A New Measure of Herding Behavior on Financial Markets

Puriya Abbassi†, Christopher Boortz‡,a,b and Simon Jurkatis*,a,b

†Deutsche Bundesbank, ‡DIW Berlin, *BDPEMS, aFreie Universität Berlin, bCRC649

18th July 2015
Subgroups of traders have been suspected to move in and out of the same stocks at the same time, possibly destabilizing prices.
Subgroups of traders have been suspected to move in and out of the same stocks at the same time, possibly destabilizing prices

▶ in the 90’s mutual funds
Subgroups of traders have been suspected to move in and out of the same stocks at the same time, possibly destabilizing prices

▶ in the 90’s mutual funds

▶ 2000’s retail traders
Subgroups of traders have been suspected to move in and out of the same stocks at the same time, possibly destabilizing prices

- in the 90’s mutual funds
- 2000’s retail traders
- more recently, fire-sales of banks
We, therefore, ask: How to measure coordinated trading of subgroups that is potentially harmful?
We, therefore, ask: How to measure coordinated trading of subgroups that is potentially harmful?

Theoretical herding literature says: coordination is not bad *per se*
We, therefore, ask: How to measure coordinated trading of subgroups that is potentially harmful?

Theoretical herding literature says: coordination is not bad *per se*

- if 70% of a subgroup of traders receive (fundamentally driven) private signals that would induce them to, e.g., buy a stock, then \( \approx 70\% \) observed buys would be “good”
We, therefore, ask: How to measure coordinated trading of subgroups that is potentially harmful?

Theoretical herding literature says: coordination is not bad *per se*

- if 70% of a subgroup of traders receive (fundamentally driven) private signals that would induce them to, e.g., buy a stock, then $\approx 70\%$ observed buys would be “good”

- deviations of the buy-ratio from 70% $\Rightarrow$ not all of the traders follow their signals $\Rightarrow$ potentially bad
So far, the empirical literature . . .

- lead by the work of Lakonishok, Shleifer, Vishny (1992), *JFE*
So far, the empirical literature . . .

- lead by the work of **Lakonishok, Shleifer, Vishny (1992), JFE**
- while acknowledging the above argument
So far, the empirical literature ...

- lead by the work of Lakonishok, Shleifer, Vishny (1992), *JFE*
- while acknowledging the above argument
- does not attempt to control their measures for an *ex ante* coordination of traders due to their private signals (idiosyncratic for different stocks and times)
So far, the empirical literature ... 

- lead by the work of Lakonishok, Shleifer, Vishny (1992), JFE
- while acknowledging the above argument
- does not attempt to control their measures for an \textit{ex ante} coordination of traders due to their private signals (idiosyncratic for different stocks and times)
- instead, any deviation of observed buy-ratios, \(|br - \bar{br}|,(\bar{br} \approx 0.5)\) is interpreted as potentially destabilizing coordination
So far, the empirical literature . . .

- lead by the work of **Lakonishok, Shleifer, Vishny (1992), JFE**
- while acknowledging the above argument
- does not attempt to control their measures for an *ex ante* coordination of traders due to their private signals (idiosyncratic for different stocks and times)
- instead, any deviation of observed buy-ratios, \( |br - \bar{br}| \), \((\bar{br} \approx 0.5)\) is interpreted as potentially destabilizing coordination

We show that this frequently leads to erroneous conclusions
Literature Review

- **Lakonishok, Shleifer, Vishny (1992), JFE** - Introduce the LSV measure to test for coordinated trade behavior of pension fund money managers

- **Wermers (1999), JF** - Employs LSV measure more literally as a measure for herding of mutual funds

- **Dorn et al. (2008), JF** - Use LSV to detect correlated trading behavior of German retail investors

- **Barber et al. (2009), RFS** - Use LSV to confirm that individual investor trading behavior is coordinated and time persistent
other measures of coordinated trading of subgroups, e.g. **Sias (2004)**, **RFS** and **Sias and Choi (2009)**, **JFE**, build on the same idea as **LSV**
other measures of coordinated trading of subgroups, e.g. Sias (2004), RFS and Sias and Choi (2009), JFE, build on the same idea as LSV

early studies find relatively weak evidence of coordinated trading
other measures of coordinated trading of subgroups, e.g. Sias (2004), RFS and Sias and Choi (2009), JFE, build on the same idea as LSV

early studies find relatively weak evidence of coordinated trading

today: strong consensus that there exist non-negligible deviations from independent trading
other measures of coordinated trading of subgroups, e.g. Sias (2004), RFS and Sias and Choi (2009), JFE, build on the same idea as LSV.

early studies find relatively weak evidence of coordinated trading.

today: strong consensus that there exist non-negligible deviations from independent trading.

governing thought in the literature: measured coordination indicates (destabilizing) price impact.
▶ other measures of coordinated trading of subgroups, e.g. Sias (2004), RFS and Sias and Choi (2009), JFE, build on the same idea as LSV
▶ early studies find relatively weak evidence of coordinated trading
▶ today: strong consensus that there exist non-negligible deviations from independent trading
▶ governing thought in the literature: measured coordination indicates (destabilizing) price impact
▶ our view: this can only be true if the measure controls for information induced, independent trading
Outline

