Price Discovery in the Cross-Section: Leaders and Followers

Abstract

A frictionless market should allow information to be incorporated into all assets quickly and efficiently. Market frictions however make such incorporation less than perfect. What is the extent of the issue, and what are the main frictions? To answer this question, we use a novel econometric methodology that allows us to analyze the propagation of public information across all the constituents of the S&P 500 index. We show that a sizeable portion of information is captured by only a few equities, suggesting that not all assets incorporate information in a timely manner. Our analysis suggests that stocks that are slow to adjust are those with significant liquidity demand pressures and high trading costs.

Keywords: Price adjustment, Price discovery, Systematic risk

JEL classification: G14, G23.

*We thank Canadian Derivatives Institute (CDI) and SSHRC for financial support.

1 Introduction

An optimally functioning financial market should allow asset managers and other market participants to quickly adjust their positions in response to the arrival of new information. In theory, such changes should happen swiftly so that all stock prices reflect public information in a timely manner. Prior research shows that the adjustments indeed occur efficiently over medium and long periods of time (e.g., Bai, Philippon, and Savov, 2016). But how quickly do prices adjust within the day? Do adjustment speeds vary in the cross-section, and if so what are the determinants of this variation? We suggest that the speeds may indeed differ across assets due to a variety of frictions faced by market participants. For instance, if the cost associated with opening and closing of a position is high, prices may adjust slowly or not at all. A similar situation could arise when liquidity demand pressures are high, causing excessive volatility. In such cases, market participants may wait until the pressures subside, otherwise facing a more volatile and uncertain environment. As such, there is a possibility that stock prices, even in today's fast markets, do not incorporate information immediately.

Understanding dislocations of common information in stock prices is important for at least three reasons. First, it provides a new cross-sectional view of how prices impound public information, departing from previous studies that seek to understand how price discovery is done within a group of securities that share a common efficient price. For instance, Harris (1989) and Chan (1992) describe the links between stock and stock index futures markets, whereas Bhattacharya (1987) and Easley, O'Hara, and Srinivas (1993) study the links between stock and option markets. Second, the ubiquity of exchange traded fund (ETF) trading has introduced large intraday demand pressures on the underlying stocks due to arbitrage activity (Ben-David, Franzzoni, and Moussawi, 2018). This new type of intraday liquidity shocks increases stock volatility and may therefore affect the efficiency of information incorporation into prices. Third, it allows us to appreciate relative gains in price efficiency, shedding light on the way markets incorporate common information across stocks (Fama, 1970).

In this paper, we investigate if the stocks exposed to common price shocks incorporate them into prices in a timely manner. Using the largest and most liquid stocks in the US market, those underlying the S&P 500 index, we document sizeable differences in reactions to new information in the cross section.

To study how market-wide information is impounded into prices, we propose a new price discovery measure that identifies the stocks that lead information incorporation in the cross section. This measure can be viewed as an extension of Hasbrouck (1995) to a case in which information is the innovations in the market portfolio, rather than a common efficient stock price across multiple venues. Our interest with this shift in paradigm is to measure price discovery across stocks about a common component. We implement this measure using one-second data for all stocks that are part of the S&P 500 index during 2012, given that this group of stocks enjoys stable characteristics that allow for comparison across periods.

The information share measure builds on the common factor approach of De Jong and Schotman (2010) and Westerlund, Reese, and Narayan (2017). This methodology conveniently formulates the problem of price discovery as one in which the efficient price component is seen as a common factor, so that existing methods can be employed in their estimation. In our case, the common factor is the market portfolio return, which we estimate at a given point of time from the cross-sectional average of all stocks in the intersection of CRSP and DTAQ. The basic idea of the information share measure is to identify the proportion of information in a given stock relative to other stocks that are also exposed to common innovations in the market portfolio. Thus, this measure provides the proportion of information in a given stock relative to the total contribution of stocks in the panel.

Our empirical analyses generate four sets of main findings. First, when we look at the daily information share across stocks in the S&P 500 index, we find that a small group contains a disproportional amount of information about common innovations in the market portfolio. For instance, when we sort stocks according to their daily information share, we find that the top quintile contains about 70% of the total information in the common innovation. This amount is in striking contrasts with the total information contained in the lowest information share quintile, which contains on average less than 1%. These proportions are consistent when we estimate the daily information share from lower intraday frequencies, showing that this result is not a mechanical effect from stale quotes or non-synchronous trading (Campbell, Lo, and MacKinlay, 1997), but rather support the view that price discovery about the market portfolio is carried out by a small group of stocks.

Second, we show that there is an notable level of persistence among constituents of information share groups, particularly among the stocks that belong to the lowest information share quintile in a given day. We find that these stocks are in the same quintile the next day 38% of the time, much more often than the 29% observed for stocks remaining in the highest quintile. Moreover , the lower persistence in the highest quintile suggests that there is nontrivial rotation among stocks incorporating common information, but there is still a fraction of them consistently integrating information day after day. When we compare characteristics of stocks in these two information groups, we find important differences: stocks in the highest (lowest) quintile have higher (lower) market capitalization, lower (higher) transaction costs, and higher (lower) systematic risk.

Third, we use panel regressions to study the determinants of information share ratios. When we look at specifications that only account for stock characteristics unrelated to liquidity, we find that market capitalization and the proportion of systematic risk in the stocks return (R^2) are positively related with the information share. However, when these variables are used in conjunction with others related to liquidity, i.e., trading activity, demand pressure, and liquidity of the stock, we find that only the latter group of variables is significant. In particular, demand pressure plays a deterrent role in the proportion of information impounded in the price. We view this as evidence that the stock's ability to impound common information comes primary from microstructure factors associated with the stock's trading environment. These factors speak directly to the signal-to-noise ratio of the stock, as observed prices become more imprecise about fundamental quantities when these microstructure components grow in importance.

Fourth, we use Dimson (1979) regressions to compare the speed of adjustment of portfolio values to market returns for stocks sorted into groups according to their information share. The portfolio containing stocks with the largest information share are less sensitive to lagged information in the market return. In contrast, portfolio returns of stocks with the lowest information share are more sensitive, creating a lagged response to the common factor shock.

The rest of the paper is organized as follows. Section 2 presents a simple model to understand price adjustments and the measure of information share. Section 3 presents the data and descriptive statistics about this variable. Section 4 empirically assess the determinants of information shares and Section 6 extends analysis to lower frequencies. Section 6 examines that speed of adjustment of portfolios based on the proposed measure. Section 7 concludes.

2 Model and Methodology

In this section, we start by proposing a simple model to analyze how stock prices incorporate information embedded in a common market factor under market frictions. We then define a measure of the information share that a stock return has with respect to innovations in the common factor. Finally, we propose a method to estimate this share from a panel of observed returns.

2.1 Price adjustment to Market Factor Changes

We start with a benchmark model under perfect capital markets with no frictions, in which observed stock returns are a function of a common and an idiosyncratic component as described by the following equation:

$$r_{i,t} = \beta_i f_t + \varepsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ is the observed stock return for asset i in period t, β_i is the company i's loading on a common factor innovation f_t, and $\varepsilon_{i,t}$ represents idiosyncratic innovations. We assume that both types of innovations are i.i.d., with variances of σ_f^2 for the common factor innovation and σ_ε^2 for the idiosyncratic innovation.[1] Under Equation (1) with frictionless and efficient markets, stock returns immediately reflect changes in the common factor innovations f_t, and this effect depends only on each stock's level of systematic risk β_i. Thus, innovations in prices are only driven by the arrival of new information and fully reflect fundamental values at all times.

[1]Stock returns are $r_{i,t} = \pi_{i,t} - \pi_{i,t-1}$and factor innovations (returns) are similarly defined as $f_t = F_t - F_{t-1}$ where $\pi_{i,t}$ and F_t are observed stock and factor prices, respectively.

However, due to different market imperfections and sources of inefficiencies, the incorporation of new information is not immediate and some level of autocorrelation could exist.[2] Therefore, to account for the effect of these frictions, not only the stock's β_i may influence price reaction to changes in the market factor, but also an additional adjustment speed component is incorporated in the price formation process. To formalize this mechanism, let $p_{i,t}$ denote the efficient log-price of stock i at time t, and $\alpha_i = [0, 1]$ the speed of price adjustment to permanent shocks in the common market factor. In this set up, stocks with larger values of α incorporate information faster. Hence, we assume that the efficient price incorporates information according to the following process:

$$p_{i,t} = \alpha_i \beta_i F_t + (1 - \alpha_i) p_{i,t-1}. \tag{2}$$

The intuition behind Equation (2) is straightforward: a fundamental change in the common factor price will have a larger effect in stocks with larger speed of price adjustment α_i and larger level of systematic risk, as measured by its factor loading β_i. Notice that stocks with $\alpha_i = 1$ immediately adjust to the arrival of the common innovation F_t. In contrast, stocks for which $\alpha_i < 1$ do not fully incorporate information at time t, as current price changes also impound lagged adjustments. Moreover, if a stock has an $\alpha_i = 0$, the efficient price at time t does not incorporate common information and will only reflect past innovations.[3]

When incorporating market frictions in the price formation process, another possibility is

[2]Examples of how market imperfections could lead to slow incorporation of information include the mechanical effect of bid-ask bounce documented in Roll (1984), the adjustment of quotes done by liquidity providers who manage inventory risk proposed in Ho and Stoll (1981), or convergence of prices toward fully-informational values as in sequential trade models of Glosten and Milgrom (1985) and Kyle (1985).

