INSEAD The Business School for the World®

Working Paper

2024/26/FIN

Data Specialists and Market Efficiency

Massimo Massa
INSEAD, massimo.massa@insead.edu

Hong Zhang
Singapore Management University, hongzhang@smu.edu.sg

Yijun Zhou
Baruch College, City University of New York, yijun.zhou@baruch.cuny.edu

In the age of big data, investors need to process increasingly complicated, multidimensional data to decipher different aspects of a firm. How do investors deal with such multidimensional data? We find more informed institutional investors tend to specialize in subsets of firm aspects (i.e., data specialists). Such data specialization, however, may hamper market efficiency. Inattention shocks to specialists hinder price efficiency in their specialized aspects of firms; other aspects of firms may also be negatively influenced due to strategic complementarity. Specialist inattention also significantly impacts anomaly returns, impeding the price corrective effect of news arrival.

Keywords: Big Data; Institutional Investors; Market Efficiency; Data Specialists; Strategic Complementarity; Anomalies
JEL Codes: G14; G23

Electronic copy available at: https://ssrn.com/abstract=4739691

We thank Ramin Baghai, Jennie Bai, Ben Charoenwong, Darrell Duffie, Paul Ehling, Xavier Giroud, Bin Han, Pierre Hillion, Gerard Hoberg, Marcin Kacperczyk, Kai Li, Ye Li, Pedro Matos, Roni Michaely, Seth Pruitt, Adam Reed, Martin Schmalz, Lukas Schmid, Kelly Shue, Clemens Sialm, Mathew Spiegel, Laura Starks, Yuehua Tang, Laura Veldkamp, Gyuri Venter, and Liyan Yang for their helpful comments.

Working Paper is the author's intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu

Find more INSEAD papers at https://www.insead.edu/faculty-research/research

Copyright © 2024 INSEAD

1. Introduction

The notion of market efficiency in modern financial theories relies on the efficacy of sophisticated investors to process and trade firm-level information. As firms become more complicated in a modern economy, multidimensional data is often needed to describe different aspects of the same firm (e.g., Martin and Nagel 2022). Indeed, high dimensionality is a leading property of big data (Goldstein, Spatt, and Ye 2021). This data structure gives rise to new fundamental questions: Do sophisticated investors prefer to grasp all aspects of a firm by processing the entirety of data, or do they focus only on subsets of available data? Could market efficiency subsequently be affected?

To see the importance of these questions, we can draw on intuitions from an analogous case where multiple investors trade multiple assets. Traditional theories (e.g., the CAPM) posit that *all* investors behave similarly when trading *all* assets. However, since Merton (1987) and Grossman and Miller (1988), researchers have recognized that real-world investors trade subsets of assets, and that such *asset specialization* impedes market efficiency. In the context of big data, two comparable scenarios can arise when multiple investors process multidimensional data for the same firm: either *all* sophisticated investors may similarly process *all* data to decipher *all* aspects of the firm, or they may only process subsets of data to *specialize* in some aspects of the firm. Both scenarios are reasonable and align with the theoretical assumptions of, respectively, Admati and Pfleiderer (1987) and Goldstein and Yang (2015). Yet, as the two studies show, the two scenarios imply very different market efficiency. Despite the theoretical importance of these scenarios in markets with multidimensional big data, empirical tests for them are surprisingly scarce.

Our paper aims to fill this gap by examining how real-world sophisticated investors process multidimensional data and how market efficiency could subsequently be affected. To achieve this goal, we utilize the news sentiment data provided by *RavenPack* and link it to the daily trading activities of institutional investors from *ANcerno* for the period 2000–2010. *RavenPack* is suitable for our purpose, because it represents one important type of big data, as indicated by a recent JP Morgan report (Kolanovic and Krishnamachari 2017).[1] Importantly, *RavenPack* also captures the multidimensional structure of data by classifying each firm news into a comprehensive list of *News Categories*. For instance, "Dividends" and "Labor Issues" represent two such news categories. News released by a firm under these two categories can be regarded as two distinct dimensions of the firm data that investors need to process if they want to decipher the related aspects of the firm.

This feature allows us to construct a novel measure of data specialization based on the degree of news-related trading concentration across news categories. For instance, if an investor concentrates all her trading in the category of "Dividends" (and no other categories), her data processing specializes in the dividend-related aspects of the firm. Empirically, data specialization is heterogeneous across investors and persistent for given investors. Hence, we can refer to investors with the lowest (highest) degree of specialization as *data generalists* (*data specialists*).

We take several steps to examine the economics of data specialization and its implication for market efficiency. Firstly, we link data specialization to performance to test the two scenarios mentioned earlier. From a theoretical standpoint, data generalists and specialists represent the prototype information-processing investors in Admati and Pfleiderer (1987) and Goldstein and Yang (2015). Assuming that data processing skills are scalable across categories, more capable investors should process data in more news categories, giving rise to the *informed generalist hypothesis* à la Admati and Pfleiderer (1987). This hypothesis suggests that data generalists

[1] According to the report, *RavenPack* pioneered the era of big data by analyzing unstructured datasets, which include premium newswires, regulatory news providers, press releases and over 19,000 web publications, to provide structured sentiment data. The source, processing, and usage of the *RavenPack* data fit well with the prominent characteristics of big data, as summarized by the National Institute of Standards and Technology (NIST) as "volume, velocity, variety, and/or variability—that require a scalable architecture for efficient storage, manipulation, and analysis" (NIST 2019). For more information on big data, see Farboodi and Veldkamp (2023) for a recent survey.

outperform specialists because they are better informed and have broader knowledge across categories. In contrast, if there are substantial frictions in processing cross-category data, capable investors may optimally focus on a subset of categories.[2] This *informed specialist hypothesis* (à la Goldstein and Yang 2015) predicts that data specialists will outperform instead.

We conduct two tests to examine these two competing hypotheses. The first test, conducted at the investor level, examines portfolio performance across specialization quintiles. We observe a positive relationship between specialization and trading performance, suggesting that specialists are generally more informed. A long-short strategy of buying (selling) the portfolios of specialists (generalists) can generate a highly significant annualized return of 8.3% or a Daniel, Grinblatt, Titman, and Wermers (1997; hereafter, DGTW) adjusted abnormal return of 6.17%. The second test is at the stock level. Motivated by the informed trading literature (e.g., Kyle 1985), we use the demand of investors to proxy for the value of their processed information. We find that stocks deliver significant out-of-sample DGTW-adjusted returns when the demand *difference* between specialists and generalists is high, confirming that specialist demand is more informative. Using the daily trades initiated by specialists and generalists yields similar conclusions.

Jointly, these results support the *informed specialist hypothesis*. Additionally, a typical data specialist tends to concentrate on two to three news categories, indicating that data processing is quite specialized in our sample. It is worth noting that asset specialization—i.e., trading on subsets of assets following Merton (1987) and Grossman and Miller (1988)—has been shown to benefit fund managers and analysts (e.g., Kacperczyk, Sialm, and Zheng 2005; Sonney 2007) and help explain the home bias of investors (Van Nieuwerburgh and Veldkamp 2009). Our novelty is to show that, holding the same asset (e.g., a firm) fixed, more informed investors tend to trade only

[2] Institutional investors such as mutual funds can, of course, hire multiple portfolio managers when individual managers are constrained. This organizational structure, however, faces coordination costs as a different type of friction.

a subset of available data on the asset. This same-asset data specialization differs from, yet complements, asset specialization. As a result, existing measures on asset specialization fail to explain our results (we will address alternative explanations shortly).

Our second-step analysis aims to explore the implications for market efficiency. The literature suggests two important economic channels through which data specialists may affect market efficiency. First, a temporary shortage of specialist attention may directly hinder market informativeness about the aspect(s) of firms that these inattentive specialists usually process. This occurs because finding a close substitute for data specialists is difficult. Hence, data specialists' inattention shocks reduce the informativeness of the *entire* market about the corresponding aspect(s) of firms, even when the inattention shock concerns only a small fraction of informed investors. We refer to this effect as the *specialist inattention channel*.

A testable implication of this economic channel is that it will take longer for the market to incorporate newly arrived information into stock prices in a specific category faced with specialist inattention. To measure specialist inattention, we follow the intuition of Kempf, Manconi, and Spalt (2016) that "extreme" returns of some invested stocks may distract specialists from processing data for *other* stocks in their holding portfolios.[3] We then examine how market efficiency—proxied by the speed of price reaction to the sentiment of newly arrived news—is affected by specialist inattention.

In the absence of specialist inattention, the market exhibits a general underreaction pattern, as it takes time for the market to incorporate new information (e.g., Della Vigna and Pollet 2009; Hirshleifer, Lim, and Teoh 2009). Our new finding is that specialist inattention further hampers

[3] The intuition is that extreme returns are attention-grabbing shocks (Barber and Odean 2008). More recent studies on salience theory (Bordalo, Gennaioli, and Shleifer 2012; 2013; 2020) also suggest that extreme stock returns relative to the market are salient and that investors overweight (underweight) their attention to salient (nonsalient) attributes.

the market's effectiveness in incorporating information. When specialist inattention is high on the news day, the market price adjusts less. Accompanying this news-day effect is an enlarged postnews drift for these stocks: the postnews drift from t+1 to t+20, for instance, becomes 37.5% higher when specialist inattention increases by one standard deviation. After 20 days, the market price converges to its fully adjusted level, indicating that the market eventually "catches up".

Since the inattention shocks are plausibly exogenous (e.g., Kempf, Manconi, and Spalt 2016), the above results identify a causal impact of data specialists on market efficiency. In contrast, generalist inattention does not have significant influences, consistent with the notion that they are relatively less informed and can be more easily substituted for. This placebo test highlights the unique role data specialists play in affecting market efficiency.

Moreover, the influence of specialist inattention may extend beyond the focal aspects of firms, due to a second *strategic complementarity channel*. Broadly speaking, the information uncertainty faced by specialists in one news category can decrease (increase) when specialists in other categories process more (less) information about other aspects of the same firm—a phenomenon referred to as the "uncertainty reduction effect" in Goldstein and Yang (2015). In this case, a specialist's incentive to process data in her area of specialization can be reduced when other specialists jointly process *less* data for other aspects of the same firm, giving rise to strategic complementarity across different specialists.[4]

[4] The theoretical literature also proposes several other economic mechanisms that may give rise to complementarity in information acquisition. These channels include the cost structure of information (Veldkamp 2006), coordination among investors (Hellwig and Veldkamp 2009), the existence of multiple information sources (Ganguli and Yang, 2009), the concern of relative wealth (Garcia and Strobl 2011), agency problems (Huang 2015), the stochastic supply of risky assets (Avdis 2016), the link between the stock price and real investments (Dow, Goldstein, and Guembel 2017), leverage constraints (Glebkin, Gondhi, Kuong, 2021), and the mutual learning between financial markets and the real economy (Benhabib, Liu, and Wang, 2019). We focus on the uncertainty reduction effect because of its importance when multiple investors process multiple dimensions of information; we leave the empirical analysis of other channels to future research.

To investigate this second channel, we link specialists covering a specific category (i.e., “focal specialists”) to those specialists processing information in news categories *other than* the original focal category (i.e., “other specialists”). We focus on the case in which the attention of other specialists—but not that of focal specialists—becomes exogenously distracted. For instance, for the focal category of “*Dividends*,” we examine whether the inattention shock of other specialists who do not trade on dividend news may weaken the market’s efficiency on dividend news even when dividend specialists’ attention remains unchanged. This cross-category effect can arise when the strategic complementarity channel is in play.

Our empirical tests lend strong support to this channel. The speed at which stock prices incorporate new information in the focal category is significantly reduced by other specialists’ inattention. A one-standard-deviation increase in other specialists’ inattention is associated with an 11% lower market reaction on the news day. The delayed price adjustment continues for three days following the news day and converges afterward. These results confirm that even when the potential capacity of focal specialists remains unchanged in the focal news category, their efficiency can nonetheless be temporarily reduced by other specialists’ inattention.