Toy-Model

The Measure

Estimation Procedure

Simulation Results

Comparison to the LSV measure

Empirical Application

Outlook
Investors trading stock $i$ receive one of 3 private signals:

- $S_1$: bad news - sell
- $S_2$: ambiguous news, negatively biased - sell
- $S_3$: good news - buy

$P(S) = [0.5 \ 0.2 \ 0.3]$ (Distribution of signals)
Investors trading stock $i$ receive one of 3 private signals:

- $S_1$: bad news - sell
- $S_2$: ambiguous news, negatively biased - sell
- $S_3$: good news - buy

$P(S) = [0.5 \ 0.2 \ 0.3]$ (Distribution of signals)

**Independent Trading**

- Trading decisions exclusively based on private signal
- No change in decision
- buy-ratio $br = 0.3$
Investors trading stock $i$ receive one of 3 private signals:

- $S_1$: bad news - sell
- $S_2$: ambiguous news, negatively biased - sell
- $S_3$: good news - buy

$P(S) = [0.5 \ 0.2 \ 0.3]$ (Distribution of signals)

**Independent Trading**

- Trading decisions exclusively based on private signal
- No change in decision
- buy-ratio $br = 0.3$

**Learning**

- Trading decisions based on private signal and history
- $S_2$ may change trading decision
- $br \rightarrow 0.5$
If we could measure the buy-probability under independent trading, $\tilde{br}$, coordinated trading should be measured by

$$h = |br - \tilde{br}|$$
If we could measure the buy-probability under independent trading, \( \tilde{br} \), coordinated trading should be measured by

\[
h = |br - \tilde{br}|
\]

**Independent Trading**

- \( \tilde{br} = 0.3 \)
- Trading decisions exclusively based on private signal
- No change in decision
- buy-ratio \( br = 0.3 \)
- \( h = 0 \)
If we could measure the buy-probability under independent trading, $\tilde{br}$, coordinated trading should be measured by

$$h = |br - \tilde{br}|$$

**Independent Trading**

- $\tilde{br} = 0.3$
- Trading decisions exclusively based on private signal
- No change in decision
- buy-ratio $br = 0.3$
- $h = 0$

**Learning**

- $\tilde{br} = 0.3$
- Trading decisions based on private signal and history
- $S_2$ may change trading decision
- $br \rightarrow 0.5$
- $h \rightarrow 0.2$
Let us generalize the idea
Let us generalize the idea

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
<td>$p_{11}$</td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
<td>$p_{21}$</td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
<td>$p_{31}$</td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td></td>
<td>$\tilde{br}_1$</td>
</tr>
</tbody>
</table>
Let us generalize the idea

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
</tr>
<tr>
<td>$	ilde{br}$</td>
<td>$\tilde{br}_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
<td>$p_{11}$</td>
<td>$p_{12}$</td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
<td>$p_{21}$</td>
<td>$p_{22}$</td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
<td>$p_{31}$</td>
<td>$p_{32}$</td>
</tr>
<tr>
<td>$	ilde{br}$</td>
<td>$\tilde{br}_1$</td>
<td>$\tilde{br}_2$</td>
<td>$\tilde{br}_3$</td>
</tr>
</tbody>
</table>
Let us generalize the idea

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
<td>$p_{11}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
<td>$p_{21}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
<td>$p_{31}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td>$\tilde{br}_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
<td>$p_{11}$</td>
<td>$p_{12}$</td>
<td>$p_{13}$</td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
<td>$p_{21}$</td>
<td>$p_{22}$</td>
<td>$p_{23}$</td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
<td>$p_{31}$</td>
<td>$p_{32}$</td>
<td>$p_{33}$</td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td>$\tilde{br}_1$</td>
<td>$\tilde{br}_2$</td>
<td>$\tilde{br}_3$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>...</th>
<th>$V_∞$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller</td>
<td>$S_1$</td>
<td>$p_{11}$</td>
<td>$p_{12}$</td>
<td>$p_{1∞}$</td>
</tr>
<tr>
<td>Seller</td>
<td>$S_2$</td>
<td>$p_{21}$</td>
<td>$p_{22}$</td>
<td>$p_{2∞}$</td>
</tr>
<tr>
<td>Buyer</td>
<td>$S_3$</td>
<td>$p_{31}$</td>
<td>$p_{32}$</td>
<td>$p_{3∞}$</td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td>$\tilde{br}_1$</td>
<td>$\tilde{br}_2$</td>
<td>...</td>
<td>$\tilde{br}_∞$</td>
</tr>
</tbody>
</table>
Let us generalize the idea