[3]The model presented in Equation (2) has the spirit of Hasbrouck and Ho (1987), but incorporates a common shock in the stock price.

that efficient and observed prices differ. To allow for this effect, let $\pi_{i,t}$ denote the observed transaction price of stock i at time t so that the differences between efficient and observed prices are expressed with the following representation:

$$\pi_{i,t} = p_{i,t} + \epsilon_{i,t}, \tag{3}$$

where $\epsilon_{i,t}$ is an error process such as $\triangle\epsilon_{i,t}$ has a constant variance of $\sigma^2_{\epsilon_i}$.

Combining equations (3) and (2), we obtain a version of (1) that incorporates the effects of sensitivity to market shocks, microestructure effects, and the speed of price adjustments:

$$\begin{aligned} r_{i,t} &= \Delta\pi_{i,t} \\ &= \alpha_i\beta_i f_t + (1-\alpha_i)r_{i,t-1} - (1-\alpha_i)\Delta\epsilon_{i,t-1} + \Delta\epsilon_{i,t}. \end{aligned} \tag{4}$$

Notice that equation (4) converges to the model in equation (1) when the speed of price adjustment for all stocks is $\alpha_i = 1$. In this case, as in the benchmark model, stock returns immediately incorporate the effects of common innovations depending on their level of systematic risk β_i. Clearly, in case that $\alpha_i < 1$, the stock's β_i is muffled by the price adjustment speed, so stock returns reflect only partial factor innovations. Interestingly, a speed of adjustment $\alpha_i < 1$ will also imply that stock returns have some significant level of auto-correlation, captured by the ARMA terms $(1-\alpha_i)r_{i,t-1}$ and $(1-\alpha_i)\Delta\epsilon_{i,t-1}$.

The representation of price adjustments in equation (4) follows existent empirical literature documenting how both positive and negative autocorrelations in returns relate to informational inefficiency. For instance, Hendershott and Jones (2005) use first-order return autocorrelations to illustrate how price adjusts more slowly when quote transparency decreases. Comerton-Forde and Putniņŝ (2015) combine autocorrelations computed at different frequencies to measure information efficiency, which is shown to decrease with high levels of dark trading.

2.2 Stock's Information Share of the Market Factor

The above representation of observed returns is particularly useful because it illustrates how price changes driven by factor innovations are determined by stock specific parameters. Thus, cross-sectional measurements about the amount of common information absorbed by a given stock should help identify stocks that are relatively more efficient than others in the process of incorporating new information from the market factor. We now propose a formal measure that captures how information from market shocks is incorporated to different stock prices.

Suppose that f_t and $\epsilon_{i,t}$ are observable, and let $v_{i,t} = \alpha_i \beta_i f_t + \Delta\epsilon_{i,t}$ denote the sum of common ($\alpha_i \beta_i f_t$) and idiosyncratic contemporaneous ($\Delta\epsilon_{i,t}$) innovations to the observed stock returns, from equation (4). Then, as in De Jong and Schotman (2010), it is possible to identify stocks that contain a larger share of information about f_t by looking at the explanatory power of each $v_{i,t}$ in the following auxiliary reverse regression:

$$f_t = b_1 v_{1,t} + b_2 v_{2,t} + \cdots + b_N v_{N,t} + e_t, \tag{5}$$

where $b_1, \ldots, b_N$ are the regression coefficients and e_t is the regression error term. Specifically we define or measure of information share as the proportion of the R^2 explained by each stock $v_{i,t}$. As we show in the appendix, the R^2 coefficient of the regression above can be conveniently expressed as follows:

$$R^2 = \sum_{i=1}^{N} b_i \alpha_i \beta_i. \tag{6}$$

From the equation above, it is clear that each stock i contribution to R^2 is given by $b_i\alpha_i\beta_i$, thus we define our measure of information share as IS_i=$b_i\alpha_i\beta_i$. To gain further intuition about IS_i, we can asymptotically express it as function of the original model parameters as follows:[4]

$$IS_i = \frac{(\alpha_i\beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{\sum_{j=1}^{N} (\alpha_j\beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}} \tag{7}$$

The above equation states that the contribution of a given stock return to the total variation in the common innovation depends on two sources. The first one is the loading on the common innovation, $(\alpha_i\beta_i)^2$, which captures the degree of association between stock i and the market factor. The degree of association between stock returns and the common factor depends on the speed at which the stock incorporates information and the systematic risk of the company. Over short horizons, stocks may incorporate information at different speeds, but in long horizons, it is expected that $\alpha_i = 1$ for all stocks and the stock's information share will only capture stock systematic risk.

The second source is the ratio between the common innovation's variability and the one from the stock's idiosyncratic component. This second term measures the precision with which contemporaneous innovations are informative about market factor innovations, since low(high) relative values of $\sigma_{\epsilon_i}^2$ magnify(reduce) the contribution of $\alpha_i\beta_i$. In addition, the amount of information is also a function of the microstructure noise present in the observed stock price, since prices that diverge more from the efficient component would have lower

[4]A detailed derivation is presented in the appendix.

information shares.

2.3 Estimation of Stock's Information Share

We now address the question of how to estimate the information share of a given stock, as measured by Equation (7). Assume that a panel of log-returns is available for N companies, each of which having a number of T observations at a given observation frequency. All parameters required for estimation of information share can be obtained from the representation in equation (4)

$$r_{i,t} = \alpha_i \beta_i f_t + (1 - \alpha_i) r_{i,t-1} + (1 - \alpha_i) \Delta\epsilon_{i,t-1} + \Delta\epsilon_{i,t}. \tag{8}$$

This equation shows that log-returns have contemporaneous and lagged information. Accordingly, our estimation procedure involves a two-step methodology. In the first step, we estimate an ARMA model of observed log returns to isolate the contemporaneous from the auto-regressive components in returns, so the residuals from the ARMA regressions $\tilde{r}_{i,t}$ only captures the contemporaneous effects.[5] Using these residuals, we propose the following factor model:

$$\tilde{r}_{i,t} = \alpha_i \beta_i f_t + \Delta\epsilon_{i,t} = \lambda_i f_t + \Delta\epsilon_{i,t} \tag{9}$$

Notice that the model above is the model in equation (8) after removing the auto-regressive components. We re-parameterize $\alpha_i \beta_i = \lambda_i$ so the parameters in Equation (9) can be estimated using standard methods for factor models in large dimensions (Pesaran

[5]In our empirical implementation, we find that the AR(1) specification provides indistinguishable results from an ARMA(1,1), so we use the former model through the analyses.

2006, Bai 2003). We use these estimated factor loadings $\hat{\lambda}_i$, the common factor $\hat{f}_t$, and the factor and idiosyncratic variances $\hat{\sigma}_f^2$ and $\hat{\sigma}_{\epsilon_i}^2$ to estimate our measure of information share as follows:

$$\widehat{IS}_i = \frac{\hat{\lambda}_i^2 \frac{\hat{\sigma}_f^2}{\hat{\sigma}_{\epsilon_i}^2}}{\sum_{j=1}^{N} \hat{\lambda}_j^2 \frac{\hat{\sigma}_f^2}{\hat{\sigma}_{\epsilon_i}^2}} \tag{10}$$

In the appendix we provide additional estimation details and show that under suitable assumptions our $\widehat{IS}_i$ estimator is consistent as N and $T \rightarrow \infty$.

3 Data Sources and Sample Overview

3.1 Intraday Stock Return Sample

We focus on the set of stocks that compose the S&P500 index for the entire year of 2012, since these stocks are some of the most liquid securities in the market and together constitute one of the most important benchmarks in financial markets. For our study, cross-sectional diversities in this sample provide an appropriate setting to study the differential speed with which stocks react to the arrival of common information. We obtain the S&P 500 index constituents from the Center fpr Research in Security Prices (CRSP) and keep those that are part of the index through 2012 and for which there are daily data in the DTAQ database. This selection criterion leaves us with 474 stocks.

For the list of eligible companies, we take the NBBO file in DTAQ and join at a millisecond level with quotes from NASDAQ that are not in this file. The resulting sample is used to construct one-second log-returns from the last recorded NBBO in the observation interval. To isolate opening or closing effects, observations between 9:35 a.m and 3:55 p.m. are the only ones consider. We remove from our sample days in which the market was not operating full day. To construct the market factor at the intraday level, we use NBBO mid-quotes for each stock in the sample and compute an equally- and value-weighted factor by cross-sectionally averaging log-returns at a one-second frequency.[6]

3.2 Information Share of S&P 500 Stocks

Daily estimates of the Information Share are computed for each company following the methodology presented in Section 2.3. These values are then averaged across the sample period. Table 1 presents statistics of this variable for the cross-section of stocks in the S&P 500 index. We observe that the average stock has an information share of 0.21%, which corresponds to a value close to 1/500. This value represents a situation where information about the market factor is equally impounded in the sample of stocks studied. From the skewness and percentiles in the table, it is clear that the unconditional distribution of IS is rather symmetric, showing that there are stocks with lower (larger) information shares than what would otherwise the typical level for a stock. Panel B in the table shows that this conclusion is also obtained when the market factor is computed using value weights.