An alternative explanation for this cross-category effect could be that focal specialists dynamically shift their attention away from their focal category to fill the attention gap created when other specialists are distracted. This explanation is conceptually unappealing because informed data specialists are difficult to substitute for. Nonetheless, we can empirically scrutinize this alternative by providing a diagnostic analysis of trading activities. We find that focal specialists trade less in their focal news category when inattention among other specialists is high, in line with reduced incentives. However, focal specialists also reduce their trading elsewhere (albeit insignificantly), which contradicts the notion of attention shifting. Jointly, our results support the *strategic complementarity channel* in impeding market efficiency.

Thus far, our channel analysis has suggested that data specialization may hinder the market's ability to accurately price various aspects of firms. Given that firm aspects are closely related to firm characteristics and that mispricing of the latter is known to result in characteristic-based anomalies (see, among others, McLean and Pontiff 2016; Harvey, Liu, and Zhu 2016; Engelberg, McLean, and Pontiff 2018; Hou, Xue, and Zhang 2020), it stands to reason that data specialization might also affect anomaly returns.

Our next step of analysis aims to investigate this important asset pricing implication. We build on the observation of Engelberg, McLean, and Pontiff (2018) that anomaly returns are driven by biased expectations and that news arrivals can help correct such biased expectations. From this perspective, if investors mistakenly underestimate the expected value of a firm aspect, such as dividends, the market will correct this mistake by generating positive stock returns when new data indicates favorable dividend policies. However, our earlier analysis suggests that if dividend specialists experience inattention shocks, this correction effect will be impeded. This *reduced-correction hypothesis* illustrates the influence of data specialization on anomalies.

To empirically investigate this hypothesis, we construct a stock-level anomaly proxy by summing the number of long-side and short-side anomaly portfolios to which a firm belongs, following the method of Engelberg, McLean, and Pontiff (2018). This summation employs a long-minus-short strategy based on a list of 212 characteristics. Subsequently, we link a stock's news-day returns to this anomaly proxy, as well as its interactions with specialist inattention. The main findings are two-fold. First, we observe that anomalies are associated with significantly positive returns on news days, confirming the correction effect noted by Engelberg, McLean, and Pontiff

(2018). A one-standard-deviation increase in anomaly is associated with a 10 basis points higher daily return on news days. This magnitude is on par with the estimation made by these authors.

Second and unique in our analysis, inattention shocks of data specialists significantly hinder the correction effect. A one-standard-deviation increase in inattention shocks is associated with an approximately 20% reduction in price correction. This result not only remains robust across various empirical specifications but also provides insights into the economic foundations of anomalies. In principle, anomalies can arise from three economic sources: mispricing, data mining, and systematic risk. Given that inattention shocks are specialist-specific, they arguably don't pertain to either systematic risk or *ex-post* data mining. This leaves mispricing the most plausible explanation for the portion of anomaly returns affected by specialist shocks, reinforcing Engelberg, McLean, and Pontiff's (2018) mispricing interpretation of anomaly returns. These results suggest that data specialization may have profound implications for asset pricing.

As the final step of our analysis, we conduct a battery of tests to shed light on the economics of data specialization. We start with alternative explanations for data specialist's outperformance. First, fund size is known to affect both operational costs and fund returns due to the diseconomies of scale (Berk and Green 2004). As this effect may also influence the cost structure of data processing, a spurious relationship between the fund return and data specialization may arise as a result. Second, active trading strategies (e.g., Cremers and Petajisto 2009; Kacperczyk, Sialm, and Zheng 2008; Amihud and Goyenko 2013) may coincide with the effect of data specialization. Last but not least, known proxies of information utilization and asset specialization, particularly the reliance on public news (Kacperczyk and Seru 2007) and industry concentration (Kacperczyk, Sialm, and Zheng 2005), may be related to data specialization.

We take a two-stage approach to parsimoniously study these alternative explanations. In the first stage, we examine how data specialization is influenced by the aforementioned effects by regressing the former on a list of contemporaneous proxies for the latter. The proxies include portfolio size, quarterly trading volume, quarter-end holding rebalance, the number of days with news-related trading within the quarter, the trading reliance on news, the industry concentration, the R squared measure of Amihud and Goyenko (2013), and the active share using S&P as a benchmark (Cremers and Petajisto 2009). Among these proxies, we use quarterly dollar trading volume and quarter-end percentage holding rebalance to capture the interim active trading of Kacperczyk, Sialm, and Zheng (2008). Additionally, we include the number of within-quarter news-trading days to capture the exposure of (in)active strategies to news days. We find that these alternative channels jointly explain only 24.7% of data specialization, and the explanatory power stems primarily from the reliance on public news. Since such a reliance signals uninformed trading (Kacperczyk and Seru 2007), it is unlikely to account for the outperformance of data specialists.

In the second stage, we use *characteristics-adjusted specialization*, proxied by the residuals of the first-stage regression, to revisit our baseline results, to control for the influences of alternative channels. Our main conclusions remain highly robust. For instance, the specialist-minus-generalist portfolio delivers a highly significant annualized return of 5.67%.[5] Characteristics-adjusted specialization also impedes market efficiency via both the inattention and strategic complementarity channels. These results suggest that the effect of data specialization is unlikely a byproduct of asset specialization or known trading strategies of informed investors.

[5] To further investigate the influence of news reliance and industry concentration in our setup, we also double-sort investors into 25 groups by our data specialization measure and one of the two alternative measures. Unreported results confirm that each captures independent characteristics of institutional investors.

To better understand the skill sets and incentives of data specialists, we investigate how they deal with market-wide information. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014) point out that investors may rationally shift attention from firm-specific news to market-wide information in bear markets. To test this implication, we adopt Daniel and Moskowitz's (2016) approach to examine investors' market timing intensity in bull and bear markets—a successful bear market timing reveals a shift in attention from firm-specific data to market-wide information in bear markets. Our results show that generalists shift their attention in this manner, consistent with Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014). In contrast, more informed specialists do not reallocate attention, reaffirming their concentrated focus on a subset of available data. This behavior is consistent with the model of Van Nieuwerburgh, and Veldkamp (2009), in which investors benefit from specializing in areas where they have some initial advantage.[6]

Lastly, we employ alternative news categories (by combining clustered news categories) or specialization measures (based on the dollar value of news-related trading) to alter the empirical methods of inferring data specialization. Across all these different specifications, our results on performance and market efficiency remain highly robust, confirming that data specialization is predominantly the strategy of more informed investors and that data specialists can affect market efficiency due to their segmented information processing.

Our study is related to several strands of the literature. A nascent literature explores the financial implications of big data (e.g., Begenau, Farboodi, and Veldkamp 2018; Dugast and

---

[6] Zambrana and Zapatero (2021) document that asset managers who invest in single and multiple investment objectives exhibit better selection and market timing ability, respectively. Although their specialization focuses more on the asset side, their results are broadly consistent with our observation that specialists tend to process a subset of firm-specific information. Despite the difference between asset and data specialization, informed investors benefit from specialization. This similarity suggests that the two effects may share a similar economic ground. For instance, Van Nieuwerburgh, and Veldkamp (2009) indicate that when local investors have a small initial information advantage on local assets, they benefit from learning more about local assets—i.e., specializing in what investors already know is a more profitable strategy. The same intuition may apply to the processing of multi-dimensional information: investors who have a small initial advantage in processing a particular dimension of news may find it beneficial to specialize in this dimension.

Foucault 2018; Farboodi and Veldkamp 2020; Jones and Tonetti 2020; Balasubramanian and Yang 2019; Martin and Nagel 2022; Farboodi et al. 2022; Farboodi and Veldkamp 2023 provide a recent survey). In particular, the high dimensionality of big data often poses challenges to traditional financial tools (see, e.g., the editorial comments from Goldstein, Spatt, and Ye 2021). We investigate the emerging big-data question of how multiple investors process multidimensional data. In the classical setting where multiple investors trade multiple assets, it is often observed that investors deviate from the representative-investor paradigm by specializing in subsets of assets (e.g., Merton 1987; Grossman and Miller 1988). Our novelty is to show that an analogous specialization in processing multidimensional data may exist, which could potentially impede market efficiency. This specialization and its impact may provide a heuristic when data becomes increasingly complicated. It is also consistent with leading implications of big data—e.g., investors value the same data differently (Farboodi et al. 2023).

Our findings are closely related to the seminal work of Goldstein and Yang (2015), which lays the theoretical foundation for understanding multidimensional data. They show that market inefficiency can arise when different traders are informed on subsets of fundamentals of the same security, in contrast to Admati and Pfleiderer (1987).[7] To the best of our knowledge, we are the first to provide evidence on such specialization and its strategic complementarity. Our findings have important normative implications. For instance, while the market is known to underreact to news when its investors are distracted, we show that the distraction of a particular group of data specialists—who may represent only a small fraction of informed investors in the market—may suffice to reduce the informativeness of the *entire* market in their specialized fields. The market is thus segmented for different aspects of firms.

[7] Lundholm (1991), Paul (1993) and Lee (2010) discuss more implications of representative vs. specialized information processing.

Expanding on this perspective, data processing frictions on different aspects of firms may also be related to characteristic-based multidimensional anomalies. Engelberg, McLean, and Pontiff (2018) point out that anomaly returns may be driven by biased expectations, which are subsequently corrected by news arrivals. We further hypothesize and empirically confirm that data specialist inattention impedes this correction. Our results not only demonstrate the relevance of data specialization in asset pricing but also contribute to the ongoing debate about the economic ground of anomalies (e.g., McLean and Pontiff 2016; Harvey, Liu, and Zhu 2016; Hou, Xue, and Zhang 2020; Jensen, Kelly, and Pedersen 2023), leaning toward mispricing.

Our results are also related to studies examining the trading and attention of institutional investors. Vast evidence shows that institutional investors are skillful in processing firm-level information (e.g., Chemmanur, He, and Hu 2009; Puckett and Yan 2011; Henry and Koski 2017; Huang, Tan, and Wermers 2020).[8] Such skill allows their attention to affect asset prices (Ben-Rephael, Da, and Israelsen 2017; Chen et al. 2018) and corporate actions (Kempf, Manconi, and Spalt 2016; Gilje, Gormley, and Levit 2020).[9] We contribute by zooming in on how institutional investors penetrate the black box of firm-specific information, which provides a novel perspective to understand market efficiency.

The remainder of the paper is organized as follows: Section II presents the data that we employ and the main variables constructed for the analysis. Section III describes the empirical findings regarding the informativeness of generalists vs. specialists. The implications for market efficiency are discussed in Section IV and the implications for anomaly returns are detailed in Section V.

[8] Existing asset pricing studies typically focus on certain categories of news to infer the skills of institutional investors, such as cash flow news around earnings announcements and analyst downgrades (e.g., Christophe, Ferri, and Angel 2004; Christophe, Ferri, and Hsieh 2010).

[9] Barber and Odean (2008) discuss the attention difference between retail and institutional investors. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014, 2016) examine how institutional investors shift attention in different states of the market.

Section VI provides additional tests and robustness checks, and a brief conclusion follows in Section VII.

2. Data and Construction of the Variables

We obtain our data from several sources. We start with the news sentiment data provided by *RavenPack*. As discussed in Kolanovic and Krishnamachari (2017), *RavenPack* provides an interesting example of big data because it offers structured sentiment data from analyzing unstructured datasets such as premium newswires, regulatory news providers, press releases, and over 19,000 web publications. This sentiment data aligns with NIST's (2019) definition of big data, which is characterized by "volume, velocity, variety, and/or variability—that require a scalable architecture for efficient storage, manipulation, and analysis". We focus on the firm-level news processed by *RavenPack*, which started in 2000. For our purpose, *RavenPack* also captures the multidimensional structure of data by classifying firm news into a comprehensive list of *News Categories*.[10] Therefore, it provides an ideal ground to study questions related to the multidimensions of data.[11] Hence, we retrieve the standardized sentiment index, labeled Event Sentiment Score *(ESS)*, together with its category information from *RavenPack*.