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller $S_1$</td>
<td>$p_{11}$</td>
</tr>
<tr>
<td>Seller $S_2$</td>
<td>$p_{21}$</td>
</tr>
<tr>
<td>Buyer $S_3$</td>
<td>$p_{31}$</td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td>$\tilde{br}_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller $S_1$</td>
<td>$p_{11}$</td>
<td>$p_{12}$</td>
<td>$p_{13}$</td>
<td></td>
</tr>
<tr>
<td>Seller $S_2$</td>
<td>$p_{21}$</td>
<td>$p_{22}$</td>
<td>$p_{23}$</td>
<td>...</td>
</tr>
<tr>
<td>Buyer $S_3$</td>
<td>$p_{31}$</td>
<td>$p_{32}$</td>
<td>$p_{33}$</td>
<td></td>
</tr>
<tr>
<td>$\tilde{br}$</td>
<td>$\tilde{br}_1$</td>
<td>$\tilde{br}_2$</td>
<td>$\tilde{br}_3$</td>
<td></td>
</tr>
</tbody>
</table>

$\forall V \in [0, 1]$

$\forall p = f_S(V)$ (continuous signal)

$\forall \tilde{br} \in (0, 1)$

$\tilde{br}$ different for different realizations of $V$

$\Rightarrow \tilde{br} \sim \text{Beta}(\alpha, \beta)$
we are not able to observe or estimate $\tilde{br}$...
we are not able to observe or estimate $\tilde{b}r$ . . .

but maybe $\text{Beta}(\alpha, \beta)$

Does this help us to assess the deviation from independent trading, $h = |br - \tilde{b}r|$?
we are not able to observe or estimate \( \tilde{br} \ldots \)

but maybe \( \text{Beta}(\alpha, \beta) \)

Does this help us to assess the deviation from independent trading, \( h = |br - \tilde{br}| \) ?

Yes, we can form an expectation over the deviation by
we are not able to observe or estimate $\tilde{br}$ . . .

but maybe $\text{Beta}(\alpha, \beta)$

Does this help us to assess the deviation from independent trading, $h = |br - \tilde{br}|$? Yes, we can form an expectation over the deviation by

$$H = \mathbb{E}[|br - p|] - AF = \int_{0}^{1} f(p|\alpha, \beta)|br - p|dp - AF \quad (1)$$

- $f(\cdot)$: Beta distribution
- $AF$: adjustment factor to center $H$ on zero under “random deviation”
\[ H = \int_0^1 f(p|\alpha, \beta) |br - p| dp - AF \]
Introduction

Toy-Model

The Measure

Estimation Procedure

Simulation Results

Comparison to the LSV measure

Empirical Application

Outlook

\[ H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - AF \]

given \( br \), in expectation 20 out of 100 trades were more buys or sales than under independent trading.
\[ H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - AF \]

Given \( br \), in expectation 20 out of 100 trades were more buys or sales than under independent trading.

\[ AF = \]
\[ H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - AF \]

\[ \sum_{k} \binom{T}{k} \bar{p}^k \bar{p}^{T-k} \left| \frac{k}{T} - p \right| \]

- given \( br \), in expectation 20 out of 100 trades were more buys or sales than under independent trading.

- adjusting for variation in \( br \) due to finite number of draws.
\[ H = \int_0^1 f(p|\alpha, \beta) |br - p|dp - 0.2 \]

- given \( br \), in expectation 20 out of 100 trades were more buys or sales than under independent trading

\[ AF = \int_0^1 f(\tilde{p}|\alpha, \beta) \sum_k^T \binom{T}{k} \tilde{p}^k \tilde{p}^{T-k} \frac{k}{T} - p|d\tilde{p} \]

- adjusting for variation in \( br \) due to finite number of draws
- adjusting for variation in \( br \) due to being drawn from a beta distribution
\[
H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - \underbrace{AF}_{0.2}
\]

- given \(br\), in expectation 20 out of 100 trades were more buys or sales than under independent trading

\[
AF = \int_0^1 f(p|\cdot) \int_0^1 f(\tilde{p}|\alpha, \beta) \sum_k \binom{T}{k} \tilde{p}^k \tilde{p}^{T-k} \left[ \frac{k}{T} - p \right] |d\tilde{p}dp
\]

- adjusting for variation in \(br\) due to finite number of draws
- adjusting for variation in \(br\) due to being drawn from a beta distribution
\[ H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - AF \]

given \( br \), in expectation 20 out of 100 trades were more buys or sales than under independent trading

\[ AF = \int_0^1 f(p|\cdot) \int_0^1 f(\tilde{p}|\alpha, \beta) \sum_k \binom{T}{k} \tilde{p}^k \tilde{p}^{T-k} \left| \frac{k}{T} - p \right| d\tilde{p} dp \]

adjusting for variation in \( br \) due to finite number of draws

adjusting for variation in \( br \) due to being drawn from a beta distribution
Introduction

Toy-Model

The Measure

Estimation Procedure

Simulation Results

Comparison to the LSV measure

Empirical Application

Outlook

\[
H = \int_{0}^{1} f(p|\alpha, \beta)|br - p| dp - AF
\]

\[
AF = \int_{0}^{1} f(p|\cdot) \int_{0}^{1} f(\tilde{p}|\alpha, \beta) \sum_{k} \binom{T}{k} \tilde{p}^{k} (1 - \tilde{p})^{T-k} \frac{k}{T} - p| dp\tilde{p} dp
\]

\[
= \int_{0}^{1} f(p|\cdot) \sum_{k} Beta-Bino(k| T, \alpha, \beta)| \frac{k}{T} - p| dp
\]

