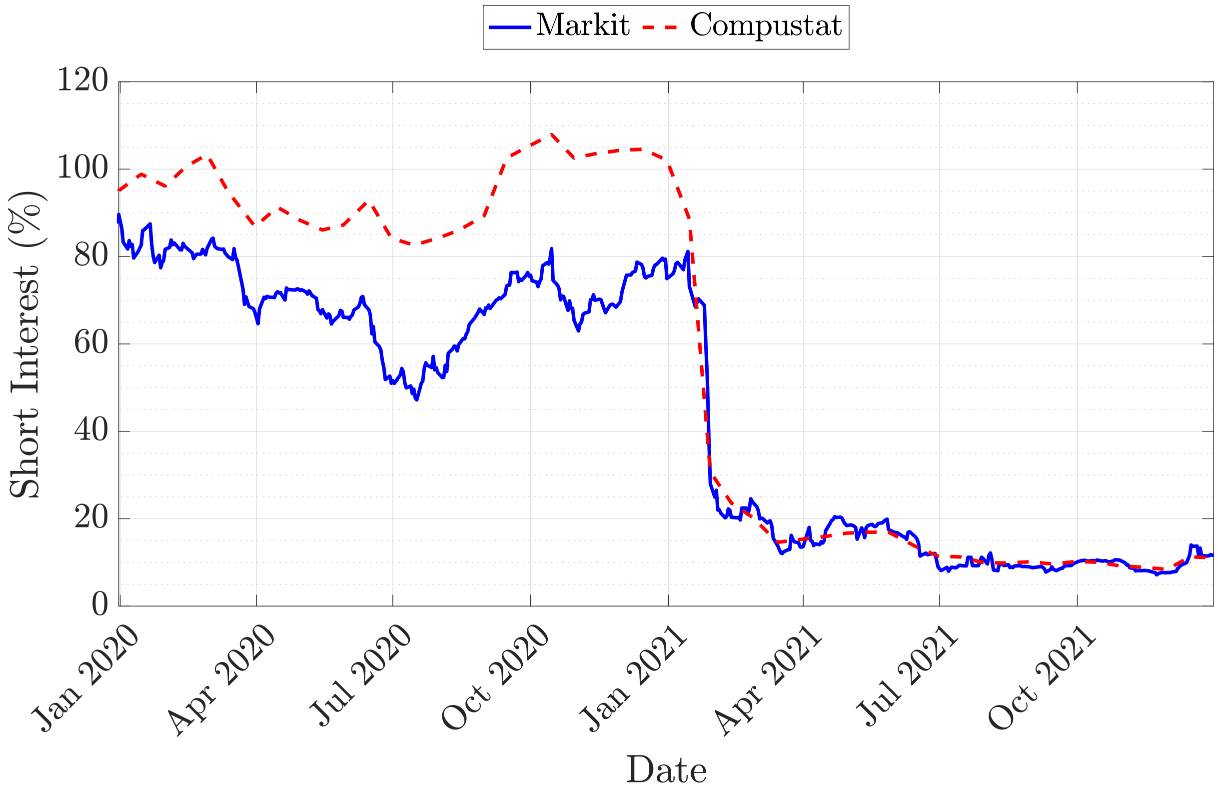



**Figure A1. Short interest of GameStop from IHS Markit versus Compustat.** This figure compares the short interest of GameStop computed using IHS Markit data versus that using Compustat data, for the period from January 1, 2020 to December 31, 2021. Short interest is defined as the ratio of the number of shares sold short to the number of shares outstanding (equation (6)). The solid blue line is the short interest computed using daily data on the number of shares sold short from IHS Markit. The dashed red line is the short interest computed using the mid-month and month-end number of shares sold short from Compustat.