We now look at the relative information share of stocks across the sample period. To this end, every day we sort stocks into 5 groups according to their IS for a given day. The group with label 0 represents stocks in the lowest information group, while the label 4 is given to the group with the highest information share. We present in Table 1 statistics about this variable as well as statistics about the number of days that stocks stay in the most (least) informed groups. Although stocks spend a similar number of days in both groups, we observe a larger proportion of stocks in the lowest information group, as evidenced by the positive

[6]In unreported results, we also consider the case where the market factor is estimated from all the cross-section of stock returns, and obtain similar results to those reported in the following sections. We also run the same analyses independently for 2010 and 2011, and our main conclusions hold.

Table 1: **Information Share Characteristics.**
This table presents summary statistics about the information share that companies in the S&P 500 have with respect to the market portfolio. Panel A shows summary statistics across stocks using as market portfolio an equally-weighted average of common stocks in the intersection of CRSP and DTAQ. The variable *IS* represents the average information share (in percentage) as defined in Equation (10) for a given stock across the sample period. The variable *IS Group* refers to the average quintile in which a stock is classified according to its daily IS (the lowest information group is 0 and the highest one is 4). The variable *# Days - Lowest IS* provides the number of days that a given stock belongs to the lowest IS quintile. The variable *# Days - Highest IS* provides the number of days in the highest IS quintile. Panel B shows results based on a value-weighted market factor. The sample period is 2012 with 246 trading days.

Panel A: Equal-Weighted Market Index

	Nobs	Median	Mean	StdDev	Skew	P10	P90
IS	474	0.19	0.21	0.13	0.94	0.07	0.40
IS Group	474	2.05	2.00	0.63	-0.49	1.13	2.78
# Days - Lowest IS	474	39.00	48.98	40.35	1.70	10	100
# Days - Highest IS	474	46.50	49.10	30.63	0.45	11	93

Panel B: Value-Weighted Market Index

	Nobs	Median	Mean	StdDev	Skew	P10	P90
IS	474	0.19	0.21	0.14	1.22	0.06	0.40
IS Group	474	2.07	2.00	0.67	-0.47	1.11	2.83
# Days - Lowest IS	474	37.00	48.98	42.64	1.74	9	103
# Days - Highest IS	474	46.00	49.10	32.00	0.54	10	95

skewness and larger standard deviation.

To appreciate the level of information across groups, for each day in the sample, we add within each group the information share attributed to each stock. Table 2 shows the median value for the total information share across the five groups. We observe a large dispersion in the amount of total information share present in each group. Whereas the lowest group captures less than 1% of the total variation in the market factor for a given day, the group with the largest IS captures about 70%. In Figure 1, we plot the daily information share of each group across the sample period. From this figure, it is clear that these differences are persistent across time, illustrating that a small group of stocks carries most of the information about the market factor during a given day. Nonetheless, the fact that the average stock has an IS of .21% shows that it is not necessary true that the same stocks are being attributed with the largest source of variation in the market factor component.

Table 2 also reports characteristics of stocks in each of the information groups. Each day, we compute variables such as market capitalization, CAPM-beta obtained from regressing the previous 250 daily returns on CRSP's daily market factor, and the volatility as measured by the standard deviation of the previous 250 daily returns. In addition, we report measures of trading activity on the stock such as total number of trades and dollar volume, and measures of stock's trading costs like the average quoted and effective percentage spread. These last four variables are obtained from WRDS. From the table, the group with the highest information share are large, have high systematic risk, and compared to the group with the lowest information share, experience a relative low level of trading activity and trading costs.

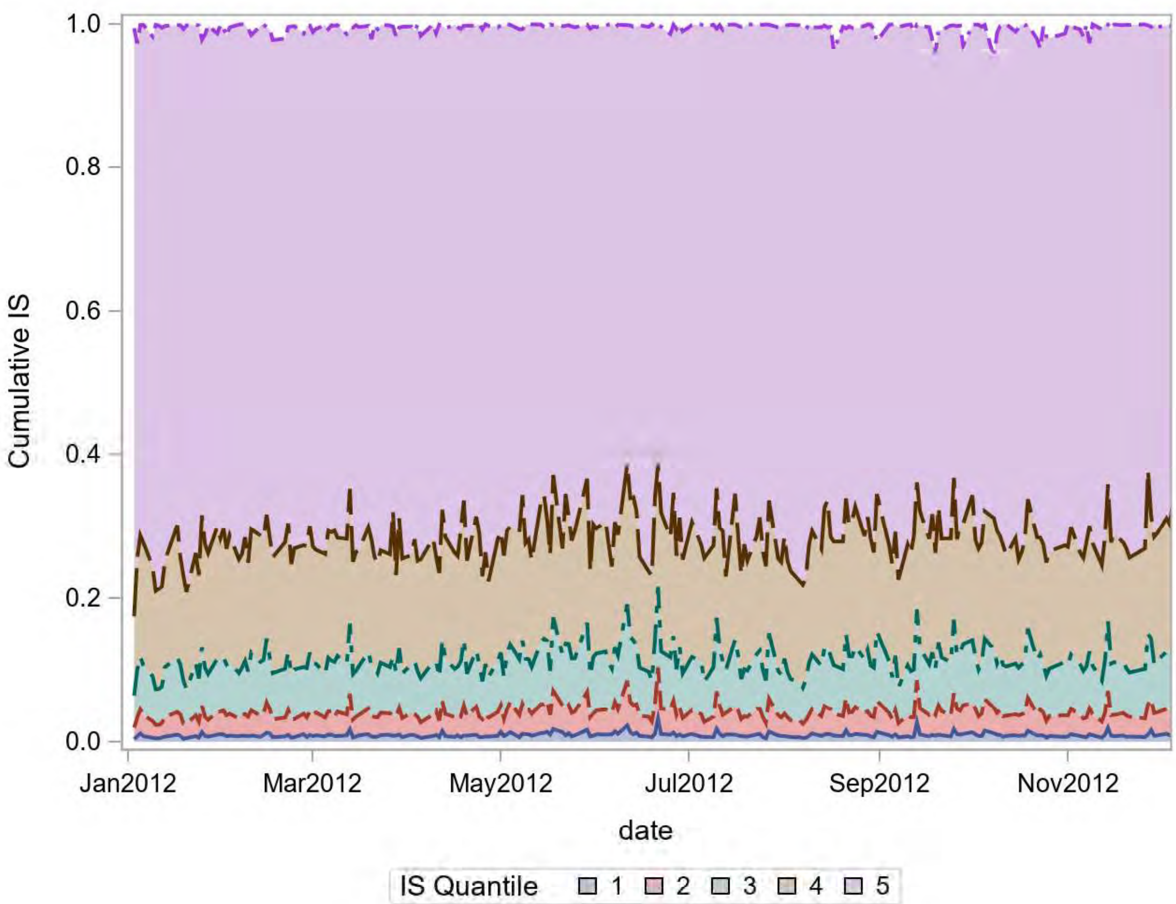

Figure 1: **Total Information Share by Quintiles.** The figure shows the total information share of IS quintile across the sample. For each day, the proportion of information in each quintile is cumulatively added and depicted with a different color. The IS quintile 1 represents the group of stocks with the lowest information share (bottom colored region in the figure), while the IS quintile 5 represents the one with the highest information share (top colored region in the figure). The sample period is 2012.

Table 2: **Information Share Group Characteristics.**
This table presents characteristics for stock groups formed by the level of information share. Every day, we sort stocks into five groups with cut-off points determined by quintiles of the IS distribution. Within each group, we compute the total sum of IS -in percentage- (*Total IS*) and obtain median values of the stock's market value (*Size*), CAPM beta (*Beta*), volatility (*Total IS*), total number of trades (*Number Trades*), total volume (*Volume*), quoted percentage spread (*Quoted Spread*), and effective percentage spread (*Effective Spread*). Variables *Size* and *Volume* are measured in millions; spreads are presented in basis points. This table presents median values of each variable across the 246 trading days of 2012. There are 447 stocks in the cross-section. Panel A presents results for the case an equal-weighted market index is used in the construction of IS. Panel B presents results when the value-weighted market index is employed.

Panel A: Equal-Weighted Market Index						
	Lowest IS	Q2	Q3	Q4	Highest IS	All
Total IS	0.83	2.87	6.70	15.76	65.84	6.70
Size	9.10	10.86	12.31	12.93	15.63	12.02
Beta	0.98	1.02	1.05	1.09	1.15	1.05
Volatility	0.31	0.31	0.32	0.32	0.33	0.32
Number Trades	17,555	14,358	14,338	14,306	15,361	14,987
Volume	3.39	2.48	2.36	2.34	2.47	2.53
Quoted Spread	4.78	3.94	3.66	3.54	3.31	3.74
Effective Spread	4.67	3.81	3.61	3.46	3.29	3.66
Panel B: Value-Weighted Market Index						
	Lowest IS	Q2	Q3	Q4	Highest IS	All
Total IS	0.77	2.76	6.59	15.58	66.69	6.59
Size	8.64	10.72	12.25	13.27	16.54	12.06
Beta	0.99	1.02	1.05	1.09	1.14	1.05
Volatility	0.32	0.31	0.32	0.32	0.33	0.32
Number Trades	17,529	14,337	13,973	14,380	15,902	15,023
Volume	3.32	2.38	2.34	2.37	2.59	2.52
Quoted Spread	4.96	3.94	3.65	3.50	3.24	3.75
Effective Spread	4.80	3.81	3.61	3.44	3.25	3.65

3.3 Information Persistence

Next, we examine the persistence of stocks' information share by looking at the empirical distribution of stock transitions between IS quintiles across days. Table 3 presents 5×5 transition matrices that indicate subsequent information share levels (columns) as a function of initial ones (rows). Panel A (B, D) displays 1-day (3-day, 5-day) transition probabilities from overlapping time periods.