- given \(br\), in expectation 20 out of 100 trades were more buys or sales than under independent trading

- adjusting for variation in \(br\) due to finite number of draws

- adjusting for variation in \(br\) due to being drawn from a beta distribution

- we would expect 10 out of 100 trades to be more buys or sales than under independent trading simply by chance
\[ H = \int_0^1 f(p|\alpha, \beta)|br - p|dp - \text{AF} = 0.1 \]
\[ H = \int_{0}^{1} f(p|\alpha, \beta)|br - p|dp - \text{AF} = 0.1 \]

- given \( br \), we expect 10 out 100 trades to be more buys or sales than we would have expected under random, independent trading.
\[ H = \int_{0}^{1} f(p|\alpha, \beta)|br - p|dp - AF = 0.1 \]

- \( \bar{H} < 0 \rightarrow \) “underdispersion”

- given \( br \), we expect 10 out of 100 trades to be more buys or sales than we would have expected under random, independent trading.
\[
H = \int_0^1 f(p|\alpha, \beta) |br - p| dp - \cancel{AF} = 0.1
\]

- given \(br\), we expect 10 out 100 trades to be more buys or sales than we would have expected under random, independent trading

- \(\overline{H} < 0 \rightarrow \) “underdispersion”
- \(\overline{H} > 0 \rightarrow \) “overdispersion” or
\[ H = \int_0^1 f(p|\alpha,\beta) |br - p| dp - AF = 0.1 \]

- given \(br\), we expect 10 out of 100 trades to be more buys or sales than we would have expected under random, independent trading.

- \(\bar{H} < 0 \rightarrow \) “underdispersion”
- \(\bar{H} > 0 \rightarrow \) “overdispersion” or
- shift in the mean of the distribution
How to estimate $\text{Beta}(\alpha, \beta)$?
How to estimate Beta(\(\alpha, \beta\))?

- consider \(X^\tau\) to be the number of buys up to the \(\tau\)-th trade, \(A_\tau \in \{\text{Buy, Sell}\}\)
How to estimate Beta(\(\alpha, \beta\))?

- consider \(X^\tau\) to be the number of buys up to the \(\tau\)-th trade, \(A_\tau \in \{\text{Buy}, \text{Sell}\}\)
- market microstructure theory tells us, \(\exists \tau : 1 \leq \tau \leq T\) and all \(A_j\)'s are independent trades for \(j = 1, \ldots, \tau\)
How to estimate Beta(\(\alpha, \beta\))? 

- consider \(X^\tau\) to be the number of buys up to the \(\tau\)-th trade, \(A_\tau \in \{\text{Buy}, \text{Sell}\}\)
- market microstructure theory tells us, \(\exists \tau: 1 \leq \tau \leq T\) and all \(A_j\)'s are independent trades for \(j = 1, \ldots, \tau\)
- then, \(X^\tau | \tilde{b}_r \sim \text{Bino}(\tau, \tilde{b}_r)\) and \(X^\tau \sim \text{Beta-Bino}(\tau, \alpha, \beta)\)
How to estimate Beta(α, β)?

- consider $X^\tau$ to be the number of buys up to the $\tau$-th trade, $A \tau \in \{\text{Buy, Sell}\}$
- market microstructure theory tells us, \( \exists \tau : 1 \leq \tau \leq T \) and all $A_j$'s are independent trades for $j = 1, \ldots, \tau$
- then, $X^\tau | \tilde{\beta} r \sim \text{Bino}(\tau, \tilde{\beta} r)$ and $X^\tau \sim \text{Beta-Bino}(\tau, \alpha, \beta)$
- $\hat{\alpha}, \hat{\beta} = \arg\max_{\alpha, \beta} L(X^\tau | \tau, \alpha, \beta)$, with $L(\cdot)$ the log-likelihood function of the data, $\{X^\tau\}$, assuming the beta-binomial distribution
How to estimate $\text{Beta}(\alpha, \beta)$?

- consider $X^\tau$ to be the number of buys up to the $\tau$-th trade, $A_\tau \in \{\text{Buy}, \text{Sell}\}$
- market microstructure theory tells us, $\exists \tau: 1 \leq \tau \leq T$ and all $A_j$’s are independent trades for $j = 1, \ldots, \tau$
- then, $X^\tau|_{\tilde{b}_r} \sim \text{Bino}(\tau, \tilde{b}_r)$ and $X^\tau \sim \text{Beta-Bino}(\tau, \alpha, \beta)$
- $\hat{\alpha}, \hat{\beta} = \arg\max_{\alpha, \beta} L(X^\tau|\tau, \alpha, \beta)$, with $L(\cdot)$ the log-likelihood function of the data, $\{X^\tau\}$, assuming the beta-binomial distribution
- how to chose $\tau$? small but not too small, e.g. intra-day setting: $\tau = 10$ trades (ad hoc; refinement for future work)
PROPOSITION