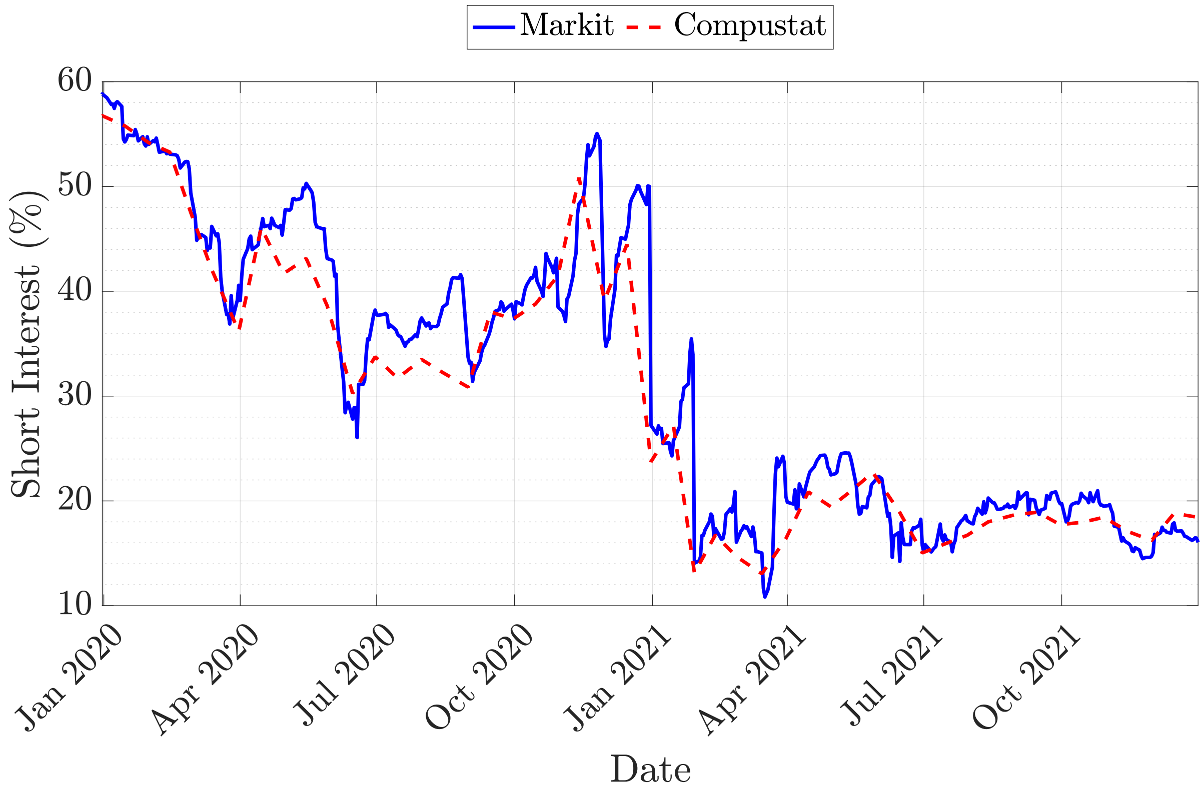



**Figure A2. Short interest of AMC from IHS Markit versus Compustat.** This figure compares the short interest of AMC computed using IHS Markit data versus that using Compustat data, for the period from January 1, 2020 to December 31, 2021. Short interest is defined as the ratio of the number of shares sold short to the number of shares outstanding (equation (6)). The solid blue line is the short interest computed using daily data on the number of shares sold short from IHS Markit. The dashed red line is the short interest computed using the mid-month and month-end number of shares sold short from Compustat.