To investigate how investors process the multidimensional data of the same firm, we next utilize the *Ancerno* database of institutional investors (see Puckett and Yan, 2011, for a detailed description of the data), from which we keep track of the daily trades of each institutional investor in each stock from 1999 to 2010. Lastly, we collect data on stock prices and the accounting

[10] *Ravenpack* identifies firms mentioned in the news and assigns a relevance score ranging from 0 to 100 to indicate how strongly the firm is related to the underlying news. Then *Ravenpack* classifies each piece of news to one (and only one) news category. This classification occurs primarily when the news demonstrates a significant relevance to the firm (with an average relevance score of 94.9). Our analysis focuses on the sample of firm-level news with assigned *Ravenpack* category.

[11] The use of *RavenPack* categorization alleviates the concern of data mining related to category construction by researchers. Later sections will also show that modifying *RavenPack* categorization (by combining clustered categories) will not change our results.

information of each listed company in the U.S. from the Center for Research in Security Prices (*CRSP*) and *COMPUSTAT*.

We then match institutional investors' daily trades on all stocks from *Ancerno* with *RavenPack* as follows. From the *Ancerno* database, we first aggregate the daily trading (including multiple buy or sell transactions) by an investor in a stock to a single trade. We then label a trade *news-related* if there is news reported for the stock (firm) either on the day of the trade or the day before.[12] The *RavenPack* categorization of firm-specific news enables us to classify each news-related trade of an institutional investor into news categories. We focus on the top 15 news categories, which helps us understand whether an investor bases her trading on firm-specific information from a limited set of news categories or a more extensive range.[13] The list of the news categories and the frequency of each news category are reported in Table 1, with Column 4 further presenting the number of investor-quarter with trades in each news category.

Next, we construct a *news category concentration index* for each investor each quarter to capture how each investor allocates her trading across different *news categories* of firm-specific information. For a given investor i in quarter q, we calculate her percentage of news-related trading in news category c as her number of news-related trades in that category divided by her total number of news-related trades across all news categories in that quarter. We can construct her news category concentration index as the sum of the squared percentages of news-related trades in each news category (i.e., the Herfindahl index):

[12] We also account for news from the day before the trade, recognizing that news might be released after trading hours and consequently relate to the subsequent day's trading.
[13] We group the remaining smaller categories into a distinct category ("Other categories"). Our results remain unchanged if we discard this additional category.

$$News\ category\ concentration\ index_{i,q} = \sum_{c=1}^{N} \left(\frac{\#\ of\ news\ trading_{c,i,q}}{\#\ of\ news\ trading_{i,q}} \right)^2, \tag{1}$$

The news category concentration index measures how each investor specializes her trading in the multidimensional data of firms, proxied by the multidimensional news categories. A high index value reveals a higher degree of data specialization.

For each quarter, we sort investors into five quintiles based on the concentration index, ranging from the lowest to the highest. Table 2 Panel A reports the summary statistics of the news category concentration index for each quintile. We can observe large heterogeneities in the trading behavior among different traders in terms of their specialization in news categories. The average concentration index of investors in quintile 5 is 44.90%, while that of investors in quintile 1 is 13.32%. Roughly speaking, these values are equivalent to quintile 5 (quintile 1) investors processing information in two to three (seven to eight) news categories if they allocate trading equally across categories. We refer to traders in the highest (lowest) concentration quintile as *data specialists* (*generalists*).

Given the large heterogeneity among institutional investors, it is important to understand how persistent their trading behavior is. Therefore, we investigate the transition rates of investors across concentration quintiles over time. As we can see from the transition matrix presented in Panel B, the diagonal elements are much larger than the off-diagonal ones, suggesting that this classification is quite persistent. For instance, an investor classified as a data generalist in one period (quarter) has a probability of 49.27% of being classified as a generalist in the next period. This probability is approximately ten-fold greater than the probability that the same investor switches to a specialist in the next quarter (which is 4.95%). Likewise, an investor classified as a data specialist in one period has a probability of 48.74% (5.25%) of being classified as a data specialist (generalist) in the next quarter. Since the degree of data specialization is both prevalent and persistent, we can treat it as a fundamental characteristic of institutional investors to investigate its performance and efficiency implications.

Panel C presents the trading characteristics of the investors across each quintile. No clear trends emerge in terms of monthly trading dollars and monthly turnover when considering data specialization, indicating that both data specialists and generalists likely operate under similar funding or liquidity conditions. Despite these similarities, the informativeness of their trading may differ.[14] To investigate this possibility, we proceed to explore the performance implications of data specialization.

3. Informativeness of Specialists vs. Generalists

3.1. Portfolio-level performance of specialists vs. generalists

This section differentiates between the *informed generalist* and the *informed specialist hypotheses* by linking data specialization to performance. Specifically, the *informed generalist hypothesis* posits that information-processing skills are scalable across different categories. In this case, more capable investors are expected to cover more news categories, giving rise to the predominance of informed generalists (Admati and Pfleiderer 1987) and a *negative* relationship between data specialization and performance.

In contrast, the *informed specialist hypothesis* posits that skilled investors face severe barriers or costs in processing multidimensional firm-specific information. In the presence of costly cross-

[14] Although their total dollar trading volumes are similar, our Online Appendix (Table A1) shows that specialists typically allocate a greater portion of total trading dollars into news-related trades compared to generalists, a pattern consistent across all news categories. Considering the informed trading literature (e.g., Kyle 1985), which suggests that an informed investor's total demand increases with the value of her signals, this discrepancy may imply that data specialists process superior information.

category information processing, skilled investors choose the optimal number of categories in which to process information by equalizing the marginal benefits and marginal costs of expanding to another category. When the barriers and costs are very high, skilled investors specialize their information processing and trading in a very limited number of categories, giving rise to the predominance of informed specialists (Goldstein and Yang 2015) and a *positive* relationship between data specialization and performance.

Since the two hypotheses have different performance predictions, we can distinguish them by examining the relationship between data specialization and performance. To achieve this goal, we assess the trading performance of data specialists and generalists as follows. For each investor, we keep track of her trades (including both buy and sell decisions) in each stock over time, following the first-in-first-out rule. The shares purchased are kept in her portfolio until she sells them (i.e., the buy position is closed). In other words, we track the actual trades of the investor in replicating her portfolio holdings and estimating her portfolio performance.

At the beginning of each month, we sort investors into five quintiles based on their data specialization estimated in the previous calendar quarter. This time convention ensures that our performance tests are out-of-sample, and it will also be adopted in our future tests. Next, we aggregate the portfolios of all investors in each quintile and compute the quintile-level out-of-sample portfolio return for the month. Table 3 reports the monthly performance of such tracking portfolios for each quintile during the sample period from 2000:04–2010:12.

We observe that institutional investors generate positive and significant monthly returns on average. More importantly, the average monthly return monotonically increases with the data specialization quintiles. In particular, generalists have an average monthly return of 1.23% with a t statistic of 2.77, while specialists have an average monthly return of 1.90% with a t statistic of

3.83. The average monthly return difference between specialists and generalists amounts to 0.67% (annualized 8.3%), with a highly significant t statistic of 5.08.

Similar patterns also apply to DGTW characteristics–adjusted or market-adjusted abnormal returns across the five quintiles. For instance, the DGTW-adjusted return also monotonically increases from the generalist quintile to the specialist quintile, with the monthly DGTW return difference between specialists and generalists amounting to 0.50% (annualized 6.17%), with a *t* statistic of 7.66.[15] Although specialists load more on market risk, the specialist–generalist market beta spread (0.08) is almost negligible in the economic magnitude of its influence on returns. The specialist portfolio also exhibits a higher Sharpe ratio and lower crash risk (i.e., less negative skewness). These observations strongly suggest that data specialists are more informed than data generalists in their investments, supporting the *informed specialist hypotheses*.

3.2. Stock-level predictivity of specialists vs. generalists

We further construct stock-level tests to investigate how informed specialists and generalists are. The informed trading literature suggests that the total demand of an informed investor reflects the value of the signal that she observes (e.g., Kyle 1985). Hence, the stock-level return prediction power of investor demand reveals the informativeness of their trading signals. Accordingly, we measure the aggregate demand of specialists or generalists for each stock in each month as the aggregate holding positions of each type of investor scaled by the average monthly CRSP trading volume in that stock over the previous six months. We then test whether the common signals

[15] In rare cases, we also observe that investors sometimes sell a stock without an existing holding position (as inferred from previous trades), which resembles short selling. However, such trading may also reflect the data limitation that we cannot observe each investor's complete trading history. Hence, we exclude these trading positions from our main analysis and scrutinize them separately (including them in our main analysis does not change our main conclusions). In unreported tests, we also directly compare these "short-selling" portfolios of generalists and specialists and found that specialists also slightly outperform generalists on the short-selling side. The DGTW-adjusted monthly return difference amounts to 0.24% (t statistic 2.14), which is smaller in magnitude than its counterpart in our main results.

processed by all investors (proxied by the aggregate demand of both specialists and generalists) and the signal uniquely processed by specialists (proxied by the demand *difference* between specialists and generalists) can help predict the DGTW-adjusted return of individual stocks.

Table 4 presents the results. First, the aggregate demand of data specialists and generalists has significant return prediction power. In other words, a stock will have higher future returns in the next month if both specialists and generalists hold it in their portfolio.[16] This observation is consistent with our previous results on the general informativeness of institutional investors.

More interestingly, when specialists' demand is higher than that of generalists, the stock delivers significant out-of-sample returns. In particular, a one-standard-deviation increase in the demand *difference* between specialists and generalists (7.17%) predicts a 0.125% higher out-of-sample DGTW-adjusted stock return in the following month (or 1.5% annualized returns). Moreover, such return prediction power lasts for six months without any reversal. Hence our stock-level test also confirms that specialist demand is relatively more informed than that of generalists.

As a robustness check, we also adopt an alternative approach to estimating investor demand as the new trades that investors initiated. In this case, we can estimate the informativeness of any particular group of investors by testing whether the stocks they have recently bought (or sold) yield positive (or negative) returns. Since the conclusions are similar, we tabulate the empirical tests in our Online Appendix (Table A2) and summarize only the key findings here. We observe that stocks exhibit significantly positive (or negative) DGTW-adjusted daily returns when data specialists execute a net buy (or sell) of these stocks in the preceding week. Generalist buys fail to predict stock returns, again suggesting that specialist (generalist) demand is better (less) informed.

[16] A one-standard-deviation increase in the aggregate long position (7.39%) in a stock is associated with a 0.23% increase in the one-month DGTW-adjusted future return of the stock.

Overall, both investor- and stock-level tests confirm that data specialists appear to be more informed. This observation suggests that real-world informed investors may face severe constraints in processing multidimensional data, giving rise to the information processing dynamics described by the *informed specialist hypothesis* (à la Goldstein and Yang 2015).

4. Implications for Market Efficiency

4.1. Specialist distraction and market efficiency

We now investigate how data specialization influences market efficiency via the two proposed economic channels. We start with the *specialist inattention channel,* where a temporary shortage of data specialists' attention may hinder the market's informativeness about the aspect(s) of firms in which these investors specialize. To test this channel, we first notice that it generally takes time for the market to process information and incorporate such information into stock prices. Post-earnings-announcement drift provides a familiar example (e.g., Ball and Brown 1968; and Ben-Rephael, Da, and Israelsen 2017). As a result, we expect the market price to generally underreact to news on the news day, followed by a price drift in the post-news period.

Next, we consider the implication of data specialization. When data specialists who process information in a particular news category become distracted by exogenous shocks, other categories of specialists cannot easily fill the information gap due to the severe constraints or costs of processing multidimensional data. Although generalists may feel interested in filling in the gap, they are less informed and cannot replace the information role of distracted specialists, either. As a result, the market will experience a temporary shortage of skilled investors to process information in this category. A testable implication, in this case, is that the market price underreacts *more* to news on the news day and then "catches up" through an *enhanced* postnews drift, which we refer to as the *specialist inattention channel*.