Let \( \{X^{\tau_i}\}_i \) be iid with \( X^{\tau_i} \sim \text{Beta-Bino}(\tau_i, \alpha, \beta) \) and \( 1 \leq \tau_i \leq T_i \), and \( \{b_i\}_i \) with \( b_i \in [0, 1] \). Define

\[
\hat{H}_i := \int_0^1 f(p|\hat{\alpha}, \hat{\beta}) b_i - p \, dp - \hat{AF}_i, \\
\text{with } \hat{AF}_i := \int_0^1 f(p|\hat{\alpha}, \hat{\beta}) \sum_{k} \text{Beta-Bino}(k|T_i, \hat{\alpha}, \hat{\beta}) \frac{k}{T_i} - p \, dp
\]

and \( \hat{\alpha}, \hat{\beta} = \arg \max_{\alpha, \beta} L(X^{\tau_i} | \tau_i, \alpha, \beta) \). Then

\[
\hat{H}_i \xrightarrow{p} H_i, \text{ for } I \to \infty.
\]
We test convergence and finite sample performance in simulations

- **Procedure:**
  1. set $\alpha, \beta$
We test convergence and finite sample performance in simulations

- Procedure:
  1. set $\alpha, \beta$
  2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
We test convergence and finite sample performance in simulations

- **Procedure:**
  1. set $\alpha, \beta$
  2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
  3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{br}_i$)
We test convergence and finite sample performance in simulations

Procedure:

1. set $\alpha, \beta$
2. draw $\{\tilde{b}_r_i\}$ from Beta($\alpha, \beta$)
3. draw $\{X^\tau_i\}$ from Bino($\tau_i, \tilde{b}_r_i$)
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
We test convergence and finite sample performance in simulations

- Procedure:
  1. set $\alpha, \beta$
  2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
  3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{br}_i$)
  4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
  5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
We test convergence and finite sample performance in simulations

Procedure:

1. set $\alpha, \beta$
2. draw $\{\tilde{br}_i\}$ from $\text{Beta}(\alpha, \beta)$
3. draw $\{X^{\tau_i}\}$ from $\text{Bino}(\tau_i, \tilde{br}_i)$
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
6. compute $H_i$
We test convergence and finite sample performance in simulations

- **Procedure:**
  1. set $\alpha, \beta$
  2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
  3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{br}_i$)
  4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
  5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
  6. compute $H_i$
  7. estimate $\alpha, \beta$ on $\{X^{\tau_i}\}$ by ML and compute $\hat{H}_i$
We test convergence and finite sample performance in simulations

Procedure:

1. set $\alpha, \beta$
2. draw $\{\tilde{b}_r_i\}$ from Beta($\alpha, \beta$)
3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{b}_r_i$)
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
5. compute $b_r_i = \tilde{b}_r_i + h_i$ for all $i$ (if necessary truncate to zero or one)
6. compute $H_i$
7. estimate $\alpha, \beta$ on $\{X^{\tau_i}\}$ by ML and compute $\hat{H}_i$
8. compute $H_i - \hat{H}_i$ for all $i$ and RMSE($H, \hat{H}$)
We test convergence and finite sample performance in simulations

 Procedure:

1. set $\alpha, \beta$
2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{br}_i$)
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
6. compute $H_i$
7. estimate $\alpha, \beta$ on $\{X^{\tau_i}\}$ by ML and compute $\hat{H}_i$
8. compute $H_i - \hat{H}_i$ for all $i$ and RMSE($H, \hat{H}$)

$\tau_i = 10 \forall i$, $T_i = T \sim U[50, 1000]$
We test convergence and finite sample performance in simulations

Procedure:

1. set $\alpha, \beta$
2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{br}_i$)
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
6. compute $H_i$
7. estimate $\alpha, \beta$ on $\{X^{\tau_i}\}$ by ML and compute $\hat{H}_i$
8. compute $H_i - \hat{H}_i$ for all $i$ and RMSE($H, \hat{H}$)

$\tau_i = 10 \forall i$, $T_i = T \sim U[50, 1000]$

different setups for choosing $\alpha, \beta$ and $h_i$:

- different shapes of beta distribution
- different degrees of distortion in $\tilde{br}_i$
We test convergence and finite sample performance in simulations

Procedure:

1. set $a, b$
2. draw $\{\tilde{b}r_i\}$ from Beta($a, b$)
3. draw $\{X^{\tau_i}\}$ from Bino($\tau_i, \tilde{b}r_i$)
4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
5. compute $br_i = \tilde{b}r_i + h_i$ for all $i$ (if necessary truncate to zero or one)
6. compute $H_i$
7. estimate $a, b$ on $\{X^{\tau_i}\}$ by ML and compute $\hat{H}_i$
8. compute $H_i - \hat{H}_i$ for all $i$ and RMSE($H, \hat{H}$)

$\tau_i = 10 \forall i$, $T_i = T \sim U[50, 1000]$

different setups for choosing $a, b$ and $h_i$:

- different shapes of beta distribution
- different degrees of distortion in $\tilde{b}r_i$

each setup repeated 50 times for $I = 75, 250, 500, 1000, \ldots, 16000$
We test convergence and finite sample performance in simulations

- **Procedure:**
  1. set $\alpha, \beta$
  2. draw $\{\tilde{br}_i\}$ from Beta($\alpha, \beta$)
  3. draw $\{X_{\tau_i}\}$ from Bino($\tau_i, br_i$)
  4. set distortion in independent buy ratio $h_i$ and number of trades $T_i$ for all $i$
  5. compute $br_i = \tilde{br}_i + h_i$ for all $i$ (if necessary truncate to zero or one)
  6. compute $H_i$
  7. estimate $\alpha, \beta$ on $\{X_{\tau_i}\}$ by ML and compute $\hat{H}_i$
  8. compute $H_i - \hat{H}_i$ for all $i$ and RMSE($H, \hat{H}$)

- $\tau_i = 10 \forall i, T_i = T \sim U[50, 1000]$
- different setups for choosing $\alpha, \beta$ and $h_i$:
  - different shapes of beta distribution
  - different degrees of distortion in $\tilde{br}_i$
- each setup repeated 50 times for $I = 75, 250, 500, 1000, \ldots, 16000$

(we also implemented a market microstructure model where $\tau_i$ and $h_i$ arise endogenously and cannot be observed)
Figure: $H_i - \hat{H}_i$ based on $50 \times l$ observations for $l \in 75, 250, 16000$ (left, middle, right)
Figure: RMSE\((H, \hat{H}) \times 100\) based on \(50 \times I\) observations for \(I = 75, 250, 500, 1000, \ldots, 16000\)
How does our measure compare to LSV?
How does our measure compare to \( LSV \)?

\[
LSV_i = |br_i - \bar{br}| - AF_{iLSV},
\]

(3)

with \( AF_{iLSV} = \sum_{k} \left( \begin{array}{c} T_i \\ k \end{array} \right) \bar{br}_t^k (1 - \bar{br})^{T_i-k} \left| \frac{k}{T_i} - \bar{br} \right| \)

(4)
How does our measure compare to \textit{LSV}?

\[
LSV_i = |br_i - \bar{br}| - AF_i^{LSV},
\]

with \[
AF_i^{LSV} = \sum_k \binom{T_i}{k} \bar{br}^k (1 - \bar{br})^{T_i-k} |\frac{k}{T_i} - \bar{br}|
\]

- interpretation is similar to our measure, but ...
How does our measure compare to LSV?

\[ LSV_i = |br_i - \bar{br}| - AF_{i}^{LSV}, \]

with \[ AF_{i}^{LSV} = \sum_{k}^{T_i} \left( \begin{array}{c} T_i \\ k \end{array} \right) \bar{br}_k (1 - \bar{br})^{T_i-k} \left| \frac{k}{T_i} - \bar{br} \right| \]

- interpretation is similar to our measure, but ...
- assumes implicitly that \( X_{T_i}^{H_0} \sim \text{Bino}(T_i, p) \) for all \( i \)
How does our measure compare to LSV?

\[
LSV_i = |br_i - \bar{br}| - AF_i^{LSV},
\]

with \( AF_i^{LSV} = \sum_{k} \binom{T_i}{k} \bar{br}_t^k (1 - \bar{br})^{T_i-k} \frac{k}{T_i} - \bar{br} |T_i - \bar{br}| \)

- interpretation is similar to our measure, but ... 
- assumes implicitly that \( X^{T_i} \overset{H_0}{\sim} \text{Bino}(T_i, p) \) for all \( i \) 
- that is, LSV assumes that \( p \) is the same for all stocks under independent trading
How does our measure compare to \textit{LSV}?

\[ LSV_i = |br_i - \bar{br}| - AF_i^{LSV} \] \hspace{1cm} (3)

with

\[ AF_i^{LSV} = \sum_k \left( \begin{array}{c} T_i \\ k \end{array} \right) \bar{br}_i^k (1 - \bar{br})^{T_i - k} \left| \frac{k}{T_i} - \bar{br} \right| \] \hspace{1cm} (4)

- interpretation is similar to our measure, but ...
- assumes implicitly that \( X^{T_i} \overset{H_0}{\sim} \text{Bino}(T_i, p) \) for all \( i \)
- that is, LSV assumes that \( p \) is the same for all stocks under independent trading
- \( p \) is estimated by ML on \( \{X^{T_i}\} \), given by \( \bar{br} = \sum X^{T_i} / \sum T_i \)
How does our measure compare to LSV?

\[
LSV_i = |br_i - \bar{br}| - AF_{LSV}^i, \tag{3}
\]

with \( AF_{LSV}^i = \sum_{k} \left( \begin{array}{c} T_i \\ k \end{array} \right) \bar{br}_t^k (1 - \bar{br})^{T_i-k} | \frac{k}{T_i} - \bar{br}| \tag{4} \]