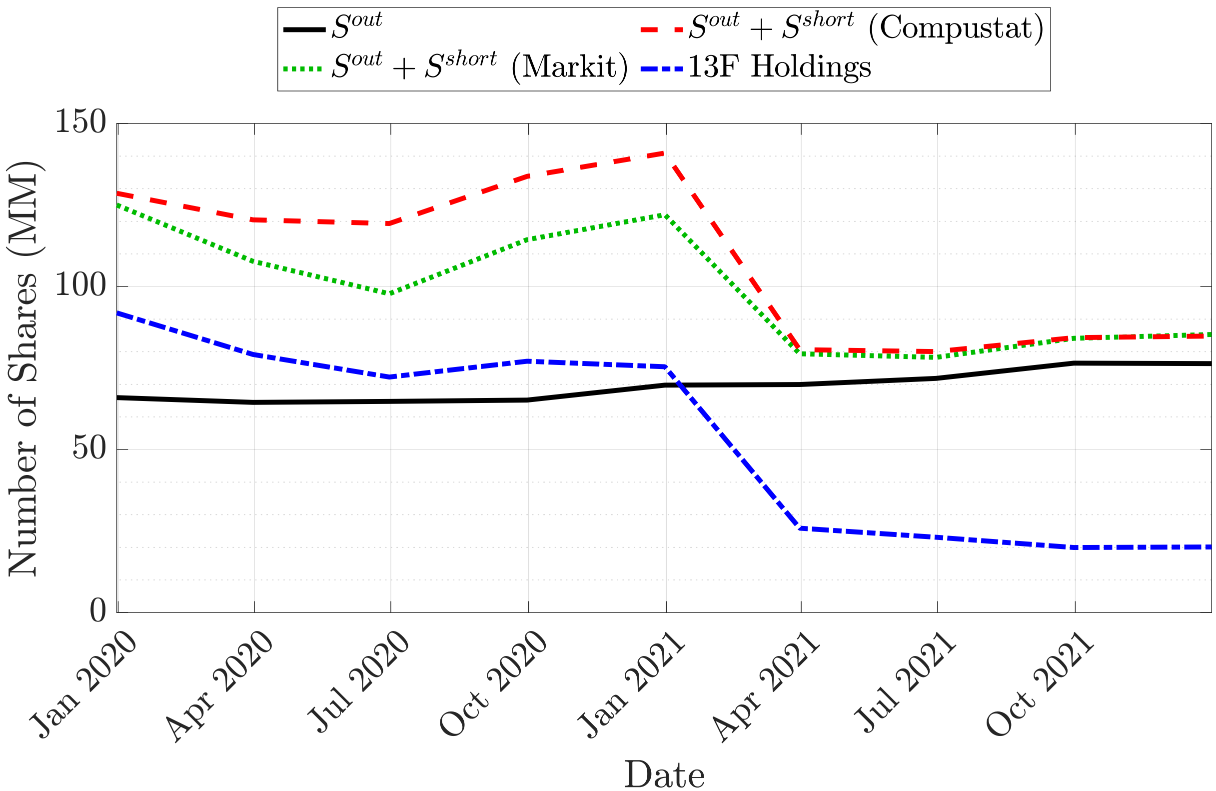



**Figure A3. Shares outstanding and 13F institutional ownership of GameStop.** This figure compares the number of shares outstanding of GameStop with 13F institutional ownership of GameStop, for the period from January 1, 2020 to December 31, 2021. The solid black line is the number of shares outstanding. The dashed red line is the number of shares outstanding plus the number of shares sold short from Compustat. The dotted green line is the number of shares outstanding plus the number of shares sold short from IHS Markit. The dash-dotted blue line is the number of shares held by 13F institutions.

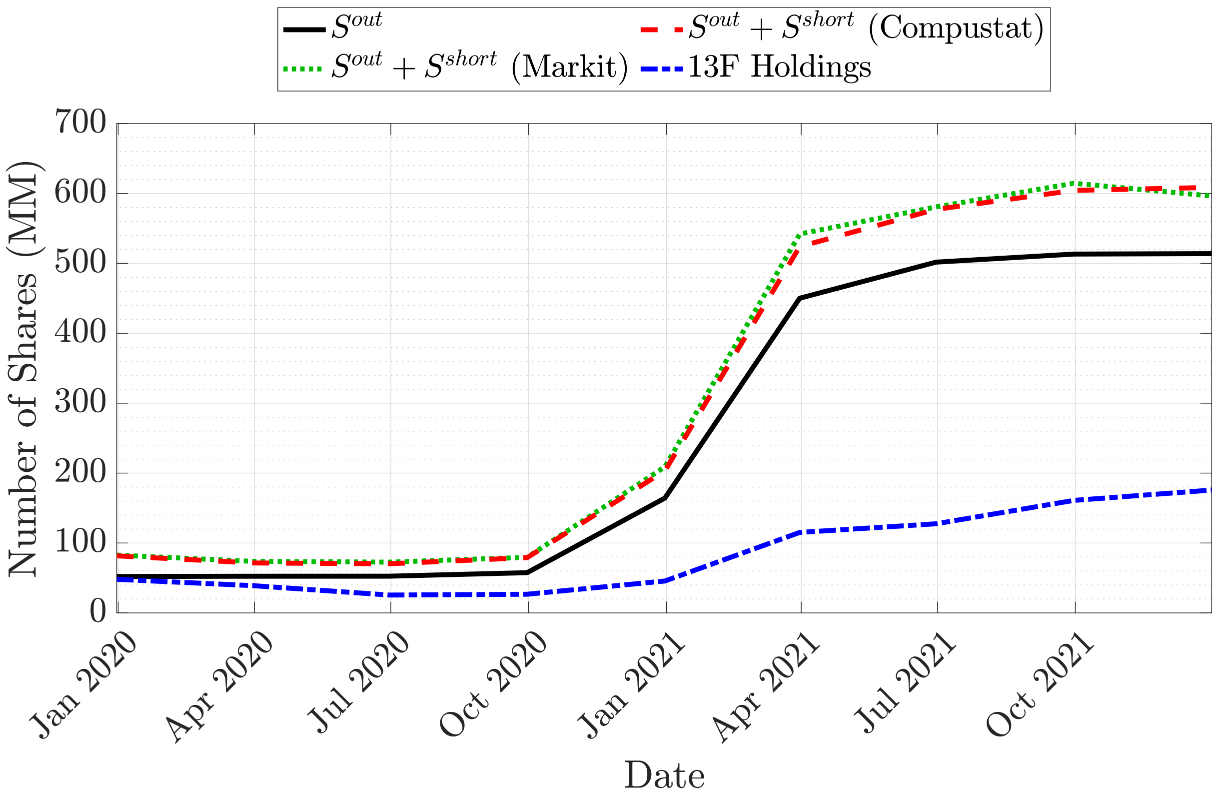



**Figure A4. Shares outstanding and 13F institutional ownership of AMC.** This figure compares the number of shares outstanding of AMC with 13F institutional ownership of AMC, for the period from January 1, 2020 to December 31, 2021. The solid black line is the number of shares outstanding. The dashed red line is the number of shares outstanding plus the number of shares sold short from Compustat. The dotted green line is the number of shares outstanding plus the number of shares sold short from IHS Markit. The dash-dotted blue line is the number of shares held by 13F institutions.