The test this channel, we follow Kempf, Manconi, and Spalt (2016) and use the fraction of stocks that experienced extreme returns as the measure of exogenous inattention for each institutional investor. The main intuition, in the spirit of Barber and Odean (2008), is that extreme returns (both positive and negative) are attention-grabbing shocks that distract investors' attention elsewhere. Based on this intuition, Kempf, Manconi, and Spalt (2016) show that extreme industry returns distract institutional investors from monitoring corporate insiders.

Although Kempf, Manconi, and Spalt (2016) focus on extreme industry returns, the recent literature points out that extreme stock returns similarly affect investor attention. Indeed, both salience theory (Bordalo, Gennaioli, and Shleifer 2012; 2013; and Bordalo et al. 2020) and asset pricing tests (Cosemans and Frehen 2021) suggest that extreme stock returns relative to the market are salient. Investors correspondingly overweight their attention to such salient attributes and underallocate attention to nonsalient stocks.

Applying the above intuition to a specialist who allocates attention to her invested stocks, we argue that the presence of more salient stocks in her portfolio (those with extreme stock returns relative to the market) leads her to shift attention from nonsalient to salient stocks. Moreover, the stock return salience is relative to the market and is thus caused by idiosyncratic shocks. Hence, a large fraction of stocks with extreme returns serves as a plausibly exogenous shock that distracts the specialist's attention from the information of her *remaining* stocks.

For each news category of a focal stock each month, we calculate the inattention of a specialist based on the stocks in her portfolio and aggregate it to the news category level to obtain a specialist inattention measure for each news category. More explicitly, for any date, we identify stocks with extreme returns—i.e., returns belonging to the highest or lowest return decile in the cross-section—in the previous month (the ranking period). We then calculate a specialist's inattention as her end-of-ranking period portfolio weight invested in the stocks with realized extreme returns. We further aggregate specialist inattention for each news category as the value-weighted individual specialist inattention. The weight is the dollar value of news-related trading in this particular news category (in the same ranking period).

We then pool all the events of firm news releases and examine the impact of specialists' inattention on the postnews price drift. The testing period is over the 20 days following the news of firm i in category *c*, released on day t. Note that our sample focuses on the events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. More explicitly, we examine the following specification:

$$DGTW_{i,c,t:\,t+x} = \alpha + \beta_1 \times ESS_{i,c,t} + \beta_2 \times Inattention_{c,t} + \gamma_1 \times ESS_{i,c,t} \times Inattention_{c,t} + \delta \times M_{i,t} + \epsilon_{i,c,t:t+x}, \tag{2}$$

where $DGTW_{i,c,t:\,t+x}$ is the DGTW-adjusted abnormal return of the stock of firm i accumulated over the period from day t to $t+x$ (or otherwise specified), $ESS_{i,c,t}$ is the sentiment of the news from *RavenPack*, and $Inattention_{c,t}$ refers to the inattention of the specialists who typically process the news in category *c* (the inattention measure is calculated based on stock returns realized in the month before day t). $ESS_{i,c,t}$ and $Inattention_{c,t-1}$ are normalized to have a standard deviation of one. We also include stock-level control variables ($M_{i,t}$) and an array of (firm, news category, and time) fixed effects.

The results are tabulated in Table 5. Column (1) reports a positive and significant coefficient on *ESS*, confirming that the market reacts strongly to the news on day *t*. Moreover, the remaining columns show that the price continues to drift in the same direction as the news sentiment over the next 20 trading days. These observations confirm that it generally takes time for the market to absorb firm-specific news.

More importantly, column (1) reports a significantly negative interaction between *ESS* and specialist inattention, suggesting that specialist inattention *reduces* the speed of incorporating firm-specific news on the news day. Meanwhile, as columns (2) to (8) illustrate, the interaction terms between *ESS* and specialist inattention become positive and significant after the news day. Hence, specialist inattention enhances the postnews price drift. The magnitude of the latter effect is sizable: a one-standard-deviation increase in specialist inattention is associated with 37.5% larger postnews drift from days $t+1$ to $t+20$. These observations lend support to the *specialist inattention channel*.

Figure 1 visualizes the effect of specialists' inattention on the market reactions to the news. We plot the cumulative abnormal returns from days t to $t+20$ for four distinct groups. These four groups are 1) news with positive sentiment and low specialist inattention (at the 25th percentile of the distribution); 2) news with positive sentiment and high specialist inattention (at the 75th percentile); 3) news with negative sentiment and low specialist inattention; and 4) news with negative sentiment and high specialist inattention. We plot the sentiment-related abnormal returns for each group based on the regression coefficients from Table 5, realized sentiment values, and the corresponding inattention percentile values.[17]

From Figure 1, the market reactions on the news-release day t are much smaller for groups with high specialist inattention (depicted with the dashed lines) compared to those with low

[17] For instance, we estimate the abnormal return for the first group as $\widehat{DGTW}_{i,c,t:\,t+x} = \hat{\beta}_1 \times \overline{ESS}^{+}_{i,c,t} + \hat{\gamma}_1 \times \overline{ESS}^{+}_{i,c,t} \times \overline{Inattention}^{25th}_{c,t}$, where $\hat{\beta}_1$ and $\hat{\gamma}_1$ are the estimated regression coefficients, $\overline{ESS}^{+}_{i,c,t}$ is the average value of positive sentiment, and $\overline{Inattention}^{25th}_{c,t}$ is the 25th percentile value of specialist inattention.

specialist inattention (represented by the solid lines). Moreover, the two high-specialist-inattention groups exhibit a larger postnews drift. As a result, their cumulative abnormal returns gradually catch up and converge to the returns of the other two low-distraction groups. These observations clearly illustrate the delayed manner in which the market processes information under high specialist inattention, consistent with the mechanism of the *specialist inattention channel*.

Table 6 provides a placebo test by examining whether generalists' inattention can hinder information dissemination. Under the *informed specialist hypothesis*, generalists play a less important role in processing information. Moreover, if certain generalists are distracted by exogenous shocks, other generalists can easily fill the attention gap. Hence, we expect that generalist inattention should play a much less significant role. This is indeed what we observe from the data. When we replace specialist attention with generalist attention in the postnews drift test, the interaction term between *ESS* and generalist inattention remains largely insignificant throughout the postnews period that we test. The insignificance of generalist inattention also confirms that our postnews drift test has adequate power to reject the role of unimportant players.

Taken together, Tables 5 and 6 suggest that specialists' inattention can significantly hinder information dissemination, while generalists' inattention does not. These results lend support to the specialist inattention channel and strongly suggest that the market for processing different aspects of firms could be segmented due to data specialization.

4.2. Tests on strategic complementarity

The influence of data specialization is not limited to the *specialist inattention channel*. In the spirit of Goldstein and Yang (2015), a *strategic complementarity channel* can further distort market efficiency. Broadly speaking, the uncertainty faced by one category of data specialists decreases (increases) when other specialists process more (less) information about other aspects of the same firm—an uncertainty reduction effect proposed by Goldstein and Yang (2015). Under this channel, a specialist's incentive to process data may still diminish, even with full attention, if distractions from other specialists result in less information about other aspects of the same firm.

To investigate this economic channel, we extend our analysis from specialists covering a specific category (i.e., focal specialists) to those specialists processing information in *other* news categories excluding the focal category (i.e., other specialists). In particular, we focus on the case in which the attention of other specialists—but not that of focal specialists—becomes exogenously distracted. This empirical design excludes the effects of the *specialist inattention channel* and the potential impact of common inattention shocks, allowing us to focus on how other specialist inattention shocks affect the incentives of focal specialists. If the *strategic complementarity channel* is operative, even when focal specialists' attention remains unchanged, the focal news category may nonetheless become less informative under other specialists' inattention shocks.

We test this effect by replacing focal specialist inattention with other specialist inattention in equation (2) as follows:

$$\begin{aligned} DGTW_{i,c,t:\,t+x} = \alpha + \beta_1 \times ESS_{i,c,t} + \beta_2 \times Inattention_{-c,t} + \gamma_1 \times ESS_{i,c,t} \times Inattention_{-c,t} \\ + \delta \times M_{i,t} + \epsilon_{i,c,t:t+x}, \qquad (3) \end{aligned}$$

where we use $Inattention_{-c,t}$ to denote the inattention of other specialists. Other specifications and variables remain the same. To construct $Inattention_{-c,t}$, we first estimate the inattention of other specialists as before using the fraction of her invested stocks that have realized extreme returns in the month before the news-release day. We then aggregate the inattention of all other specialists for the focal news category. Similar to the previous test, our sample includes only news events when the focal stocks did not experience extreme stock returns in the previous month. However, we further restrict the sample to news events when the focal specialists experience no inattention shocks (i.e., their inattention is in the bottom-decile among all the news categories each month). As a result, the sample size of this test is smaller than that of the previous test.

We report the results in Table 7. Panel A documents that the interaction term between *ESS* and other specialist inattention is negative and significant on news day *t*. A one-standard-deviation increase in other specialist inattention is associated with an 11% lower market reaction on the news day. Although the underreaction pattern is similar to what we have observed before, the underlying economic rationale is very different. Indeed, since the focal specialists experience no inattention shocks by construction, the observed underreaction pattern is consistent with the reduced incentives of these specialists, as predicted by the *strategic complementarity channel*.

The speed of price convergence in this test also differs from the *specialist inattention channel.* Empirically, the interaction term remains negative and significant until $t + 3$, indicating that the incentives of the focal specialists remain low these days. After this period, however, the interaction term becomes insignificant, suggesting that the focal specialists resume information processing. The price drift statistically converges as a result. The duration of the *strategic complementarity channel* is noticeably shorter than that of the *specialist inattention channel*. This difference is reasonable, as focal specialists do not need to wait until the attention of other specialists (and the uncertainty reduction effect they provide) has fully recovered. A reasonable recovery in the uncertainty reduction effect is sufficient for the focal specialists to regain their incentives.

Thus far, our results document reduced market reactions induced by the inattention of other specialists. This cross-category effect is consistent with the *strategic complementarity channel* and, due to the market segmentation across different categories, is difficult to explain by alternative economic grounds. Indeed, static specialist attributes alone (including attention) are unlikely to generate the cross-category spillover effect that we observe. The closest alternative explanation could be dynamic attention shifting—i.e., focal specialists might dynamically shift their attention away from their focal category to fill the attention and trading gap left by distracted other specialists. Conceptually speaking, such dynamic attention shifting is unlikely to occur in an economy with data specialization because specialists are difficult to substitute for. Nonetheless, we can empirically scrutinize this alternative explanation to shed more light on specialist behavior.

The idea is to check whether the focal specialists shift their trading away from the focal news categories to other categories. If focal specialists do shift attention to fill the gap, we should observe simultaneously decreased trading in their focal category and increased trading in other firms that do not release news in their focal category.

Panel B tabulates the results of this diagnostic test. For each news event examined in Panel A (i.e., when firm i releases news in category c on day t), we split the (dollar-valued) trades of the focal specialists into two groups. The first group is their trades in the focal category and related firms; the second is their trades in other stocks that have no news in category c. Columns (1) to (4) and columns (5) to (8) report the results for the two groups of trades. From columns (1) to (4), we observe that trades in the first group are reduced when the inattention of other specialists is high. This reduction is consistent with reduced incentives. Columns (5) to (8) further show that trades in the second group are not significantly affected—if not reduced. Hence, focal specialists do not shift their trading elsewhere, which rules out dynamic attention shifting. Taken together, our postnews drift and diagnostic tests strongly support the *strategic complementarity channel*.

Collectively, our channel tests demonstrate that the market's ability to accurately price various aspects of firms may be impeded by both the *specialist inattention channel* and the *strategic* *complementarity channel*. Since this pricing efficiency is at the core of financial markets, data specialization may have profound implications. We next move on to examine such effects.

5. Asset Pricing Implications for Anomaly Returns

When the market becomes less informative, mispricing may arise as a consequence. Since characteristic-based anomalies provide a prominent indicator of mispricing, we now turn to investigate how data specialization influences anomalies to illuminate the pivotal role that data specialization may play in asset pricing.