- interpretation is similar to our measure, but...
- assumes implicitly that \( X^{T_i} \sim H_0 \text{ Bino}(T_i, p) \) for all \( i \)
- that is, LSV assumes that \( p \) is the same for all stocks under independent trading
- \( p \) is estimated by ML on \( \{X^{T_i}\} \), given by \( \bar{br} = \sum X^{T_i} / \sum T_i \)
- \( X^{T_i} \) may already be affected by dependent trading
Our measure includes $LSV$ as a special case
Our measure includes *LSV* as a special case

**PROPOSITION**

Let \( \{X_{\tau i}\} \) and \( \{X_{T i}\} \) be iid with \( X_{\tau i} \sim \text{Beta-Bino}(\tau_i, \alpha, \beta) \) and \( X_{T i} = X_{o T i} + \varepsilon_i \) where \( X_{o T i} \sim \text{Beta-Bino}(T_i, \alpha, \beta) \). If \( \varepsilon_i = 0 \ \forall \ i \ \vee \sum_i \varepsilon_i = 0 \), then

\[
\hat{H}_i \xrightarrow{I} LSV_i, \quad \text{with } \alpha, \beta \to \infty
\]
Our measure includes $LSV$ as a special case

**PROPOSITION**

Let $\{X^{\tau_i}\}$ and $\{X^{T_i}\}$ be iid with $X^{\tau_i} \sim \text{Beta-Bino}(\tau_i, \alpha, \beta)$ and $X^{T_i} = X^{T_i}_o + \varepsilon_i$ where $X^{T_i}_o \sim \text{Beta-Bino}(T_i, \alpha, \beta)$. If $\varepsilon_i = 0 \forall i \lor \sum_i \varepsilon_i = 0$, then

$$\hat{H}_i \xrightarrow{I} \infty \text{ LSV}_i, \text{ with } \alpha, \beta \to \infty$$

(5)

If the conditions in the proposition are not satisfied, how should we expect our measures to differ?
Our measure includes \emph{LSV} as a special case

**PROPOSITION**

Let \( \{X^{\tau_i}\} \) and \( \{X^{T_i}\} \) be iid with \( X^{\tau_i} \sim \text{Beta-Bino}(\tau_i, \alpha, \beta) \) and \( X^{T_i} = X^{T_i}_0 + \varepsilon_i \) where \( X^{T_i}_0 \sim \text{Beta-Bino}(T_i, \alpha, \beta) \). If \( \varepsilon_i = 0 \forall i \lor \sum_i \varepsilon_i = 0 \), then

\[
\hat{H}_i \xrightarrow{I \to \infty} \text{LSV}_i, \quad \text{with } \alpha, \beta \to \infty
\]  

(5)

If the conditions in the proposition are not satisfied, how should we expect our measures to differ?

- if \( \varepsilon_i = 0 \forall i \), but \( \alpha, \beta \ll \infty \): \( \hat{H} = 0 \), but \( \text{LSV} > 0 \)
Our measure includes $LSV$ as a special case

PROPOSITION

Let $\{X_{\tau i}\}$ and $\{X_{Ti}\}$ be iid with $X_{\tau i} \sim Beta-Bino(\tau_i, \alpha, \beta)$ and $X_{Ti} = X_{oTi} + \varepsilon_i$ where $X_{oTi} \sim Beta-Bino(T_i, \alpha, \beta)$. If $\varepsilon_i = 0 \forall i \vee \sum_i \varepsilon_i = 0$, then

$$\hat{H}_i \xrightarrow{I} \infty \ LSV_i, \ \text{with} \ \alpha, \beta \to \infty$$

If the conditions in the proposition are not satisfied, how should we expect our measures to differ?

▶ if $\varepsilon_i = 0 \forall i$, but $\alpha, \beta \ll \infty$: $\hat{H} = 0$, but $LSV > 0$

▶ if $\alpha, \beta \to \infty$, but $\sum_i \varepsilon_i \neq 0$:
  e.g. $p = 0.7, \ br_1 = 0.8, \ br_2 = 0.8$, then $\bar{b}r = 0.8$ and $LSV = 0$ assuming
  ($T_i \to \infty$ s.t. $AF_i^{LSV} \to 0$ for $i = 1, 2$), whereas $\hat{H} > 0$
We apply our measure and LSV to German equity transaction data
We apply our measure and $LSV$ to German equity transaction data

- Prime Standard stocks traded on XETRA during 2008
- Transaction data provided by BaFin/Bundesbank ($\S$ 9 WpHG, $\S$ 5 FinStabG):
  - Execution Price
  - Quantity
  - Timestamp ‘2008–01–02 09:50:01’
  - Agent-Principal ‘E’/’K’
  - Trade ‘K’/’V’
  - Trader ID
- XETRA order book data provided CRC 649:
  - Best bid/ask quotes
  - Quantity
  - Timestamp ‘2008–01–02 09:50:01.94’
  - Bid-ask-flag ‘B’/’A’
Empirical Setup