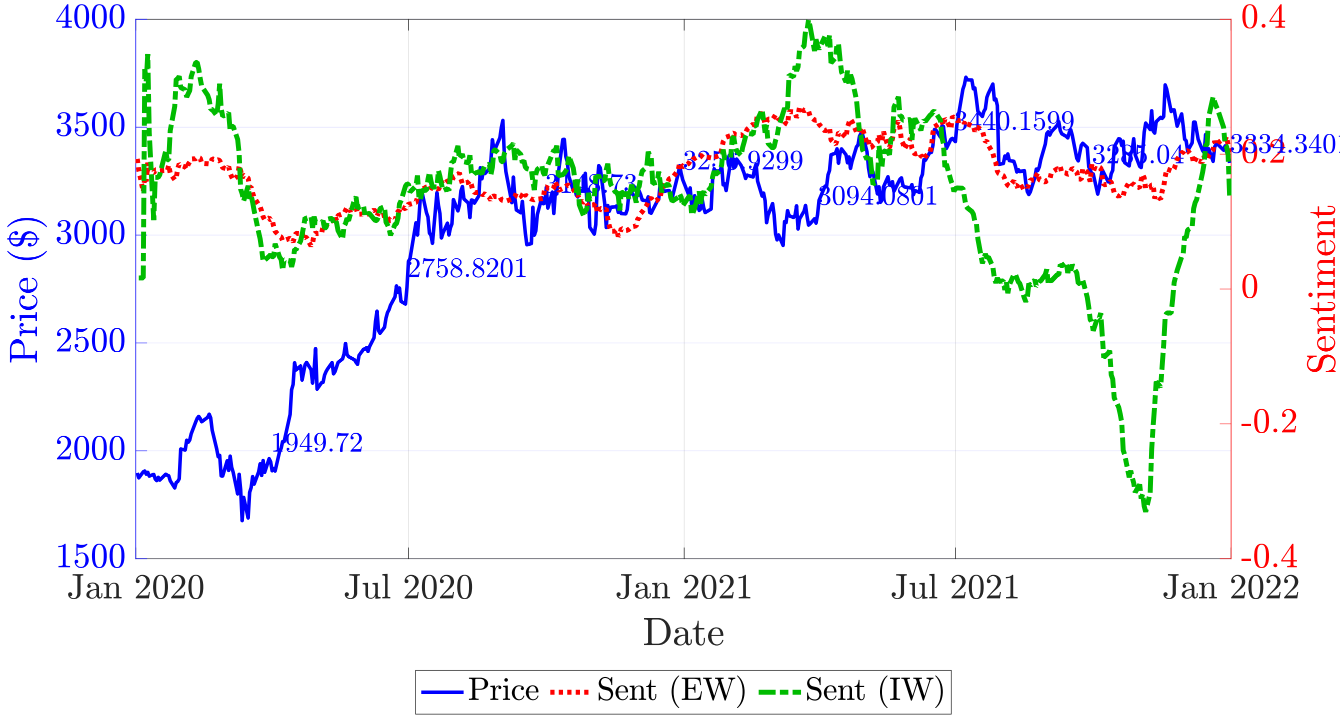



**Figure A5. Price and sentiment of Amazon.** This figure shows the daily close price (left $y$-axis) and the daily WSB sentiment measures (right $y$-axis) of Amazon, for the period from January 1, 2020 to December 31, 2021. The solid blue line plots the close price, the dotted red line plots the equal-weighted sentiment defined in equation (4), and the dash-dotted green line plots the influence-weighted sentiment defined in equation (5). The sentiment series are 30-day moving averages.