Specifically, an important strand of the literature demonstrates that a large number of firm characteristics can predict the cross-section of stock returns. This excessive return predictability imposes a "multidimensional challenge" to the asset pricing literature (Cochrane 2011) and invokes heated debate on the economic drivers of such anomalies—see, among others, McLean and Pontiff (2016), Harvey, Liu, and Zhu (2016), Engelberg, McLean, and Pontiff (2018), Hou, Xue, and Zhang (2020), and Kelly and Pedersen (2022) for recent evidence.

Anomalies can be particularly related to data specialization for two reasons. First, the firm characteristics used in the anomaly literature underscore the importance of the multidimensional data structure analyzed in our study. Indeed, we can intuitively think of a particular aspect of firm-level data—such as "Dividends"—as a multitude of firm characteristics. On the one hand, the content of data related to "Dividends" may be quantified by the list of dividend-related firm characteristics, such as dividend yield, dividend initiation, dividend omission, and dividend seasonality—using the characteristics list provided by Chen and Zimmermann (2020). On the other hand, dividend specialists are likely to process information in all dividend-related firm characteristics. In this case, if dividend specialists experience inattention shocks, asset returns associated with all dividend characteristics can be potentially affected.

Second, mispricing is at the core of anomaly returns. The anomaly literature identifies three major economic sources of anomaly returns: systematic risk, (return-based) data mining, and mispricing. Data specialization-induced mispricing can be particularly related to anomaly returns generated on corporate news dates, as explored by Engelberg, McLean, and Pontiff (2018). These authors find that, compared to regular days, anomaly returns are 50% higher on corporate news dates. Furthermore, they show that their results can be explained by mispricing driven by biased expectations and the correction of such mispricing due to the arrival of corporate news. Take dividends as an example. In their framework, anomalies arise when investors mistakenly underestimate the expected dividend value. The market corrects this mistake by generating positive stock returns when new data reveals favorable dividend policies.

We argue that the mispricing arising from data segmentation may significantly affect asset returns in Engelberg, McLean, and Pontiff's (2018) framework. In particular, if dividend specialists experience inattention shocks, this correction effect documented by these authors will be impeded. More generally speaking, specialist inattention shocks should reduce the market correction when new data—i.e., the announcement of corporate news—arrives. This effect, which we refer to as the *reduced-correction hypothesis*, reveals a first-order influence of data specialization on anomaly returns that can be directly tested.

To empirically test this hypothesis, we follow Engelberg, McLean, and Pontiff (2018) and construct a stock-level anomaly proxy. For each stock-month observation, we first create a Long (Short) measure by summing the number of long-side (short-side) anomaly portfolios that the stock belongs to, from a list of 212 characteristics.[18] Engelberg, McLean, and Pontiff (2018) calculate their stock-level anomaly proxy, Net, as $Long$ minus $Short$ of these positions. We further normalize Net to have a standard deviation of one. We measure Net of each stock at the beginning of each month and then link it to the news-day returns of stocks. Our specification is as follows:

$$Return_{i,c,t} = \alpha_t + \beta_1 \times Net_{i,t} + \beta_2 \times Inattention_{c,t}$$
$$+\gamma_1 \times Net_{i,t} \times Inattention_{c,t} + \delta \times M_{i,t} + \epsilon_{i,c,t}, \tag{4}$$

where $Return_{i,c,t}$ refers to the return of stock i on day t following news in category c, and $Inattention_{c,t}$ refers to the inattention shock of the specialists who typically process data in category c. The inattention shock arises due to abnormal stock returns in the portfolio holdings of the specialists, as specified in the *specialist inattention channel*. In all regressions, we include control variables of lagged 1-month stock return and lagged 1-month volatility. Year-month fixed effects, news category fixed effects, and date fixed effects are included in different specifications.

It is worth noting that Engelberg, McLean, and Pontiff (2018) examine anomaly returns on all dates, which allows them to compare corporate news date returns with other dates and interpret news-day returns as the correction of mispricing. Since the processing and trading of corporate news are at the core of data analysis, we directly focus on corporate news-day results. Accordingly, the sample of our test includes corporate news dates (i.e., the release date of the corporate news). Furthermore, we exclude the news dates of salient stocks—i.e., stocks that experienced extreme stock returns in the previous month. This is because these stocks attract specialist attention and hence cannot be used to test the impact of specialist inattention shocks (although empirically including these stocks does not change the main results).

[18] We use the data of stock characteristics from Chen and Zimmermann (2020) at https://www.openassetpricing.com/data/. Our results are robust to the sample of multidimensional anomalies.

The results of our analysis are tabulated in Table 8. In Columns (1) to (4), the dependent variables are the news-day DGTW-adjusted returns. Among these models, the first two aim to replicate the main result of Engelberg, McLean, and Pontiff (2018) in our setup—i.e., the anomaly return realized on corporate news dates. The next two models further report the full model by interacting $Net_{i,t}$ with specialist inattention. To be comparable to Engelberg, McLean, and Pontiff (2018), we also report the results of unadjusted news-day returns in Columns (5) to (8).

We first observe that anomalies are associated with significantly positive returns on news days, confirming the correction effect of new data arrival as documented in Engelberg, McLean, and Pontiff (2018). Both Models (1) and (5) report the coefficient of the anomaly proxy (β_1) as 0.0010. With all independent variables normalized, this β_1 coefficient can be directly interpreted as the anomaly return associated with a one-standard-deviation increase in the anomaly proxy ($Net_{i,t}$). In other words, a one-standard-deviation increase in the anomaly proxy ($Net_{i,t}$) is associated with an additional 10 basis points daily return on news days. This economic magnitude is on par with the estimation made by Engelberg, McLean, and Pontiff (2018).[19]

More importantly, we observe that the interaction term between the anomaly proxy ($Net_{i,t}$) and specialist inattention is significantly negative, consistent with the notion that specialist inattention shocks hinder the correction effect. To estimate the economic magnitude, we can interpret the ratio γ_1/β_1 as the marginal impact of (one-standard-deviation higher) specialist inattention on the correction effect of new data arrival. In this context, Models (3) and (4) report that a one-standard-deviation increase in inattention shocks is associated with approximately a 20%

[19] For instance, Engelberg, McLean, and Pontiff (2018) report that a Net value of 10 (about 1.5 standard deviations) is associated with a return of 3.84 basis points higher on nonnews days, 5.62 basis points higher on news days, and 21.64 basis points higher on earnings announcement days. Our estimations do not differentiate between earnings and non-earnings news. As a result, the economic magnitude of our estimation falls between their estimates for these two types of events.

reduction in price correction using DGTW-adjusted return. Models (7) and (8) indicate a similar effect with a slightly smaller economic magnitude.

Collectively, our empirical tests lend strong support to the *reduced-price correction hypothesis*. This verification provides important insights into the economic foundations of anomalies. Given that inattention shocks are specialist-specific, it is highly implausible that they can be attributable to systematic risk or result from *ex-post* data mining. Hence, of the three economic sources that may contribute to anomaly returns—i.e., mispricing, data mining, and systematic risk—mispricing provides the most plausible explanation to account for the part of anomaly returns affected by inattention shocks.[20] In this regard, our results provide direct evidence to support the mispricing interpretation of anomaly returns and underscore the potential influence that data specialization may have on asset prices.

6. Additional Analyses

The previous sections have shown that informed investors process multidimensional data in a specialized manner and that limited specialist attention may impede market efficiency. We now conduct a battery of tests to further shed light on the economics of data specialization.

6.1. Alternative explanations and characteristics-adjusted data specialization

We first investigate what affects institutional investors' data specialization and whether the superior performance of specialists can be explained by other characteristics or known trading strategies of institutional investors.

[20] For instance, one caveat noted by Engelberg, McLean, and Pontiff (2018) is that data mining may serve as a potential alternative explanation, even though this channel alone may struggle to account for the high anomaly returns on news days. Our results are even less susceptible to data mining contamination because the latter is typically conducted by researchers ex-post, without considering the potential influence of data specialists.

The literature indeed suggests a list of alternative explanations. First of all, fund size, via the channel of diseconomies of scale, is known to affect the cost of operation and, as a result, fund returns (Berk and Green, 2004) may also be affected. Since the operational costs may also encompass expenses related to data processing in the era of big data, a spurious relationship between the fund return and data specialization may arise as a result channeled through fund size.

Second, active trading strategies reflected in active share (Cremers and Petajisto 2009), interim trading (Kacperczyk, Sialm, and Zheng 2008), and deviations from a factor model (Amihud and Goyenko 2013) may also potentially generate our results. A fund that deviates more from a factor model by adopting concentrated active trading, for instance, may exhibit trading concentration on one hand and higher expected fund returns on the other hand.

Finally, known proxies of information utilization and asset specialization, particularly the reliance on public news (Kacperczyk and Seru 2007) and industry concentration (Kacperczyk, Sialm, and Zheng 2005), may be related to data specialization. Kacperczyk and Seru (2007) show that uninformed mutual funds rely on public information processed by analysts, whose trades may spuriously concentrate on news categories. In addition, asset specialization, such as industry concentration, is known to deliver better performance (e.g., Kacperczyk, Sialm, and Zheng 2005). Although data specialization conceptually differs from asset specialization, the two may be empirically correlated.

To deal with these alternative explanations, we take a parsimonious two-step approach. In the first step (first-stage analysis), we regress the data specialization index of each investor each quarter on a set of contemporaneous proxies for these effects. In the second step, we build a characteristics-adjusted data specialization index, which is proxied by the residual from the first-stage regression. We then revisit our main tests using adjusted data specialization.

The proxies for alternative explanations include portfolio size, quarterly trading volume, quarter-end holding rebalance, the number of days with news-related trading within the quarter, the R squared measure of Amihud and Goyenko (2013), the active share using S&P as a benchmark (Cremers and Petajisto, 2009), the reliance on public news (Kacperczyk and Seru 2007), and the industry concentration (Kacperczyk, Sialm, and Zheng 2005). Among these proxies, we include the number of news-trading days to capture the exposure of (in)active strategies to news days. Since we do not directly observe the benchmarks of investors, we use the S&P 500 index, the most popular mutual fund benchmark, to calculate the active share of investors. We further use quarterly dollar trading volume and quarter-end percentage holding rebalance to capture the degree of active interim trading of the portfolio (Kacperczyk, Sialm, and Zheng 2008), respectively, in absolute and relative terms. To proxy for each investor's trading reliance on firm news, we divide an investor's news-related trading by her total trading (including news- and nonnews-related trading) each quarter. We also calculate the investor's portfolio trading in each industry and construct the industry concentration proxy as the Herfindahl index.

The first-stage regression results are reported in the Online Appendix (Table A3). Even though each characteristic alone is significantly correlated with the news category concentration index, their marginal explanatory power, as revealed by the adjusted R-square, is very small (approximately 2%). The only exception is the reliance on news, which is negatively related to our measure and explains 21.3% of our measure. Given that reliance on public news predominately characterizes uninformed mutual funds (Kacperczyk and Seru 2007), this variable is unlikely to absorb the return-predicting power of our measure.

Among the remaining variables, data specialization is negatively correlated with the quarterly trading volume and the total number of days with news-related trading, suggesting that specialists on average incur lower trading volume and have news-related trading on fewer days than generalists. Data specialization is also positively correlated with proxies that indicate managerial skills, such as industry concentration and active share, consistent with the notion that data specialists are informed. Collectively, however, these alternative channels explain only 24.7% of data specialization in terms of R-squared. Adding to the observation that the explanatory power comes primarily from reliance on public news, it is unlikely that the observed superior performance of data specialists can be attributed mainly to these alternative variables.

To validate the above conjecture, we next examine characteristics-adjusted data specialization. We proxy for data specialization by the residual from the regression in Table A3 Column (9). Specifically, we calculate a predicted news category concentration index by summing the product of each characteristic's coefficient and the corresponding characteristic for every investor each quarter. Subtracting this predicted value from the actual concentration index yields the characteristics-adjusted data specialization index.