For a 1-day period, we observe that 38% of stocks in the lowest information share quintile remain in this quintile one day after (row one, column one in Panel A). This number slightly decreases to 35.6% and 34.8% when looking at the same matrix entry in the 3- and 5-day transition probabilities, suggesting a high degree of persistence. Stocks with low information share are likely to exhibit low information shares ex post at 1-,3-, and 5-day horizons. This persistence contrasts with that observed in the quintile of stocks with the highest information share. When we look at the last row and column of transition probabilities, we observe lower percentage of stocks staying in this group. For instance, at a 1-day level, a stock in the highest information share quintile remains in this category 30.7% of the time. We also observe from the transition probabilities that firms with less (more) information shares are less likely to transition into a group with a higher (lower) information share. This asymmetric result shows that a large group of stocks are not likely to incorporate information about the market factor for a number of consecutive days, illustrating that the dynamic process of incorporating information about this common factor is relegated to few stocks.

Table 3: **Transition Probabilities.**
This table reports transition probabilities of IS quintiles across days. Every day, we group observations into quintiles based on their information share. The result is a 5 × 5 matrix for which we compute the empirical distribution of transitions between a given day and a specific subsequent period. In Panels A, B, and C, we examine 1-, 3-, and 5-day transition probabilities, respectively. The sample period is 248 trading days in 2012.

	Lowest	Quintile 2	Quintile 3	Quintile 4	Highest
Panel A: 1-Day Transition Probabilities					
Lowest Information Share	0.380	0.224	0.172	0.140	0.084
Quintile 2	0.219	0.220	0.204	0.196	0.161
Quintile 3	0.169	0.205	0.208	0.212	0.207
Quintile 4	0.142	0.193	0.209	0.218	0.238
Highest Information Share	0.086	0.161	0.209	0.237	0.307
Panel B: 3-Day Transition Probabilities					
Lowest Information Share	0.356	0.217	0.179	0.148	0.101
Quintile 2	0.215	0.217	0.203	0.195	0.169
Quintile 3	0.179	0.206	0.206	0.203	0.206
Quintile 4	0.146	0.196	0.208	0.217	0.233
Highest Information Share	0.099	0.167	0.206	0.240	0.288
Panel C: 5-Day Transition Probabilities					
Lowest Information Share	0.348	0.219	0.178	0.150	0.105
Quintile 2	0.216	0.216	0.205	0.193	0.169
Quintile 3	0.177	0.205	0.203	0.209	0.206
Quintile 4	0.151	0.192	0.210	0.214	0.232
Highest Information Share	0.103	0.171	0.205	0.236	0.285

4 Determinants of Information Shares

In this section, we empirically assess which stock characteristics relate with the level of information about the market factor contained in stock prices. To select these variables, we look at the two main components of IS in equation (7) and highlight which characteristics could play an important role in this variable empirically.

We start with the term $\alpha_i\beta_i$. This term captures the stock's systematic risk and the speed at which it incorporates this information. From this perspective and given the parsimony of the model, the first set of variables to be tested correspond to variables related to the systematic component of a stock. The first one we include is CAPM beta, which provides a standardized measure of co-movement between the stock and the market portfolio. The second one is the R^2 resulting from the regression employed in the estimation of beta. This variable helps disentangle the proportion of systematic risk present in the stock return. We compute these variables for each day in the sample by using the previous 250 daily excess returns of a given stock and regressing them over the daily market factor obtained from CRSP. Given cross-sectional differences in company's size and the marked interest in the largest companies of S&P 500 index, we also include size in this set of variables. This variable also controls for the fact that as market capitalization increases, the potential profits associated with acquiring information increase, so stocks for large companies will be more likely to incorporate new information faster.

Next, we look at variables that could relate to the speed that information is impounded in the stock. Given the role of algorithmic trading in the dissemination of information in equity markets, we consider different variables that proxy for this activity. First, following Weller (2018), we use three variables associated with the presence of algorithmic trading: odd lot volume ratios, trade-to-order volume ratios, and cancel-to-trade ratios. Odd lot volume ratio is the total volume executed in quantities smaller than 100 shares dividied by the total volume traded, the variable trade-to-order volume ratio is the total volume traded divided by the total volume across all orders placed, and the cancel-to-trade ratio corresponds to the number of full or partial cancellations divided by the number of trades. A higher odd lot and cancel-to-trade ratios signal more algorithmic trading, as well as a lower trader-to-order volume ratio. Second, we look at the ratio between the total number of intermarket sweep orders employed in the day divided by the total number of trades. This variable is meant to capture information from trading activity by agents looking for faster execution.

Another component determining the stock's information share is the microstructure noise present in observed stock prices, σ_{ϵ_i}. The first variable we include is the daily dollar volume of the stock (in logarithm), which captures trading activity in the stock. The second one is the demand pressure exerted on the stock, measured by the absolute difference between buyer and seller initiated trades divided by the total number of trades. Finally, regarding liquidity costs, we include the daily percent quoted spread following Holden and Jacobsen (2014). All of these variables are obtained from WRDS. Descriptive statistics of these variables are provided in Table 4.

Next, to analyze these effects, we estimate the equation

$$IS_{i,t} = a + \mu_i + \gamma' X_{i,t} + bIRet_t + e_{i,t}, \tag{11}$$

where IS is the information share for stock i on day t, μ_i denotes a firm fixed effect, and X corresponds to a vector with stock-level characteristics. The model also includes monthly-time fixed effects in some of the specifications. We estimate the model using ordinary least squares, clustering observations by stock when calculating standard errors. We only work

Table 4: **Descriptive Statistics.**
This table presents descriptive statistics for our sample of XXX firm-years with available characteristics. Columns (1), (2), and (3) present the median, mean, and standard deviation, respectively. Columns (4) and (5) present the 10th and 90th percentiles. The variable (*Size*) represents the total market capitalization computed from closing prices (in logarithm). To compute the variables α_{CAPM} and β_{CAPM}, for a given day and stock, we estimate the equation $R^e = a + \beta_{CAPM} R^{M,e} + u_t$ using the previous 250 daily observations of stock and market excess returns. The variable R^2 represents coefficient of adjustment in this regression. The variable *Volume* is the dollar volume of the stock (in logarithm). The variable *Quoted Spread* corresponds to the average of quoted percentage spread computed from all NBBO in a given day (multiplied by 100). The variable *Absolute OIB* is the absolute difference between the number of buyer minus seller initiated trades, divided by the sum of the two. The variable *ISO Orders* is the total number of intermarket seep orders divided by the total number of trades for the stock. Odd lot volume ratio (*Odd lot ratio*) is the total volume executed in quantities smaller than 100 shares divided by the total volume traded, the variable trade-to-order volume ratio (*Trade-to-Order ratio*) is the total volume traded divided by the total volume across all orders placed, and the cancel-to-trade ratio (*Cancel-to-Trade ratio*) corresponds to the number of full or partial cancellations divided by the number of trades. There are 447 stocks in the cross-section.

	Median	Mean	StdDev	Skew	P10	P90
Size	16.284	16.429	1.028	0.737	15.216	17.838
β	1.06	1.09	0.4	0.373	0.56	1.63
Systematic risk	0.54	0.519	0.173	-0.407	0.27	0.734
Volume	14.75	14.809	1.104	0.228	13.457	16.228
Quoted spread	0.038	0.047	0.037	5.475	0.023	0.076
Absolute OIB	0.042	0.052	0.043	1.488	0.008	0.11
ISO orders	0.454	0.451	0.071	-0.232	0.358	0.541
Trade-to-Order ratio	0.209	0.234	0.124	4.246	0.118	0.373
Cancel-to-Trade ratio	2.586	2.824	1.338	1.238	1.366	4.557
Odd order ratio	0.073	0.085	0.06	2.156	0.025	0.154

with stocks for which all variables are available during the sample period, which reduces the cross-section from 474 to 447 stocks.

Table 5 shows results for three specifications of the above model. The first specification, Column (1), contains characteristics related to the stock's systematic component. In this case, we observe a positive and significant relation of stock's information share with size, and no significant relation with the other two variables. To assess the economic significance of this coefficient, we multiply the variable's standard deviation by its coefficient in the regression, and find that an increase in one standard deviation of the variable size leads to an increase of 0.095% in the stock's information share. Recall from Table 1 that the standard deviation of IS is 0.13%, meaning that an increase of 0.095% represents roughly three quarters of a typical change in this variable.