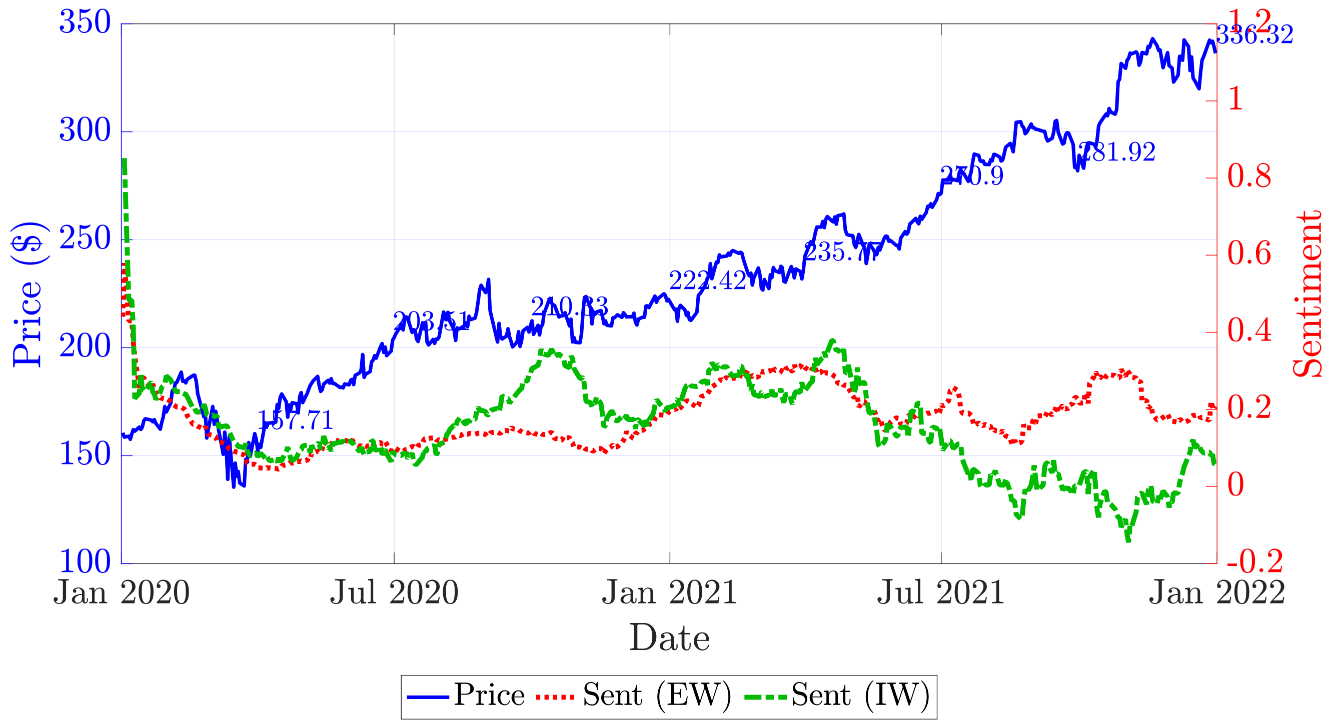



**Figure A6. Price and sentiment of Microsoft.** This figure shows the daily close price (left $y$-axis) and the daily WSB sentiment measures (right $y$-axis) of Microsoft, for the period from January 1, 2020 to December 31, 2021. The solid blue line plots the close price, the dotted red line plots the equal-weighted sentiment defined in equation (4), and the dash-dotted green line plots the influence-weighted sentiment defined in equation (5). The sentiment series are 30-day moving averages.


**Figure A7. Price and sentiment of AMC.** This figure shows the daily close price (left $y$-axis) and the daily WSB sentiment measures (right $y$-axis) of AMC, for the period from January 1, 2020 to December 31, 2021. The solid blue line plots the close price, the dotted red line plots the equal-weighted sentiment defined in equation (4), and the dash-dotted green line plots the influence-weighted sentiment defined in equation (5). The sentiment series are 30-day moving averages.

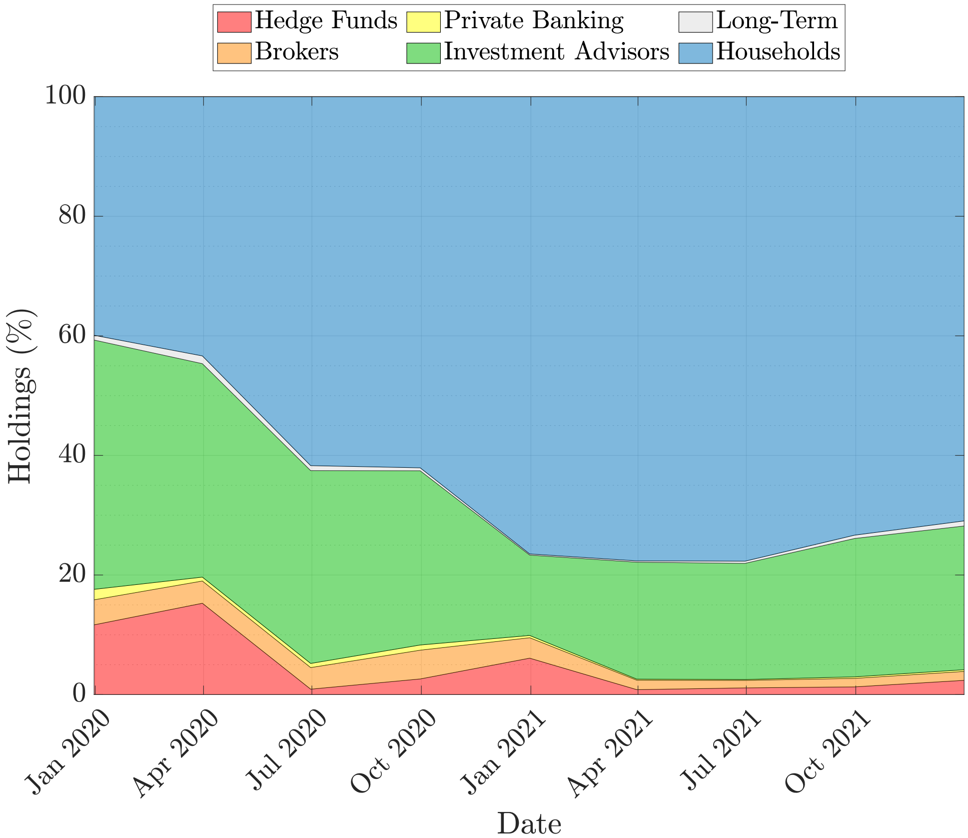
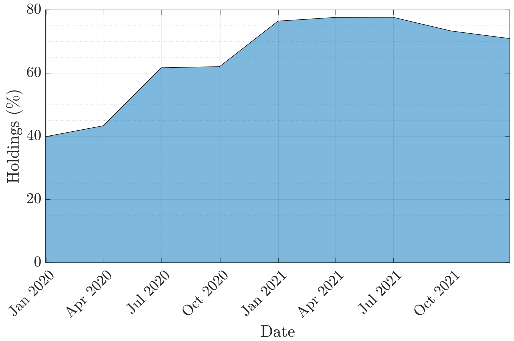
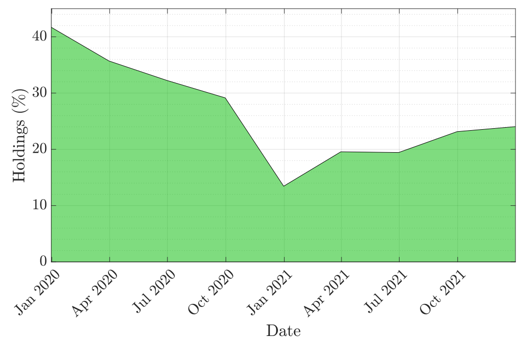