With the characteristics-adjusted specification, we first revisit the performance implication of data specialization. The results, as reported in Panel A of Table 9, are very similar to the baseline results in Table 3. The average monthly return monotonically increases with the quintiles based on the characteristics-adjusted index. Generalists have an average monthly return of 1.44% (with a t statistic of 3.11), while specialists have an average monthly return of 1.90% (with a t statistic of 3.87). The average monthly return difference between specialists and generalists is lower than that in Table 3 but still amounts to 0.46% (annualized 5.67%), with a highly significant t statistic of 4.84. We also observe similar patterns in DGTW–adjusted or market-adjusted abnormal returns across the five characteristics-adjusted data specialization quintiles, suggesting that alternative explanations do not fully account for the outperformance of data specialists.

Since reliance on public news and industry concentration are the two foremost alternative explanations in our first-stage analysis, we dedicate special attention to them by conducting a double-sort test. For each of the two variables, we independently double-sort all investors into 25 groups according to our measure and one of these two measures. For brevity, we tabulate the average monthly DGTW-adjusted returns for these portfolios in the Online Appendix (Table A4) and only discuss our main findings here.

Our main observations are twofold. First, when holding our concentration index fixed, we observe that portfolio returns generally decline across reliance-on-public news quintiles and increase across industry-concentration quintiles. These observations are consistent with the literature documenting that fund managers relying more on public news and exhibiting more asset concentration are, respectively, less and more informed (Kacperczyk and Seru 2007; Kacperczyk, Sialm, and Zheng 2005). Secondly, portfolio returns increase in data specialization within each reliance-on-news or industry-concentration quintile, allowing data specialists to significantly outperform generalists in each case. These results demonstrate that data specialization captures distinct economic behaviors of institutional investors beyond their use of public news or asset specialization.

We then further examine the two channels of market efficiency for characteristics-adjusted data specialization. The results are tabulated in Panel B of Table 9. In particular, Columns (1) to (3) report the effect of focal specialists' inattention on market reactions to the news. Similar to previous results (Table 5), we observe a negative interaction term between *ESS* and specialist inattention on the news day t, showing that specialist inattention hinders the market reaction to the news. The positive interaction terms between *ESS* and specialist inattention after news suggest that specialist inattention enhances postnews price drift from days $t + 1$ to $t + 20$.

Columns (4) to (6) test the complementarity effect by examining the effects of other specialists' inattention on the market reactions to the news in the focal category. The results are highly robust, if not more significant. In particular, other specialists' inattention can significantly reduce market reactions to the news date when focal specialists' attention remains unchanged. The statistical power is even larger compared to our baseline results without controlling for other strategies and characteristics. These observations support the strategic complementarity channel.

Taken together, the results show that even after controlling for a range of known strategies and salient characteristics of institutional investors, there remains a remarkable variation in how they process firm-specific information in terms of data specialization. Moreover, data specialists remain more informed and play a fundamental role in influencing market efficiency.

6.2. Data specialists vs. market-wide information

To further understand the incentives and skill sets of data specialists, we investigate how they deal with market-wide information. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014) suggest that in bear markets, investors may rationally shift attention from firm-specific news to market-wide information. Yet, our previous results indicate that it may be costly for data specialists to process multidimensional data for the same firm. Given that the gap between firm-level and market-wide data is unlikely to be narrower than that of cross-category data of the same firm, we should expect data specialists to face a similar, if not more severe, barrier in processing firm-specific and market-wide data. This would likely induce them to maintain focus on specialized firm aspects rather than dynamically reallocating attention to market-wide information.

To test this notion, we follow Daniel and Moskowitz (2016) and estimate a CAPM with bear market timing ability in the following time-series regression:

$$R_{p,t} = (\alpha_0 + \alpha_B I_{B,t-1}) + (\beta_0 + I_{B,t-1}(\beta_B + I_{U,t}\beta_{B,U}))R_{m,t} + \epsilon_t, \tag{5}$$

where the independent variable $R_{p,t}$ is the monthly return in excess of the risk-free rate for investor p. We separately aggregate the holding portfolios of all generalists and all specialists to calculate the returns of these two types of investors. The dependent variables include $R_{m,t}$, the market return in excess of the risk-free rate; the indicator for bear markets, $I_{B,t-1}$, which equals one if the cumulative past two-year return on the market is negative; and the indicator for a positive market return, $I_{U,t}$, which equals one if $R_{m,t}$ is positive.

The economic interpretation of the regression parameters is as follows. First, the parameters α_0 and α_B measure the selection ability of the investor in general and under the bear market in particular. Next, β_0 and β_B measure the market exposure of the investor in general and under the bear market. Finally, $\beta_{B,U}$ captures the market timing ability of the investor during the bear market—a positive coefficient indicates positive market timing, as the investor increases beta in anticipating a positive market return. Suppose that investors relocate their attention from stock selection to market-wide information in bear markets. In that case, we would expect to observe positive bear market timing to reflect the benefit of processing market-wide information (i.e., a positive $\beta_{B,U}$) and a corresponding decline in bear market selection (i.e., a negative α_B).

The results are reported in Table 10, with the first (last) two columns showing the results for generalists (specialists). We first observe that although both types of investors have a certain level of selection ability, specialists' selection is markedly superior. Indeed, the parameter α_0 for specialists (e.g., 0.0111, which means 1.11% abnormal return per month) is almost double that for generalists (0.0055).

Second, column (2) suggests that generalists have significant market timing ability during bear markets ($\beta_{B,U} > 0$). However, this ability is achieved at the price of selection—generalists exhibit a significantly negative selection during the same period ($\alpha_B < 0$). In other words, generalists sacrifice stock selection to benefit from market timing during bear markets. This time-series pattern is indeed consistent with investors rationally shifting their attention from firm-specific information to market-wide information in bear markets, as described in Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014).[21]

Column (4) suggests that, in contrast, data specialists have consistent selection ability and do not resort to market timing during bear markets. Hence, specialists do not significantly reallocate their attention in bear markets. Interestingly, these investors specialize not only in the cross-section but also in time series. They consistently rely on superior (yet specialized) firm-level information processing ability to generate performance regardless of general market conditions. This behavior is more consistent with the model of Van Nieuwerburgh, and Veldkamp (2009), in which investors benefit from specializing in areas where they have some initial advantage.

In the model of Van Nieuwerburgh, and Veldkamp (2009), when local investors possess a small initial information advantage on local assets, they benefit from learning more about local assets—i.e., specializing in what investors already know is a more profitable strategy. Although this argument centers on asset specialization, the same intuition could extend to the processing of multi-dimensional information. Investors with a small initial advantage in processing a particular dimension of data may find it beneficial to specialize in that dimension. This parallel highlights the importance of understanding data specialization. It also suggests that a better comprehension

[21] The tradeoff between market timing and selection can also be interpreted as generalists buying call options on the market during bear markets (Daniel and Moskowitz 2016). Hence, generalists benefit from the upside of the market while paying the cost.

of the market equilibrium in the big data era might require a theoretical foundation that synchronizes the effects of both data specialists and generalists, with a particular emphasis on the incentives and impacts of data specialization.

6.3. Alternative specialization measures as robustness checks

In our main analysis, we use the fifteen most frequently occurring news categories as defined by *RavenPack* to alleviate the concern of data mining related to category construction. To test whether our results are sensitive to the classifications of news categories, we now conduct a clustering analysis combining closely related categories. We first calculate the closeness of any pair of two categories in terms of trading. In particular, we calculate the common span of any given investor in trading in two categories in a given quarter. Empirically, two pairs of categories, "earnings"/"revenues" and "analysts rating"/"price targets", are most jointly traded by investors. Hence, we combine "earnings" and "revenues" to create a new category "earnings and revenues" and, similarly, "analysts rating" and "price targets" into "analysts rating and targets".

We then use the alternatively defined news categories to calculate the news category concentration index and control for investor characteristics as in Section 6.1. The results are reported in Table A5. Panel A examines the performance of each quintile of investors and shows that specialists deliver an average monthly return of 1.93% (with a t statistic of 3.95), which is 0.52% higher than that of generalists. Panel B tests the implications of data specialization on market efficiency and we find that both the inattention of focal specialists (as shown in Columns 1 to 3) and the inattention of other specialists (as shown in Columns 4 to 6) hinder the market efficiency.

Additionally, our main measure is based on the number of news-related trades. As a second set of robustness checks, we also use the dollar amount of news-related trades to constructure a data specialization measure, which can be considered a value-weighted measure of our original measure. Hence, we construct an alternative measure based on how investors allocate the dollar value of trades across different news categories. We assess the performance implications of the alternative measure vis-à-vis the *informed specialist hypothesis* in the appendix (Table A6 Panel A). When we sort investors into quintiles based on this new measure, the performance pattern is quantitatively the same as that under our main proxy. For instance, data generalists deliver an average monthly return of 1.3% (with a t statistic of 2.86) while data specialists deliver an average monthly return of 1.82% (with a t statistic of 3.75). Their return difference is approximately the same as that in our previous tests. We also reconduct the two market efficiency tests using the alternative measure to classify specialists and find robust results, as shown in the appendix (Table A6 Panel B).

Overall, our results are highly robust to the clustering of categories and the use of alternative definitions of data specialists. Although these alternative empirical setups may change the explicit definition of data specialists, the underlying economic implication remains consistent: data specialization is predominantly a strategy employed by more informed investors, which can affect market efficiency due to the segmented processing of information.

7. Conclusion

This paper examines how multiple investors process multidimensional data for the same asset and investigates whether market efficiency can be affected as a result. Based on a sample of institutional trades across multidimensional categories of firm-specific news, we find a *positive* relationship between data specialization and performance. This observation suggests that more informed institutional investors tend to specialize in subsets of firm aspects, lending support to the *informed specialist hypothesis* as depicted in Goldstein and Yang (2015).

Importantly, such data specialization may profoundly affect market efficiency through two channels. First, inattention shocks to data specialists hinder price efficiency in their specialized aspects of firms, even when the shocks may concern only a small fraction of specialists. Second, other aspects of firms may also be negatively influenced due to strategic complementarity. The specialists of a focal category process less information when the specialists of other categories experience inattention shocks. The market becomes less informative on the focal aspects of firms in both cases.

These results suggest that data specialization may provide a novel economic basis on which to assess market efficiency and related implications in the age of big data. Indeed, we observe that data specialization may profoundly affect asset prices by impeding the price corrective effect of news arrival on anomalies. Our findings call for more research to understand the multidimensional nature and normative implications of data.

References

Admati, A. R., and Pfleiderer, P. 1987. Viable allocations of information in financial markets. *Journal of Economic Theory* 43:76–115.

Amihud, Y., and Goyenko, R. 2013. Mutual fund's R^2 as predictor of performance. *Review of Financial Studies* 26:667-694.

Anand, A., Irvine, P., Puckett, A., and Venkataraman, K. 2011. Performance of institutional trading desks: An analysis of persistence in trading costs. *Review of Financial Studies* 25:557–598.

Anand, A., Irvine, P., Puckett, A., and Venkataraman, K. 2013. Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis. *Journal of Financial Economics* 108:773–797.

Avdis, E. 2016. Information tradeoffs in dynamic financial markets. *Journal of Financial Economics* 122:568–584.

Balasubramanian, A., and Yang, Y. D. 2019. Trading with high-dimensional data. *Available at SSRN 3583217*.

Ball, R., and Brown, P. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6:159–178.

Barber, B. M., and Odean, T. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Becker, G. S. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70:9–49.

Begenau, J., Farboodi, M., and Veldkamp, L. 2018. Big data in finance and the growth of large firms. *Journal of Monetary Economics* 97:71–87.

Ben-Rephael, A., Da, Z., and Israelsen, R. D. 2017. It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies* 30:3009–3047.

Benhabib, J., Liu, X., and Wang, P. 2019. Financial markets, the real economy, and self-fulfilling uncertainties. *Journal of Finance* 74:1503–1557.