Column (2) in Table 5 examines the role of next group of variables. The results show that all these variables are significantly related with the stock's information share. When we look at the economic significance of these variables, the largest impact in absolute value is observed for the Trade-to-Order ratio, where a change in the standard deviation of this variable leads to a decrease of 0.41% in the stock's IS: a change that corresponds to about 3.2 times the size of the IS' standard deviation. The coefficient for varible Volume has the second largest impact, as an increase in this variable translates into a decrease of 0.074% in IS. Variables such as the Odd lot ratio, the quoted spread, and ISO orders have economic effects of -0.046%, -0.03%, and 0.025%, respectively. Finally, with a lower order of magnitude the economic effect of the absolute OIB and Cancel-to-Trade ratio of -0.007% and 0.004%, respectively.

We next consider in column (3) of Table 5 a specification including all variables. We observe that the coefficient of size decreases by more than half, and that of beta increases

Table 5: **Stock's Characteristics and Information Share.**
This table reports results from panel regressions in which the dependent variable is the information share of a stock over day t. The variable (*Size*) represents the total market capitalization computed from closing prices (in logarithm). To compute the variables α_{CAPM} and β_{CAPM}, for a given day and stock, we estimate the equation $R^e = a + \beta_{CAPM} R^{M,e} + u_t$ using the previous 250 daily observations of stock and market excess returns. The variable R^2 represents coefficient of adjustment in this regression. The variable *Volume* is the dollar volume of the stock (in logarithm). The variable *Quoted Spread* corresponds to the average of quoted percentage spread computed from all NBBO in a given day. The variable *Absolute OIB* is the absolute difference between the number of buyer minus seller initiated trades, divided by the sum of the two. The variable *ISO Orders* is the total number of intermarket seep orders divided by the total number of trades for the stock. Odd lot volume ratio (*Odd lot ratio*) is the total volume executed in quantities smaller than 100 shares dividied by the total volume traded, the variable trade-to-order volume ratio (*Trade-to-Order ratio*) is the total volume traded divided by the total volume across all orders placed, and the cancel-to-trade ratio (*Cancel-to-Trade ratio*) corresponds to the number of full or partial cancellations divided by the number of trades. All coefficients, except for the one associated to Quoted spread and Cancel-to-Trade ratio, are multiplied by 100. The coefficient of Cancel-to-Trade ratio is multiplied by 10000. Standard errors are computed with robust errors clustered by stock and provided below each coefficient. There are a total of 109,165 observations in the panel. Statistical significance at the 1% and 5% levels is indicated by ** and *, respectively. There are 447 stocks in the cross-section. The sample period is 2012.

	(1)		(2)		(3)	
Constant	-1.321***	-1.27***	1.054***	1.19***	1.259***	1.353***
	(0.25)	(0.242)	(0.084)	(0.087)	(0.286)	(0.285)
Size	0.092***	0.089***			-0.016	-0.015
	(0.015)	(0.015)			(0.017)	(0.017)
β	0.016	0.022			0.055**	0.082***
	(0.025)	(0.021)			(0.027)	(0.024)
Systematic risk	0.039	0.02			-0.007	-0.08***
	(0.039)	(0.017)			(0.038)	(0.017)
Volume			-0.067***	-0.078***	-0.067***	-0.074***
			(0.005)	(0.006)	(0.005)	(0.006)
Quoted spread			-0.791**	-0.69**	-0.878***	-0.8***
			(0.31)	(0.272)	(0.318)	(0.279)
Absolute OIB			-0.168***	-0.171***	-0.167***	-0.17***
			(0.024)	(0.025)	(0.024)	(0.025)
ISO orders			0.348***	0.324***	0.349***	0.335***
			(0.03)	(0.03)	(0.03)	(0.03)
Trade-to-Order ratio			-3.274***	-2.614***	-3.318***	-2.912***
			(0.409)	(0.392)	(0.406)	(0.395)
Cancel-to-Trade ratio			0.327***	0.307***	0.329***	0.308***
			(0.064)	(0.062)	(0.064)	(0.061)
Odd lot ratio			-0.768***	-0.622***	-0.747***	-0.649***
			(0.085)	(0.086)	(0.087)	(0.088)
Month FE	YES	NO	YES	NO	YES	NO
Firm FE	YES	YES	YES	YES	YES	YES
R^2	10.32%	10.32%	13.4%	13.27%	13.41%	13.3%
Observations	109,165	109,165	109,165	109,165	109,165	109,165

and becomes statistically significant. Regarding the variables in the second specification, we find that all of them continue to be significant and exhibit similar coefficient values.

One aspect underlying the previous results is that these relations could depend on the frequency at which the Information Share is measured. For instance, at a one-second interval, if stock prices are influenced by short-lived information about the market portfolio, only those capturing this type of information would obtain larger shares of information. However, once this information has been incorporated into the price, the information share could depend on other determinants. To shed light on the role that the observation frequency plays on these determinants, we estimate the information share over a 10-minute frequency. To further contrast these results, we use a rolling window of 250 daily return and compute the information share from these data. Note that these daily returns contain intraday and overnight information.

Table 6 reports the results including all covariates used when running the third specification in Table 5. To facilitate comparisons, the first column in the table contains results for the one-second frequency. The results in this table show that the stock's information share loads on differents covariates depending on the frequency over which it is measured. For the 10 minute frequency, we find that variables such as β and odd lot ratio are no longer statistically significant, while others like Trade-to-Order ratio and Cancel-to-Trade ratio are still significant but with lower economic values. When we look at the daily frequency, we observe a different picture from the one obtained with high frequency data: only size and β are statistically significant. The changes in statistical significance, together with the fact that these variables show different signs, shows that intraday information shares similar characteristics across different frequency levels, but differs from the one contained in daily returns.

Table 6: **Stock's Characteristics and Information Share for Different Frequencies.**
This table presents the results of panel regressions in which the dependent variable is the information share of a stock computed at one of three frequencies: one second, ten minutes, and one day. The variable (*Size*) represents the total market capitalization computed from closing prices (in logarithm). To compute the variables α_{CAPM} and β_{CAPM}, for a given day and stock, we estimate the equation $R^e = a + \beta_{CAPM} R^{M,e} + u_t$ using the previous 250 daily observations of stock and market excess returns. The variable R^2 represents coefficient of adjustment in this regression. The variable *Volume* is the dollar volume of the stock (in logarithm). The variable *Quoted Spread* corresponds to the average of quoted percentage spread computed from all NBBO in a given day. The variable *Absolute OIB* is the absolute difference between the number of buyer minus seller initiated trades, divided by the sum of the two. The variable *ISO Orders* is the total number of intermarket seep orders divided by the total number of trades for the stock. Odd lot volume ratio (*Odd lot ratio*) is the total volume executed in quantities smaller than 100 shares dividied by the total volume traded, the variable trade-to-order volume ratio (*Trade-to-Order ratio*) is the total volume traded divided by the total volume across all orders placed, and the cancel-to-trade ratio (*Cancel-to-Trade ratio*) corresponds to the number of full or partial cancellations divided by the number of trades. Standard errors are computed with robust errors clustered by stock and provided under the coefficient. All coefficients, except for the one associated to Quoted spread and Cancel-to-Trade ratio, are multiplied by 100. The coefficient of Cancel-to-Trade ratio is multiplied by 10000. There are a total of 109,165 observations in the panel. Statistical significance at the 1% and 5% levels is indicated by ** and *, respectively. There are 447 stocks in the cross-section. The sample period is 2012.

	1 sec		10 min		1 day	
Constant	1.259***	1.353***	1.347***	1.372***	4.348***	4.131***
	(0.286)	(0.285)	(0.224)	(0.224)	(0.787)	(0.762)
Size	-0.016	-0.015	-0.001	-0.003	-0.272***	-0.255***
	(0.017)	(0.017)	(0.014)	(0.014)	(0.047)	(0.045)
β	0.055**	0.082***	-0.013	-0.008	0.337***	0.378***
	(0.027)	(0.024)	(0.019)	(0.017)	(0.088)	(0.087)
Systematic risk	-0.007	-0.08***	0.018	0.001	0.078	-0.044
	(0.038)	(0.017)	(0.033)	(0.015)	(0.115)	(0.049)
Volume	-0.067***	-0.074***	-0.082***	-0.081***	-0.007	-0.011
	(0.005)	(0.006)	(0.004)	(0.004)	(0.008)	(0.01)
Quoted spread	-0.878***	-0.8***	-0.164***	-0.16	-0.5	-0.521
	(0.318)	(0.279)	(0.104)	(0.103)	(0.471)	(0.451)
Absolute OIB	-0.167***	-0.17***	-0.171***	-0.173***	0.034	0.037
	(0.024)	(0.025)	(0.024)	(0.024)	(0.026)	(0.026)
ISO orders	0.349***	0.335***	0.152***	0.152***	-0.027	-0.042
	(0.03)	(0.03)	(0.023)	(0.022)	(0.047)	(0.05)
Trade-to-Order ratio	-3.318***	-2.912***	-1.274***	-1.23***	-0.093	0.038
	(0.406)	(0.395)	(0.2)	(0.2)	(0.366)	(0.443)
Cancel-to-Trade ratio	0.329***	0.308***	0.069***	0.068***	0.031	0.029
	(0.064)	(0.061)	(0.021)	(0.02)	(0.028)	(0.029)
Odd lot ratio	-0.747***	-0.649***	-0.118	-0.084	-0.147	-0.149
	(0.087)	(0.088)	(0.078)	(0.078)	(0.134)	(0.13)
Month FE	YES	NO	YES	NO	YES	NO
Firm FE	YES	YES	YES	YES	YES	YES
R^2	13.41%	13.30%	4.45%	4.43%	38.72%	38.67%
Observations	109,165	109,165	109,165	109,165	109,165	109,165

5 Information Share across Frequencies

Given the previous evidence that information captured at a one-second frequency relates to some extent with that one observed at a lower intraday frequency, we study in this section the allocation of common information across stocks as a function of the sampling frequency. This characterization provides evidence about the speed with which information is incorporated in stock prices as innovations are aggregated over time. Specifically, it allows us to observe how the disproportional amount of information captured by stocks in the largest IS quintile (see Figure 1) is incorporated at lower frequencies.