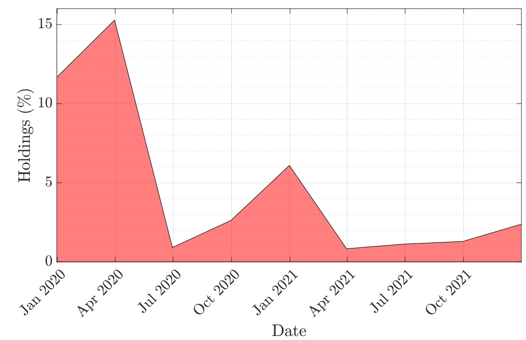

**Figure A8. Ownership of AMC by investor type.** This figure plots the end-of-quarter holdings of AMC by 13F institutions and households, for the period from 2019 Q4 to 2021 Q4. 13F holdings data are from FactSet. I aggregate 13F institutional holdings to investor-type level, using the method in Appendix A3. The five institutional investor types are: Hedge Funds (red area), Brokers (orange aread), Private Banking (yellow area), Investment Advisors (green area), and Long-Term Investors (gray area). I calculate household holdings from equation (8), using data on the number of shares sold short from Compustat. The blue area represents households. The $y$-axis is the percentage holdings defined in equation (10), which is the number of shares held by each type of investor divided by the sum of the number of shares outstanding and the number of shares sold short.



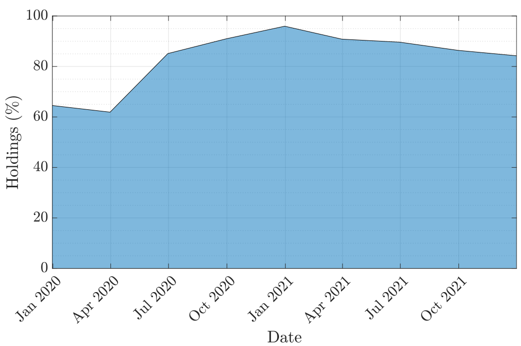

**(a)** Households / $(S^{out} + S^{short})$



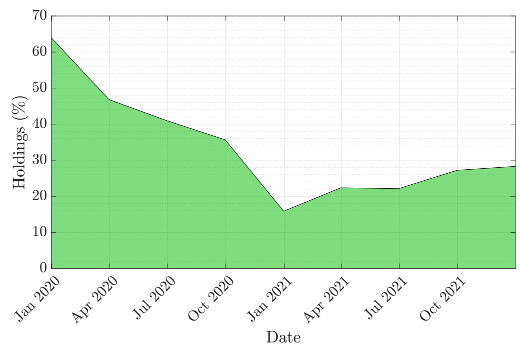

**(b)** Investment Advisors / $(S^{out} + S^{short})$



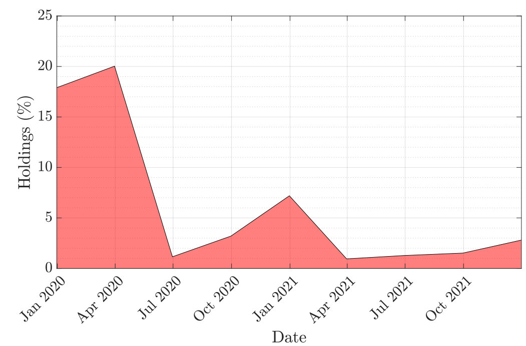

**(c)** Hedge Funds / $(S^{out} + S^{short})$




**(d)** Households / $S^{out}$




**(e)** Investment Advisors / $S^{out}$




**(f)** Hedge Funds / $S^{out}$

**Figure A9. Ownership of AMC by Households, Investment Advisors, and Hedge Funds.** This figure plots the end-of-quarter holdings of AMC by Households (panel (a) and (d)), Investment Advisors (panel (b) and (e)), and Hedge Funds (panel (c) and (f)), for the period from 2019 Q4 to 2021 Q4. 13F institutional investors are classified into Investment Advisors and Hedge Funds according to Appendix A3, and the 13F holdings data are from FactSet. Household holdings are calculated from equation (8). In panel (a), (b), and (c), the $y$-axis is the number of shares held by the investor group, divided by the sum of the number of shares outstanding and the number of shares sold short (equation (10)). Data on the number of shares sold short is from Compustat. In panel (d), (e), and (f), the $y$-axis is the number of shares held by the investor group, divided by the number of shares outstanding (equation (9)).