Bordalo, P., Gennaioli, N., and Shleifer, A. 2012. Salience theory of choice under risk. *Quarterly Journal of Economics* 127:1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. 2013. Salience and consumer choice. *Journal of Political Economy* 121:803–843.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. 2020. Overreaction in macroeconomic expectations. *American Economic Review* 110:2748–82.

Chemmanur, T. J., He, S., and Hu, G. 2009. The role of institutional investors in seasoned equity offerings. *Journal of Financial Economics* 94:384–411.

Chen, A.Y., and Zimmermann, T. 2020. Publication bias and the cross-section of stock returns. *Review of Asset Pricing Studies* 10:249–289.

Chen, H., Cohen, L., Gurun, U., Lou, D., and Malloy, C. 2018. *IQ from IP: Simplifying search in portfolio choice* No. w24801. National Bureau of Economic Research.

Christophe, S. E., Ferri, M. G., and Angel, J. J. 2004. Short-selling prior to earnings announcements. *Journal of Finance* 59:1845–1876.

Christophe, S. E., Ferri, M. G., and Hsieh, J. 2010. Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics* 95:85–106.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance 66*:1047–1108.

Cosemans, M., and Frehen, R. 2021. Salience theory and stock prices: Empirical evidence. *Journal of Financial Economics* 140:460–483.

Cremers, K. M., and Petajisto, A. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22:3329–3365.

Custódio, C., Ferreira, M. A., and Matos, P. 2013. Generalists versus specialists: Lifetime work experience and chief executive officer pay. *Journal of Financial Economics* 108:471–492.

Daniel, K., and Moskowitz, T. J. 2016. Momentum crashes. *Journal of Financial Economics* 122:221–247.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–1058.

DellaVigna, S., and Pollet, J. M. 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64:709–749.

Dow, J., Goldstein, I., and Guembel, A. 2017. Incentives for information production in markets where prices affect real investment. *Journal of the European Economic Association* 15:877–909.

Duffie, D., and Strulovici, B. 2012. Capital mobility and asset pricing. *Econometrica* 80:2469–2509.

Dugast, J., and Foucault, T. 2018. Data abundance and asset price informativeness. *Journal of Financial Economics* 130:367–391.

Engelberg, J., McLean, R.D. and Pontiff, J. 2018. Anomalies and news. *Journal of Finance* 73:1971–2001.

Erel, I., L. Stern, C. Tan, and M. S. Weisbach. 2021. Selecting directors using machine learning. *Review of Financial Studies* 34:3226–64.

Errunza, V., and Losq, E. 1985. International asset pricing under mild segmentation: Theory and test. *Journal of Finance* 40:105–124.

Farboodi, M., and Veldkamp, L. 2020. Long-run growth of financial data technology. *American Economic Review* 110:2485–2523.

Farboodi, M., and Veldkamp, L. 2023. Data and Markets. *Annual Review of Economics* 15:23–40.

Farboodi, M., Matray, A., Veldkamp, L., and Venkateswaran, V. 2022. Where has all the data gone? *Review of Financial Studies* 35:3101–3138.

Farboodi, M., Singal D., Veldkamp, L., and Venkateswaran, V. 2023. Valuing Financial Data. *Available at SSRN 3947931*

Ferreira, D., and Sah, R. K. 2012. Who gets to the top? Generalists versus specialists in managerial organizations. *RAND Journal of Economics* 43:577–601.

Ganguli, J. V., and Yang, L. 2009. Complementarities, multiplicity, and supply information. *Journal of the European Economic Association* 7:90–115.

Garcia, D., and Strobl, G. 2011. Relative wealth concerns and complementarities in information acquisition. *Review of Financial Studies* 24:169–207.

Gilje, E. P., Gormley, T. A., and Levit, D. 2020. Who's paying attention? Measuring common ownership and its impact on managerial incentives. *Journal of Financial Economics* 137:152–178.

Glebkin, S., Gondhi, N., and Kuong, J. C. F. 2021. Funding constraints and informational efficiency. *Review of Financial Studies* 34:4269–4322.

Goldstein, I., and Yang, L. 2015. Information diversity and complementarities in trading and information acquisition. *Journal of Finance* 70:1723–1765.

Goldstein, I., C. Spatt, and M. Ye, 2021. Big Data in Finance. *Review of Financial Studies* 34:3213–3225.

Greenwood, R., Hanson, S. G., and Liao, G. Y. 2018. Asset price dynamics in partially segmented markets. *Review of Financial Studies* 31:3307–3343.

Grinblatt, M., and Titman, S. 1993. Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business* 66:47–68.

Gromb, D., and Vayanos, D. 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics* 66:361–407.

Grossman, S. J., and Miller, M. H. 1988. Liquidity and market structure. *Journal of Finance* 43:617–633.

Grossman, S. J., and Stiglitz, J. E. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393–408.

Harvey, C. R., Liu, Y., and Zhu, H. 2016. … and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Hellwig, C., and Veldkamp, L. 2009. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies* 76:223–251.

Henry, T. R., and Koski, J. L. 2017. Ex-dividend profitability and institutional trading skill. *Journal of Finance* 72:461–494.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance* 64:2289–2325.

Hou, K., Xue, C., and Zhang, L. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–2133.

Huang, A. G., Tan, H., and Wermers, R. 2020. Institutional trading around corporate news: Evidence from textual analysis. *Review of Financial Studies* 33:4627–4675.

Huang, S. 2015. Delegated information acquisition and asset pricing. *Available at SSRN 2666878*.

Jensen, T.I., Kelly, B. and Pedersen, L.H. 2023. Is there a replication crisis in finance?. *Journal of Finance* 78:2465–2518.

Jiang, L., Liu, J., Peng, L. and Wang, B. 2022. Investor attention and asset pricing anomalies. *Review of Finance* 26:563–593.

Jones, C. I. and Tonetti, C. 2020. Nonrivalry and the economics of data. *American Economic Review* 110:2819–2858.

Kacperczyk, M., and Seru, A. 2007. Fund manager use of public information: New evidence on managerial skills. *Journal of Finance* 62:485–528.

Kacperczyk, M., Sialm, C., and Zheng, L. 2005. On the Industry Concentration of Actively Managed Equity Mutual Funds. *Journal of Finance* 60:1983–2011.

Kacperczyk, M., Sialm, C., and Zheng, L. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21:2379–2416.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. 2014. Time-varying fund manager skill. *Journal of Finance* 69:1455–1484.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. 2016. A rational theory of mutual funds' attention allocation. *Econometrica* 84:571–626.

Kempf, E., Manconi, A., and Spalt, O. 2016. Distracted shareholders and corporate actions. *Review of Financial Studies* 30:1660–1695.

Kolanovic M, Krishnamachari RT. 2017. Big data and AI strategies: Machine learning and alternative data approach to investing. Rep., Glob. Quant. Deriv. Strategy, J.P. Morgan, New York.

Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–1335.

Lee, S. 2013. Active investment, liquidity externalities, and markets for information. *Liquidity Externalities, and Markets for Information (June 13, 2013)*.

Lundholm, R. J. 1991. Public signals and the equilibrium allocation of private information. *Journal of Accounting Research* 29:322–349.

Martin, I. W., and Nagel, S. 2022. Market efficiency in the age of big data. *Journal of Financial Economics* 145:154–177.

McLean, R. D., and Pontiff, J. 2016. Does academic research destroy stock return predictability?. *Journal of Finance* 71:5–32.

Merton, R. 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance* 42:483–510.

Morck, R., Yeung, B., and Yu, W. 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements?. *Journal of Financial Economics* 58:215–260.

Murphy, K. J., and Zabojnik, J. 2004. CEO pay and appointments: A market-based explanation for recent trends. *American Economic Review* 94:192–196.

Paul, J. M. 1993. Crowding out and the informativeness of security prices. *Journal of Finance* 48:1475–1496.

Peng, L. 2005. Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis* 40:307–329.

Peng, L., and Xiong, W. 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80:563–602.

Puckett, A., and Yan, X. 2011. The interim trading skills of institutional investors. *Journal of Finance* 66:601–633.

Schultz, T. W. 1961. Investment in human capital. *American Economic Review* 51:1–17.

Sonney, F. 2007. Financial analysts' performance: Sector versus country specialization. *Review of Financial Studies* 22:2087–2131.

Stapleton, R. C., and Subrahmanyam, M. G. 1977. Market imperfections, capital market equilibrium and corporation finance. *Journal of Finance* 32:307–319.

Titman, S., and Grinblatt, M. 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Bussines 62*: 393–416.

Van Nieuwerburgh, S., and Veldkamp, L. 2009. Information immobility and the home bias puzzle. *Journal of Finance* 64:1187–1215.

Veldkamp, L. L. 2006. Media frenzies in markets for financial information. *American Economic Review* 96:577–601.

Zambrana, R., and Zapatero, F. 2021. A tale of two types: Generalists vs. specialists in asset management. *Journal of Financial Economics* 142:844–861.

Table 1: RavenPack news categories

This table presents the list of the 15 most frequently occurring news categories within the *RavenPack* database, together with the frequency of each news category. The categorization of each news is based on the *RavenPack* database, which assigns a category to each news of every stock. We label a trade of institutional investors from the *ANcerno* database as *news-related* if there is news reported for the stock (firm) either on the day of the trade or the day before, and classify each news-related trade into one of the 15 news categories accordingly. Column 4 presents the number of investor-quarters with positive news-related trading in each news category.

News categories	Frequency	Frequency (%)	# of investor-quarters with positive news-related trading
Insider trading	379,371	23.48	36,543
Order imbalances	234,490	14.51	45,875
Earnings	196,965	12.19	47,146
Products services	154,560	9.56	47,362
Analyst ratings	133,101	8.24	47,083
Revenues	121,575	7.52	45,682
Labor issues	89,728	5.55	44,190
Equity actions	57,405	3.55	41,492
Acquisitions mergers	47,920	2.97	39,971
Investor relations	38,930	2.41	28,977
Dividends	38,523	2.38	32,693
Credit ratings	31,464	1.95	33,194
Stock prices	26,217	1.62	40,246
Price targets	10,311	0.64	21,051
Other categories	55,490	3.43	41,494

Table 2: News category concentration

Panel A of this table presents the summary statistics of the news category concentration index for each quintile of investors, classified by their news category concentration index value from low to high. We label a trade of institutional investors from the *ANcerno* database as *news-related* if there is news reported for the stock (firm) either on the day of the trade or the day before, and classify each news-related trade into one of the 15 news categories accordingly. Then, we construct a news category concentration index for each investor each quarter as the sum of the squared percentages of news-related trades in each news category. We refer to investors in the highest (lowest) concentration quintile as data *specialists* (*generalists*). Panel B reports the transition rates of investors across concentration quintiles over time. The transition rate represents the probability that an investor classified into quintile *i* in this period is classified into quintile *j* in the next quarter. In Panel C, we present the trading characteristics of investors across each quintile, including the average monthly trading dollars (log) and average monthly turnover.