We start by estimating the stock's information share for different intraday frequencies. For each frequency, we sort stocks into groups according to IS quintiles and compute the total amount of information allocated to each group. The results are reported in Panel A of Table 7. We find that the group of stocks with the highest information share always account for the largest share of information across groups. The highest proportion is observed at a one-second frequency, with about 71% of the total information impounded in the top quintile. At the 30-second frequency, this value monotonically decreases to 66.8%, and then it reaches 66.5% at the 10-minute frequency. This pattern is different from the one observed in other quintiles, for which we observe that the total information share increases as the frequency decreases. Panel B in Table 7 shows the total information share per quintile obtained with daily data. At this frequency we still observe that almost two thirds of the total information is contained in the top quintile of stocks. This result suggests that different frequencies contain distinct information, but some of the information arriving at a higher frequency is slowly impounded into lower ones.

To shed further light into the amount of information flowing between frequencies, we look at how stocks with the highest information share move across estimation frequencies.

Table 7: **Information Share for Different Sampling Frequencies.**
This table presents average values of information shares fo quintile groups formed at different frequencies. Every day and for a given frequency, we sort stocks into five groups according to their information share. Within each group, we compute the total sum of IS -in percentage- and compute the average value across the sample period for each group. In Panel A, we present results based on eight different intraday frequencies over which the value of IS is computed. In Panel B, the information share is computed based on daily returns. In this case, we use a rolling window of the previous 250 days to estimate the information share of a stock for a given day. The sample period is 248 trading days in 2012. There are 447 stocks in the cross-section.

Panel A: Intraday Sampling Frequencies

	Lowest	Quintale 2	Quintale 3	Quintale 4	Highest
1 sec	0.95	3.22	7.33	17.05	70.93
3 sec	1.04	3.41	7.62	17.46	69.94
5 sec	1.11	3.55	7.83	17.77	69.22
10 sec	1.22	3.76	8.13	18.21	68.18
30 sec	1.39	4.04	8.58	18.68	66.80
1 min	1.47	4.21	8.87	19.01	65.92
5 min	1.44	4.33	9.08	19.28	65.33
10 min	1.13	4.01	8.80	19.01	66.49

Panel B: Daily Sampling Frequency

	Lowest	Quintale 2	Quintale 3	Quintale 4	Highest
Daily	1.54	4.22	8.98	19.84	65.05

Specifically, we look at the proportion of stocks in the highest information quintile at the one-second frequency and record the groups in which they are classified at lower observation frequencies. Figure 2 shows the flows across three frequencies. From the figure, transitions from the one-second frequency into the one-minute show that the largest proportion of stocks (53.01%) is observed for stocks migrating into the highest IS quintile (red flow in the figure). The total information share contained in these stocks at a one-second frequency is 46.62%–the average IS for a stock in this group is .71% at the one-second frequency.[7] At a one-minute frequency, the average IS for a stock in the highest quintile is about .67%, so these stocks now contain $53.01 \times .67\% = 35.51\%$ of the information share at that frequency. Given that the amount of information in the highest quintile of the one-minute frequency is 65.92% (see Table 7), we conclude that more than half of this value comes from stocks with the highest information share in the one-second frequency.

Regarding the percentage of stocks that belong to the highest quintile at the ten-minute frequency, we find that about 33.1% of stocks migrated into the largest IS quintile: 20.57% came from the highest IS quintile of the 1-min frequency and the remaining from the other 4 groups (red and green flows, respectively). At a ten-minute frequency, the average IS for these stocks is .66%, so the total information share they convey is 21.85%–one third of the total value in the highest quintile at this frequency (see Table 7).

In summary, we find that stocks with the highest information share at the one-second frequency carry a significant proportion of the information share at lower frequencies within stocks that contribute the most to systematic price discovery. Nonetheless, this proportion quickly decreases beyond one-minute, and it only represents about a fifth at a 10 minute frequency. These results show that the introduction of new systematic information does not

[7] Since the number of stocks in the quintile is about 100, the total information share contained by this group of stocks is about $53.01 \times .71\% = 37.63\%$

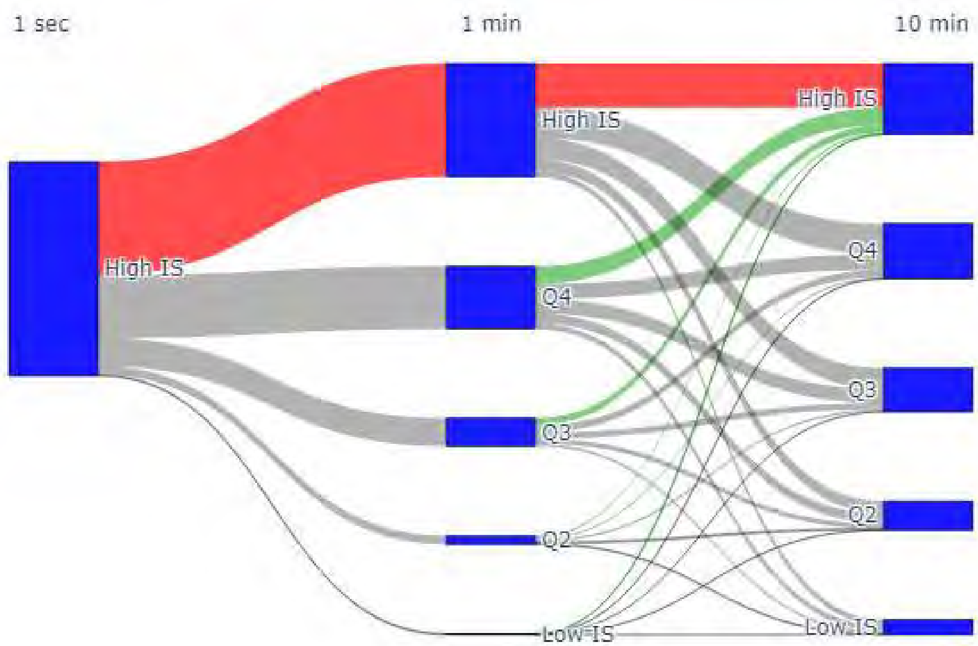

Figure 2: **Flow of Stocks from the Highest Information Quintile at the one-second frequency.** The figure plots quintile migrations of stocks in the highest information quintile at the one-second frequency. The nodes in blue represent the quintile in which stocks are grouped at a given frequency. The frequency is provided on top of the figure. A red link represents the proportion of stocks that belong to the highest quintile and transition to the the highest quintile when this group is formed at a lower frequency. A green link represents a proportion of stocks that transition into the highest quintile at the 10-minute frequency from the one-minute frequency (not including the ones transitioning from the highest quintile). Gray links represent all other transitions.

happen instantly, but

In summary, we find that stocks contributing the most to systematic price discovery at the one-second frequency carry important proportions of information share at lower frequencies. We do not pin down which part of the information in the highest frequency is shared with lower ones, but we show that an important part of the information at lower frequencies comes from stocks that impound it at higher frequencies. These results suggest that the information flow does take some time to occur and it is not immediatly assimilated by stocks.

6 Speed of Information Share Adjustment

In this section, we compare the speed of adjustment of IS quintile portfolio values to market returns. Specifically, we use Dimson (1979) regressions to determine the fraction of the information in the market portfolio returns that is present in a given quintile.

We denote by $R^n_{j,t}$ the equal-average return of stocks in the nth quintile on day j for second t. Similarly, $R^M_{j,t}$ represents the equal-average market portfolio return. For each day in our sample, we run the following regression:

$$R^n_{j,t} = a + \sum_{k=-60}^{60} b_{j,k} R^M_{j,t-k} + u_{j,t}, \quad n \in \{1, \dots, 5\}. \tag{12}$$

This specification combines lagged ($k > 0$) and lead ($k < 0$) values of the market factor to capture the total sensitivity of the quintile portfolio's return to the portfolio market. This sensitivity is computed for day j as the sum of coefficients b_k, $\beta^{All}_j = \sum_{k=-60}^{60} b_{j,k}$.