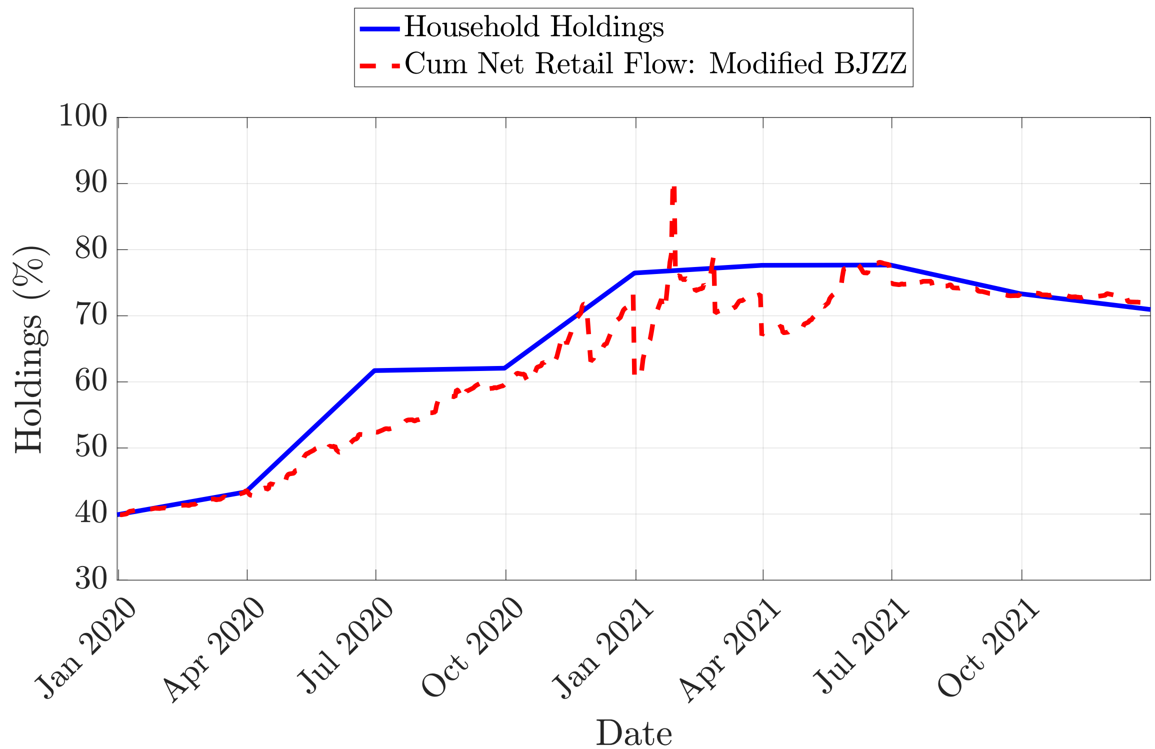



**Figure A10. Ownership by households versus cumulative net retail flow of AMC.** This figure plots the end-of-quarter percentage holdings of AMC by Households (solid blue line), and the daily cumulative net retail flow (dashed red line), for the period from January 1, 2020 to December 31, 2021. Percentage holdings by households is defined in equation (10), which is the number of shares held by households (equation (8)) divided by the sum of the number of shares outstanding and the number of shares sold short. Cumulative net retail flow is defined in equation (12), which is the cumulative net retail buy volume (equation (11)) divided by the sum of the number of shares outstanding and the number of shares sold short. Data on the number of shares sold short is from Compustat. The initial value of the cumulative net retail flow (on Dec 31, 2019) is set to be the percentage holdings by households at the end of 2019 Q4. I apply the modified BJZZ algorithm in Appendix A4 to identify retail trades from the TAQ data.


**Figure A11. Price and short interest of AMC.** This figure shows the daily close price (left $y$-axis) and the daily short interest (right $y$-axis) of AMC, for the period from January 1, 2020 to December 31, 2021. The solid blue line plots the close price. The dotted red line plots the short interest, which is defined as the ratio of the number of shares sold short to the number of shares outstanding (equation (6)). Data on the number of shares sold short is from IHS Markit.