- daily level
Empirical Setup

- daily level
- estimation on cross-section
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
- proprietary trading of reporting institutions (largely banks) and MO of customers
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
- proprietary trading of reporting institutions (largely banks) and MO of customers
- exclude trades from call-auctions (open, mid-day and closing-auctions; volatility and liquidity breaks)
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
- proprietary trading of reporting institutions (largely banks) and MO of customers
- exclude trades from call-auctions (open, mid-day and closing-auctions; volatility and liquidity breaks)
- $\tau_{i,t} = 10$ for all $i, t$
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
- proprietary trading of reporting institutions (largely banks) and MO of customers
- exclude trades from call-auctions (open, mid-day and closing-auctions; volatility and liquidity breaks)
- $\tau_{i,t} = 10$ for all $i, t$
- if $\tau_{i,t} < 10$ until 10 o’clock we excluded the stock-day from the data
Empirical Setup

- daily level
- estimation on cross-section
- market orders only (obtained via trade-classification algorithm)
- proprietary trading of reporting institutions (largely banks) and MO of customers
- exclude trades from call-auctions (open, mid-day and closing-auctions; volatility and liquidity breaks)
- \( \tau_{i,t} = 10 \) for all \( i, t \)
- if \( \tau_{i,t} < 10 \) until 10 o’clock we excluded the stock-day from the data

\( X^\tau = \) sum of buy-market orders of first 10 trades, and \( X^{T_{i,t}} = \) sum of all buy-market orders for the stock-day
We find that . . .

- strong support for the beta-binomial distribution, contrary to the binomial distribution

- both groups show generally less deviation from independent trading than one would expect

- in contrast, the LSV measure indicates persistent deviation from independent trading

- our measure indicates occasionally strong deviations from independent trading for customer trades, . . . apparently related to strong market movements

- no such relation can be observed for the LSV measure
We find that . . .

- strong support for the beta-binomial distribution, contrary to the binomial distribution
- both groups show generally less deviation from independent trading than one would expect
We find that . . .

- strong support for the beta-binomial distribution, contrary to the binomial distribution
- both groups show generally less deviation from independent trading than one would expect
- in contrast, the \( LSV \) measure indicates persistent deviation from independent trading
We find that . . .

- strong support for the beta-binomial distribution, contrary to the binomial distribution
- both groups show generally less deviation from independent trading than one would expect
- in contrast, the *LSV* measure indicates persistent deviation from independent trading
- our measure indicates occasionally strong deviations from independent trading for customer trades, . . .
We find that...

- strong support for the beta-binomial distribution, contrary to the binomial distribution
- both groups show generally less deviation from independent trading than one would expect
- in contrast, the LSV measure indicates persistent deviation from independent trading
- our measure indicates occasionally strong deviations from independent trading for customer trades, ...
- apparently related to strong market movements
We find that...

- strong support for the beta-binomial distribution, contrary to the binomial distribution
- both groups show generally less deviation from independent trading than one would expect
- in contrast, the LSV measure indicates persistent deviation from independent trading
- our measure indicates occasionally strong deviations from independent trading for customer trades, ...  
- apparently related to strong market movements
- no such relation can be observed for the LSV measure
We find overwhelming support in favor of the beta-binomial model:

\[ (a) \text{ Customer Trades} \quad (b) \text{ Proprietary Trading} \]

**Figure:** P-values from Pearson’s goodness of fit test

P-values from a KS-test on uniformly distributed p-values from the GoF-tests are 0.28 and 0.31, respectively.
Our measure does not show an apparent relation to the $LSV$ measure

Figure: $\tilde{H}_t$ vs. $\overline{LSV}_t$
For both groups, our measure shows general underdispersion, $\tilde{H}_t < 0$, 

**Figure:** $\tilde{H}_t$ of proprietary trades
For both groups, our measure shows general underdispersion, $\tilde{H}_t < 0$,
For both groups, our measure shows general underdispersion, $\tilde{H}_t < 0$.

The LSV measure would come to the opposite conclusion.

**Figure: $\tilde{H}_t$ for customer trades**
The occasionally strong deviations from independent trading by more than one would expect, seem to be related to strong market movements.

Figure: $\tilde{H}_t$ for customer trades

Figure: Prime All Share Index
\( \tilde{H}_t < 0 \) seems to be driven in most cases by

\[
\tilde{H}_t < 0 \quad \text{seems to be driven in most cases by}
\]

and \ldots
... $\hat{H}_t > 0$ seems to be driven in most cases by
... $\hat{H}_t > 0$ seems to be driven in most cases by $
abla \hat{H}_t$ overdispersed compared to binomial distribution, but we strongly reject the beta-binomial distribution as well (indicated by bootstrap GoF-tests)
Discussion and Outlook

- Further refinement of interpretability of $\hat{H}$ and $\tilde{H}$
  - Decomposition into buy and sell deviations
  - Distinguishing between contrarianism and herding
- Measure for the tails, “Deviation at Risk” à la VaR
  - $P(|br - p| - AF \geq x)$
- Can $\hat{H}$ predict, explain or be explained by stock price movements and/or volatility?
  - Conduct portfolio analysis
  - Run VAR regressions