Panel A in Table 8 reports summary statistics for the total sensitivity of each quintile portfolio. We observe a monotonic increase from quintiles with the lowest information share to the highest, which is expected given the nature of the sorting variable (the stock's infor-

Table 8: **Dimson Betas.**
This table presents results for Dimson betas, which are the coefficients obtained from regressing one-second quintile portfolio returns on lead, contemporaneous, and lagged returns of market returns for a given day. This regression is given in Equation (12). Panel A shows summary statistics of the total sensitivity of quintile returns to market returns, as defined by $\beta^{All} = \sum_{k=-60}^{60} b_{j,k}$. Panel B presents average values of leading sensitivity $\beta^{Lead} = \sum_{k=-60}^{-1} b_{j,k}$, the lagging sensitivity $\beta^{Lag} = \sum_{k=1}^{60} b_{j,k}$, and the contemporaneous sensitivity $\beta^{Ctmp} = b_{j,0}$. The sample period is 2012.

Panel A: Summary Statistics of Dimson Beta

	Nobs	Median	Mean	StdDev	Skew	P10	P90
Lowest Information Share	247	0.839	0.836	0.073	-0.035	0.740	0.924
Quintile 2	247	0.961	0.958	0.052	0.055	0.890	1.021
Quintile 3	247	1.011	1.012	0.043	0.076	0.958	1.069
Quintile 4	247	1.062	1.059	0.048	-0.189	0.993	1.120
Highest Inforamtion Share	247	1.142	1.137	0.081	-0.358	1.027	1.241

Panel B: Average Values of Different Betas

	β^{All}	β^{Lead}	β^{Ctmp}	β^{Lag}
Lowest Information Share	0.836	-0.010	0.670	0.176
Quintile 2	0.958	-0.005	0.907	0.056
Quintile 3	1.012	-0.001	1.024	-0.011
Quintile 4	1.059	0.003	1.117	-0.061
Highest Inforamtion Share	1.137	0.013	1.284	-0.160

mation share). The quintile with the lowest information share has a total sensitivity of 0.83, much lower than the 1.14 observed for the highest quintile.

In addition to the total sensitivity, we also examine the leading, lagging, and contemporaneous sensitivity for each quintile portfolio. We define the lagging sensitivity as the sum of coefficients b_k of lagged variables, $\beta_j^{Lag} = \sum_{k=1}^{60} b_{j,k}$. The leading sensitivity is defined as $\beta_j^{Lead} = \sum_{k=-60}^{-1} b_{j,k}$ and the contemporaneous one corresponds to the coefficient b_0 in the regression. Panel B in Table 8 shows average values for each quintile. We observe that the contemporaneous beta is the sensitivity with the largest component in β^{All}. For all quintile portfolios, we find that leading sensitivities are small, negative for the three portfolios with lowest IS stocks, positive for the portfolios with the two highest quantiles, which suggests a modest predictability component in these two portfolios compared to the other three. Regarding the lagging sensitivity, we find that it is negative for highest three quintiles and negative for two with the lowest information share. This suggest that these two portfolios are the one with the highest sensitivity to lagged information in the market portfolio, that is, the ones adjusting more slowly to common information. In contrast, the portfolio with the highest information share has the lowest negative sensitivity to lagged information, showing that it is the fastest to impound information. These interpretations go in line with Brennan, Jegadesh, and Swaminathan (1993) who find differential of speed adjustment to common information in portfolios of firms that are followed by different number of analysts, and with Chordia and Swaminathan (2002) who also document these differences for portfolios based on volume.

To visually contrast the speed of adjustment to market information of quintile portfolios, we compute the cumulative fraction of reaction to market information realized at a given second ($\sum_{k=-60}^{d} b_{j,k}/\beta_j^{All}$). This quantity provides the fraction of information impounded at a given point in time before ($d > 0$) or after ($d < 0$) the portfolio return is observed. Figure 3 plots the cumulative fractions for the largest and lowest information quintiles. We observe that both portfolios contain no significant information that anticipates the market ($d < -1$) and that there is some level of predictability one-second ahead of the realization of the market return. The figure shows how the largest increase in information comes when the market portfolio return is observed ($d = 0$). Once this information has arrived, cumulative fractions follow different patters as lagged information for several seconds continues to be impounded in the lowest information share portfolio. On the contrary, betas for lagged market returns are negative for the portfolio with the largest information share, so the cumulative fraction of reaction decreases with the lag.

7 Conclusion

This paper provides a measure of information share for a panel of stocks that have as common component the market portfolio. We show that the information share of stocks largely differs across stocks, pointing to inefficiencies about the way common information is impounded in a cross-section of stocks. When we look at possible determinants of this inefficiencies, our results show that stocks that experience high demand pressures are more likely to have lower shares of information about the market portfolio. Examination of adjustment speeds to information shows that stocks with the lowest share are sensitive to lagged information in the market portfolio, different from what is found for the portfolio containing the stocks with higher information share levels.

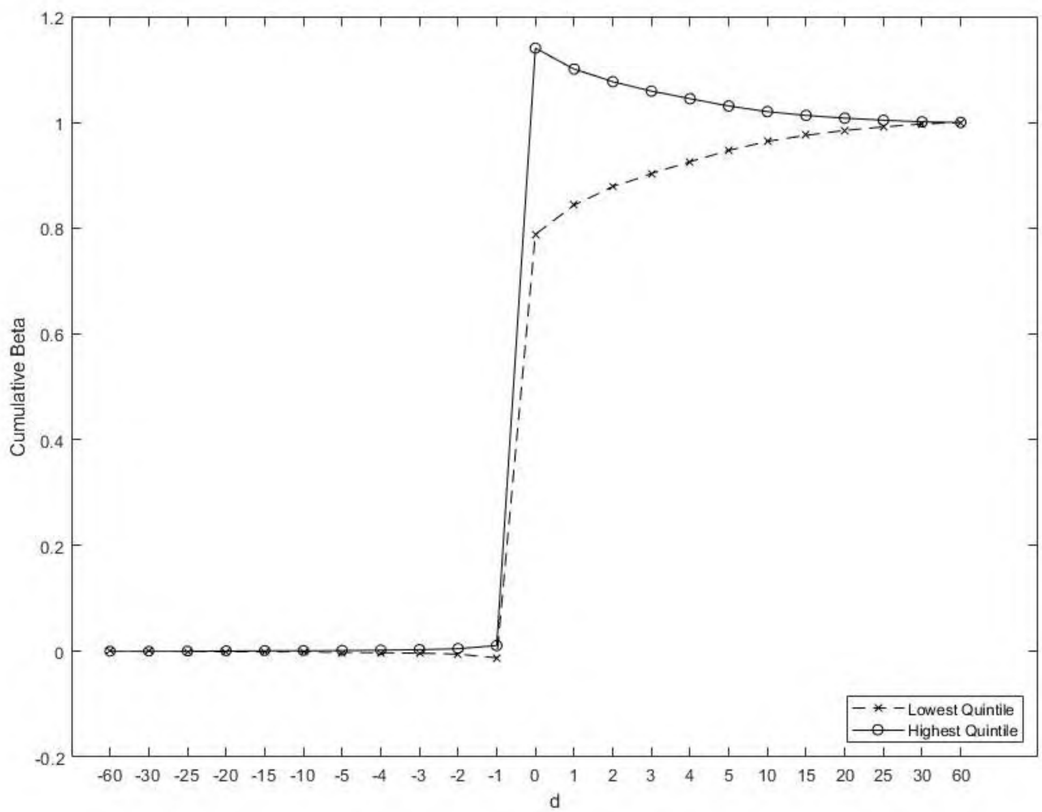

Figure 3: **Cumulative sums of coefficients in Dimson regressions.** This figure plots the average cumulative sum of coefficients in a Dimson regression for the lowest (dotted line) and the highest (continuous line) quintile portfolio returns. For each day in the sample, the one-second quintile portfolio return is regressed over several lead and lag terms of the market portfolio return. From the estimated coefficents in a given day, the cumulative beta is computed for each d in $\{-60,\ldots,60\}$ as $\sum_{k=-60}^{d} b_{j,k}/\beta_j^{All}$, where the denominator represents the sum over all lead and lag components (including the contemporaneous term). The cumulative beta presented in the figure corresponds to the average of this value over the sample period.

References

Bai, J., Philippon, T. and Savov, A., 2016. Have financial markets become more informative?. Journal of Financial Economics, 122(3), pp.625-654.

Ben-David, I., Franzoni, F. and Moussawi, R., 2018. Do ETFs increase volatility?. The Journal of Finance, 73(6), pp.2471-2535.

Bhattacharya, M., 1987. Price changes of related securities: The case of call options and stocks. Journal of Financial and Quantitative Analysis, 22(1), pp.1-15.

Brennan, M.J., Jegadeesh, N. and Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. The Review of Financial Studies, 6(4), pp.799-824.

Campbell, J.Y., Lo, A.W., and MacKinlay, A.C., 1997. The econometrics of financial markets. princeton University press.

Chakravarty, S., Jain, P., Upson, J. and Wood, R., 2012. Clean sweep: Informed trading through intermarket sweep orders. Journal of Financial and Quantitative Analysis, 47(2), pp.415-435.

Chan, K., 1992. A further analysis of the lead–lag relationship between the cash market and stock index futures market. The Review of Financial Studies, 5(1), pp.123-152.

Chordia, T. and Swaminathan, B., 2000. Trading volume and cross-autocorrelations in stock returns. The Journal of Finance, 55(2), pp.913-935.