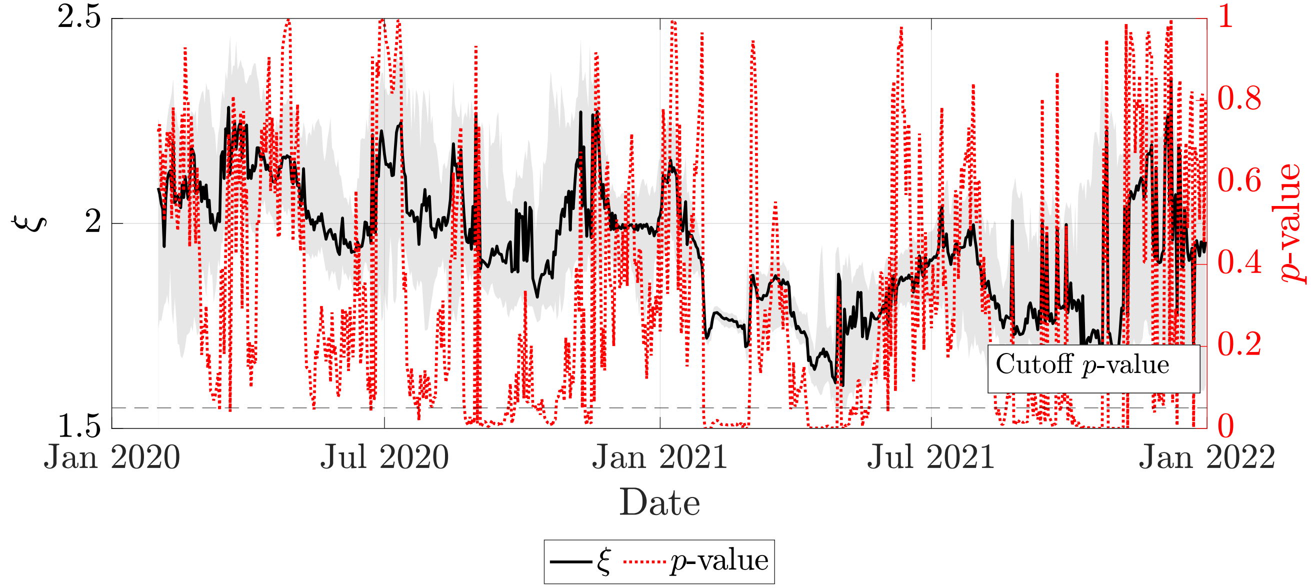



**Figure A12. $p$-value for fitting the power-law distribution.** This figure plots the daily estimate of the power-law exponent $\hat{\xi}_t$ (left $y$-axis) and the $p$-value (right $y$-axis) of the Kolmogorov-Smirnov test, for the period from January 1, 2020 to December 31, 2021. On each day $t$, I fit a power-law distribution to the vector of user influence (defined in equation (3)) and estimate the exponent $\xi$ in equation (13). The solid black line plots the $\hat{\xi}_t$ estimates from the maximum likelihood method as in Rantala (2019). The gray area shows the 95% confidence interval for the estimates, computed from the bootstrap method in Appendix A5. The dotted red line plots the $p$-value of the Kolmogorov-Smirnov test. The cutoff $p$-value is 0.05 (dashed horizontal line). Small $p$-values (less than 0.05) indicate that the test rejected the hypothesis that the original data could have been drawn from the fitted power-law distribution.

**Table A1**
**Reddit Bots Removed from the Sample**

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 |

**Table A2**
**Time-0 Equilibrium Outcomes under Different Risk Perceptions**

This table compares the time-0 equilibrium outcomes when changing investors' time-0 perceptions of risk. Column 3 shows the equilibrium outcomes when all investors believe that the size of the network at time 1 will remain the same as that at time 0, i.e., $\tilde{N}_1 = N_L = N_0$. Column 4 shows the equilibrium outcomes when all investors believe that the the size of the network will grow (deterministically) from time 0 to time 1, i.e., $\tilde{N}_1 = N_H > N_L = N_0$. The parameter values are given in Table 2.

| | | Value | |
|---|---|---|---|
| Description | Notation | $\tilde{N}_1 = N_L$ | $\tilde{N}_1 = N_H$ |
| (1) | (2) | (3) | (4) |
| Log price | $p_0$ | 4.249 | 4.612 |
| | $w_0^R$ | 1.900 | 1.024 |
| Portfolio weights | $w_0^{IL}$ | 1.759 | 1.288 |
| | $w_0^{IS}$ | $-0.250$ | 0.539 |
| | $Q_0^R$ | 60 | 34 |
| Num. shares held | $Q_0^{IL}$ | 50 | 52 |
| | $Q_0^{IS}$ | $-10$ | 14 |
| | $\alpha_0^R$ | 0.316 | 0.329 |
| Wealth shares | $\alpha_0^{IL}$ | 0.284 | 0.403 |
| | $\alpha_0^{IS}$ | 0.400 | 0.269 |
| Expected log payoff | $\mathbb{E}_0[p_1]$ | 4.469 | 5.157 |
| Variance of log return | $\sigma_0^2$ | 0.378 | 1.015 |