De Jong, F. and Schotman, P.C., 2009. Price discovery in fragmented markets. Journal of Financial Econometrics, 8(1), pp.1-28.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics, 7(2), pp.197-226.

Easley, D., O'hara, M. and Srinivas, P.S., 1998. Option volume and stock prices: Evidence on where informed traders trade. The Journal of Finance, 53(2), pp.431-465.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), pp.383-417.

Harris, L., 1989. The October 1987 S&P 500 stock-futures basis. The Journal of Finance, 44(1), pp.77-99.

Hasbrouck, J. and Ho, T.S., 1987. Order arrival, quote behavior, and the return-generating process. The Journal of Finance, 42(4), pp.1035-1048.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. The journal of Finance, 50(4), pp.1175-1199.

Holden, C.W. and Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. The Journal of Finance, 69(4), pp.1747-1785.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor errorstructure. Econometrica, 74, pp.967–1012.

Westerlund, J., Reese, S. and Narayan, P., 2017. A factor analytical approach to price discovery. Oxford Bulletin of Economics and Statistics, 79(3), pp.366-394.

A Appendix, proofs and derivation of equations

A.1 Derivation of Equation (4)

The equation comes from combining equations (3) and (2) as follows:

From equation (3): $\Delta\pi_{i,t} = \Delta p_{i,t} + \Delta\epsilon_{i,t}$

From equation (2):

$$\begin{aligned}\Delta p_{i,t} &= \alpha_i\beta_i f_t + (1-\alpha_i)(p_{i,t-1} - p_{i,t-2}) = \alpha_i\beta_i f_t + (1-\alpha_i)\Delta p_{i,t-1} \\ &= \alpha_i\beta_i f_t + (1-\alpha_i)(\Delta\pi_{it-1} - \Delta\epsilon_{it-1}) = \alpha_i\beta_i f_t + (1-\alpha_i)(r_{i,t-1} - \Delta\epsilon_{it-1})\end{aligned}$$

Thus, $r_{i,t} = \Delta\pi_{i,t} = \alpha_i\beta_i f_t + (1-\alpha_i)r_{i,t-1} - (1-\alpha_i)\Delta\epsilon_{it-1} + \Delta\epsilon_{i,t}$

A.2 Derivation of Equation (6)

From equation $v_{i,t} = \alpha_i\beta_i f_t + \Delta\epsilon_{i,t}$ we can write, $v_t = (v_{1,t}, \ldots, v_{N,t})^\top$, $\Lambda = (\alpha_1\beta_1, \ldots, \alpha_N\beta_N)^\top$, and $E = (\Delta\epsilon_{1,t}, \ldots, \Delta\epsilon_{N,t})$, and the covariance matrix of of v_t can written as

$$E\left(v_t v_t^\top\right) = \Sigma_v = \sigma_f^2 \Lambda\Lambda^\top + \Sigma_E,$$

where $\Sigma_E = diag\left(\sigma^2_{\epsilon_{1,t}}, \ldots, \sigma^2_{\epsilon_{N,t}}\right)$ under the assumptions below.

We can write (5) as $f_t = b^\top v_t + e_t$,where $b = (b_1, \ldots, b_N)^\top$, so the OLS estimate of vector

b is $\Sigma_v^{-1}\Lambda\sigma_f^2$ and the R^2 coefficient of the regression in Equation (5) is

$$\begin{aligned} R^2 &= \frac{b^\top \Sigma_v b}{\sigma_f^2} \\ &= b^\top \Lambda = \sum_{i=1}^{N} b_i \alpha_i \beta_i \end{aligned}$$

A.3 Derivation of Equation (7)

Westerlund et al. (2017) show in their Appendix A that equation (6) can be expressed as:

$$IS_i = \frac{(\alpha_i \beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{1 + \sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}$$

The sum of the IS_i measures for N companies does not equal one. However, asymptotically, as $N \to \infty$, the it will converge to 1, as show below:

$$\begin{aligned} IS_i &= \frac{(\alpha_i \beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{1 + \sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}} \\ &= \frac{(\alpha_i \beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{\frac{\sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{\sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}} + \sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}} \\ &= \frac{(\alpha_i \beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{\sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2} \left\{ \left[\sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2} \right]^{-1} + 1 \right\}} \end{aligned}$$

As $N \to \infty$, then $\left[\sum_{j=1}^{N} (\alpha_j \beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}\right]^{-1} \to 0$, and:

$$IS_i = \frac{(\alpha_i\beta_i)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}{\sum_{j=1}^N (\alpha_j\beta_j)^2 \frac{\sigma_f^2}{\sigma_{\epsilon_i}^2}}$$

that will clearly add up to one.

A.4 Consistency of $\widehat{IS_i}$

Our consistency proof is based in the recent contribution of Westerlund et al. (2017). They propose a similar information share measure for price discovery of one single asset trading in many markets. In our case our measure is designed for a one market and many assets. We show that under the assumptions below, our measure of information share has the same asymptotic characteristics as the one proposed by Westerlund et al. (2017), thus their consistency proof applies to our measure.

Our information share estimation is based on consistent estimation the parameters of the Factor model in (4)

$$\tilde{r}_{i,t} = \lambda_i f_t + \Delta\epsilon_{i,t}, \tag{13}$$

In standard factor model notation $r_{i,t}$ is the response variable, λ_i is factor loadings specific to the cross-sectional unit i, f_t is the common factor and $\Delta\epsilon_{i,t}$is the idiosyncratic error. The principal components method of (Bai and N, 2002; Bai 2003) can be used to to consistently estimate factors, and loadings, however we follow Westerlund et al. (2017) and use Pesaran (2006) and estimate the factor by the cross sectional average of the response variable. Specifically our estimator of the factor is $\hat{f}_t = N^{-1}\sum_{i=1}^N \tilde{r}_{i.t}$ and then the estimator of the loadings is the time series OLS estimator $\hat{\lambda}_i$ in the individual regressions over unit i

in equation (4) using $\hat{f}_t$. Using $\hat{\lambda}_i$ and $\hat{f}_t$ we obtain

$$\Delta\hat{\epsilon}_{i,t} = \tilde{r}_{i,t} - \hat{\lambda}_i \hat{f}_t. \quad (14)$$

Finally we compute $\hat{\sigma}_f^2 = T^{-1}\sum_{t=2}^{T} \hat{f}_t^2$ and $\hat{\sigma}_{\epsilon_i}^2 = T^{-1}\sum_{t=2}^{T} \Delta\hat{\epsilon}_{i,t}^2$

Assumptions for consistency of factor model parameters and $\widehat{IS}_i$ are as follows

A1 $E\|f_t\|^4 \leq M \leq \infty$ and $E(f_t^2) = \sigma_f^2 > 0$. $E(f_t/\mathcal{F}_{t-1}) = 0$ where $\mathcal{F}_t$ is the sigma-field generated by $\{f_j\}_{n=1}^{t}$

A2 The loading λ_i is deterministic or stochastic with $E\|\lambda_i\|^4 \leq M \leq \infty$ and $\frac{1}{N}\sum_{n=1}^{N} \lambda_i \neq 0$, for all N even if $N \rightarrow \infty$.

A3 $\Delta\epsilon_{i,t}$ is weakly correlated over time and cross-sectionally independent.

A4 λ_i, f_t, and $\Delta\epsilon_{i,t}$ are mutually independent.

According to the first assumption A1, the factor is serially uncorrelated, however this is only a simplifying assumption and it can be relaxed. A2 rules out non significant factors, or factors with trivial contribution to the variance of the response variables. Factor loadings can be assumed to be non-random variables. A3 allows for limited time series dependence in the idiosyncratic components and also heteroskedasticity. A4 is a standard assumption in factor analysis models.

Our information share estimator is constructed based on $v_{i,t} = \alpha_i\beta_i f_t + \Delta\epsilon_{i,t} = \lambda_i f_t + \Delta\epsilon_{i,t}$ that reflects the total sum of common and idiosyncratic contemporaneous innovations to observed stock returns from equation (4). Similarly, Westerlund et al. (2017) information share measure is based on $v_{i,t} = \lambda_i\eta_t + U_{i,t}$ that accounts for the sum of the shocks coming from the efficient price and market microstructure components of the model. As explained before their construction differs to ours in that their information share measures price discovery for one asset in different markets, and ours measures how different stock incorporate market information. So the parameters of the two models represent very different markets elements, specially our f_t represent the market factor common to all stocks but in their model η_t represents the efficient price common to all markets. Moreover the total shocks $v_{i,t}$ in our case represent shocks to stock returns and in theirs to stock prices. However from the technical perspective both equations represent common factor models, thus assumptions and estimation techniques are basically the same. Thus are assumptions described above are very similar to Assumption 1, (i) to (v) in Westerlund et al. (2017). Only one difference is worth mentioning is that we do not require assumption iii) in Westerlund et al. (2017) since our factor model is directly defined in price changes. If we substitute $\hat{\sigma}^2_f$ for $\hat{\sigma}^2_\eta$ and $\hat{\sigma}^2_{\epsilon_i}$for $\hat{\sigma}^2_{U_i}$ both information share measures are identical then they share the same asymptotic characteristics. As show in Westerlund et al. (2017) Proposition 1: $N|\widehat{IS}_i - IS_i| = o_p(1)$, thus $\widehat{IS_i}$it is consistent.