Panel A: Quintiles based on the news category concentration

Quintile by news category concentration	# of investor-quarters in each quintile	News category concentration				
		Mean	Median	p1	p99	Standard deviation
1 (Generalists)	10649	13.32%	12.78%	9.41%	34.32%	3.50%
2	10642	16.87%	15.83%	12.93%	43.71%	4.77%
3	10633	20.53%	19.03%	15.32%	51.90%	5.89%
4	10631	25.80%	24.17%	18.52%	59.43%	7.04%
5 (Specialists)	10611	44.90%	37.19%	23.99%	100.00%	21.39%

Panel B: Transition matrix

Quintile by news category concentration	Transition rates of investors across quintiles (%)				
	1 (Generalists)	2	3	4	5 (Specialists)
1 (Generalists)	49.27	26.09	13	6.69	4.95
2	26.85	33.68	21.27	11.09	7.11
3	12.79	22.32	31.65	21.36	11.89
4	7.18	11.32	21.93	35.94	23.63
5 (Specialists)	5.25	7.34	13.07	25.6	48.74

Panel C: Trading characteristics of each quintile

Quintile by news category concentration	Avg. monthly trading dollars (log)	Avg. monthly turnover
1 (Generalists)	15.80	0.11
2	16.20	0.12
3	16.13	0.11
4	15.81	0.11
5 (Specialists)	15.16	0.10

Table 3: Portfolio performance of specialists vs. generalists

This table reports the characteristics of the monthly portfolio performance for each quintile over the 129-month sample period from 2000:04–2010:12. For each investor, we keep track of her trades (including both buy and sell decisions) in each stock over time, following the first-in-first-out rule, and compute her actual portfolio return each month by following her actual holding periods in each stock. Then, we aggregate the portfolios of all the investors in each quintile and compute the quintile-level actual portfolio return each month. The last column of the table presents a new portfolio constructed by buying the specialist's portfolio and short-selling the generalist's portfolio. T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

	Quintile 1 (Generalists)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Specialists)	Difference (5 minus 1)
Average monthly return	1.23%***	1.35%***	1.61%***	1.88%***	1.90%***	0.67%***
	(2.77)	(3.00)	(3.52)	(3.81)	(3.83)	(5.08)
Average monthly return (DGTW-adjusted)	0.90%***	0.95%***	1.12%***	1.28%***	1.40%***	0.50%***
	(11.16)	(11.94)	(13.86)	(14.39)	(14.59)	(7.66)
σ	5.04%	5.09%	5.20%	5.62%	5.63%	1.49%
α	1.05%***	1.17%***	1.43%***	1.71%***	1.72%***	0.67%***
	(11.08)	(10.92)	(11.13)	(10.88)	(10.32)	(5.24)
β	1.01***	1.01***	1.02***	1.09***	1.08***	0.08***
	(52.03)	(46.23)	(38.73)	(33.95)	(31.78)	(2.90)
SR	0.20	0.23	0.27	0.30	0.30	0.31
Skewness	-0.23	-0.23	-0.26	-0.20	-0.18	0.77
Kurtosis	3.19	3.23	3.06	3.15	3.35	5.72

Table 4: Stock return predictivity by specialists and generalists demands

This table examines the stock-level return predictivity of specialists compared to generalists. More explicitly, we estimate the following regression using monthly stock level observations:

$$DGTW_{i,t+x} = \alpha + \beta \times Demand\ Difference_{i,t} + \delta \times M_{i,t} + \epsilon_{i,t+1},$$

where $DGTW_{i,t+x}$ represents to the DGTW-adjusted abnormal return of stock i accumulated from month t to month $t + x$. We aggregate the holding position of each type of investor for each stock in each month, scaled by the average monthly CRSP trading volume in that stock in the previous six months. The main independent variable is the difference in demand between specialists and generalists (i.e., specialists' aggregate holding minus generalists' aggregate holding), together with other control variables ($M_{i,t}$) such as the aggregate demand of both specialists and generalists, current month return, lagged 6-month cumulative return, and other firm characteristics. T statistics are reported in parentheses and are based on standard errors clustered by firm. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent variable	DGTW-adjusted return$_{i,t+x}$					
	[t, t+1m]	[t, t+2m]	[t, t+3m]	[t, t+4m]	[t, t+5m]	[t, t+6m]
	(1)	(2)	(3)	(4)	(5)	(6)
Demand Difference (Specialists – Generalists)	0.0174***	0.0341***	0.0426***	0.0569***	0.0677***	0.0782***
	(2.95)	(3.70)	(3.42)	(3.79)	(3.90)	(4.05)
Aggregate Demand (Specialists + Generalists)	0.0307***	0.0588***	0.0782***	0.1001***	0.1206***	0.1384***
	(5.21)	(6.47)	(6.29)	(6.60)	(6.80)	(6.99)
Return	-0.0222***	-0.0242***	-0.0198***	-0.0245***	-0.0259***	-0.0209***
	(-8.49)	(-6.18)	(-4.23)	(-5.22)	(-5.26)	(-3.95)
Lagged 6-month cumulative return	0.0014**	0.0037***	0.0067***	0.0077***	0.0093***	0.0097***
	(2.43)	(3.47)	(4.33)	(3.86)	(3.80)	(3.36)
TobinQ	-0.0002	-0.0014***	-0.0025***	-0.0033***	-0.0039***	-0.0048***
	(-1.46)	(-4.77)	(-6.18)	(-6.28)	(-6.10)	(-6.34)
ROA	-0.0238*	-0.0643***	-0.1009***	-0.1420***	-0.1834***	-0.2264***
	(-1.92)	(-2.90)	(-3.22)	(-3.58)	(-3.95)	(-4.19)
FCF	0.0276**	0.0691***	0.1050***	0.1438***	0.1810***	0.2164***
	(2.12)	(2.98)	(3.20)	(3.43)	(3.68)	(3.77)
Size	-0.0003*	-0.0002	-0.0000	0.0002	0.0004	0.0003
	(-1.89)	(-0.57)	(-0.01)	(0.36)	(0.64)	(0.44)
FixedAssets	0.0043*	0.0083**	0.0120**	0.0157**	0.0194***	0.0237***
	(1.92)	(2.18)	(2.31)	(2.48)	(2.78)	(3.28)
Capex	-0.0160**	-0.0260*	-0.0399**	-0.0552**	-0.0633**	-0.0710**
	(-2.15)	(-1.92)	(-2.05)	(-2.22)	(-2.16)	(-2.11)
Leverage	-0.0083***	-0.0156***	-0.0210***	-0.0250***	-0.0277***	-0.0303***
	(-4.68)	(-4.54)	(-4.15)	(-3.79)	(-3.41)	(-3.15)
Year–month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	424312	421117	417862	414516	411073	407609
R sq	0.009	0.008	0.008	0.008	0.008	0.008

Table 5: Inattention of each category's specialists and market reactions to news

This table examines how specialists' inattention affects market reactions to stock news by estimating the following regression:

$$DGTW_{i,c,t:\,t+x} = \alpha + \beta_1 \times ESS_{i,c,t} + \beta_2 \times Inattention_{c,t} + \gamma_1 \times ESS_{i,c,t} \times Inattention_{c,t} + \ \delta \times M_{i,t} + \epsilon_{i,c,t:t+x},$$

where $DGTW_{i,c,t:t+x}$ refers to the DGTW-adjusted abnormal return of stock i accumulated over the period from days t to $t + x$ following news in category c, released on day t by firm i; $ESS_{i,c,t}$ *is* the sentiment of firm i's news in category c from *RavenPack*; and $Inattention_{c,t}$ refers to inattention of the specialists who typically process news in category c that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. In all regressions, we include firm-level control variables ($M_{i,t}$), firm fixed effects, news category fixed effects, and time fixed effects. T statistics are reported in parentheses and are based on standard errors clustered by firm. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent variables:	DGTW-adjusted return$_{i,c,t:t+x}$						
	t	t + 1	[t+1, t+2]	[t+1, t+3]	[t+1, t+5]	[t+1, t+10]	[t+1, t+20]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ESS$_{i,c,t}$	0.0098***	0.0014***	0.0015***	0.0015***	0.0014***	0.0013**	0.0016**
	(16.48)	(6.77)	(6.10)	(4.80)	(3.78)	(2.35)	(2.20)
Inattention$_{c,t}$	-0.0001	0.0000	0.0002	0.0002	0.0005*	0.0012**	0.0019**
	(-0.29)	(0.26)	(1.03)	(0.81)	(1.72)	(2.37)	(2.05)
ESS$_{i,c,t}$ * Inattention$_{c,t}$	-0.0012***	0.0002***	0.0003***	0.0003***	0.0005***	0.0006***	0.0006**
	(-6.11)	(2.83)	(3.64)	(2.93)	(3.38)	(3.03)	(2.47)
Volatility	-0.0004*	-0.0004*	-0.0005	-0.0005	-0.0005	-0.0011	-0.0019
	(-1.71)	(-1.67)	(-1.27)	(-0.82)	(-0.56)	(-0.89)	(-1.13)
TobinQ	-0.0008***	-0.0004***	-0.0006***	-0.0009***	-0.0013***	-0.0023***	-0.0038***
	(-8.04)	(-5.59)	(-6.19)	(-6.57)	(-5.97)	(-6.42)	(-5.73)
ROA	0.0006	0.0011	0.0040	0.0025	0.0079	0.0131	0.0382
	(0.14)	(0.23)	(0.59)	(0.30)	(0.68)	(0.75)	(1.45)
FCF	-0.0040	-0.0023	-0.0047	-0.0027	-0.0088	-0.0143	-0.0388
	(-0.94)	(-0.47)	(-0.69)	(-0.32)	(-0.76)	(-0.81)	(-1.50)
Size	-0.0038***	-0.0025***	-0.0039***	-0.0051***	-0.0073***	-0.0126***	-0.0218***
	(-14.01)	(-11.80)	(-12.24)	(-12.95)	(-11.66)	(-12.41)	(-12.59)
FixedAssets	0.0011	0.0004	0.0011	0.0012	0.0011	-0.0002	0.0005
	(1.53)	(0.69)	(1.23)	(0.99)	(0.61)	(-0.08)	(0.10)
Capex	-0.0080***	-0.0052**	-0.0061**	-0.0069*	-0.0116**	-0.0178**	-0.0265*
	(-3.53)	(-2.56)	(-1.98)	(-1.87)	(-2.26)	(-2.15)	(-1.95)
Leverage	0.0035***	0.0025***	0.0039***	0.0049***	0.0058***	0.0122***	0.0241***
	(3.93)	(3.00)	(3.28)	(3.38)	(2.87)	(3.74)	(4.58)
Lagged 6-m cumulative return	-0.0016***	-0.0011***	-0.0019***	-0.0027***	-0.0036***	-0.0054***	-0.0082***
	(-5.13)	(-5.15)	(-4.59)	(-4.74)	(-4.10)	(-3.90)	(-3.63)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	945363	944724	944158	943752	943194	940267	933779
R sq	0.041	0.016	0.017	0.018	0.021	0.027	0.040

Table 6: Inattention of generalists and market reactions to news (placebo test)

This table examines how generalists' inattention affects market reactions to stock news by estimating the following regression:

$$DGTW_{i,c,t:\,t+x} = \alpha + \beta_1 \times ESS_{i,c,t} + \beta_2 \times Inattention_{c,t} + \gamma_1 \times ESS_{i,c,t} \times Inattention_{c,t} + \ \delta \times M_{i,t} + \epsilon_{i,c,t:t+x},$$

where $DGTW_{i,c,t:t+x}$ refers to the DGTW-adjusted abnormal return of stock i accumulated over the period from days t to $t + x$ following news in category c, released on day t by firm i; $ESS_{i,c,t}$ *is* the sentiment of firm i's news in category c from *RavenPack*; and $Inattention_{c,t}$ refers to inattention of the generalists who typically process news in category c that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. In all regressions, we include firm-level control variables ($M_{i,t}$), firm fixed effects, news category fixed effects, and time fixed effects. T statistics are reported in parentheses and are based on standard errors clustered by firm. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent variables:	DGTW-adjusted return$_{i,c,t:t+x}$						
	t	t + 1	[t+1, t+2]	[t+1, t+3]	[t+1, t+5]	[t+1, t+10]	[t+1, t+20]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ESS$_{i,c,t}$	0.0053***	0.0017***	0.0019***	0.0022***	0.0021***	0.0026***	0.0032***
	(9.62)	(8.55)	(7.15)	(6.49)	(5.08)	(4.85)	(3.91)
Inattention$_{c,t}$	0.0001	-0.0000	0.0002	0.0002	0.0006*	0.0010	0.0007
	(0.77)	(-0.02)	(0.89)	(0.75)	(1.86)	(1.53)	(0.62)
ESS$_{i,c,t}$ * Inattention$_{c,t}$	0.0003*	0.0001	0.0001	0.0001	0.0002	0.0001	0.0000
	(1.77)	(1.58)	(1.62)	(0.91)	(1.50)	(0.70)	(0.15)
Volatility	-0.0004*	-0.0004	-0.0005	-0.0005	-0.0004	-0.0010	-0.0018
	(-1.82)	(-1.65)	(-1.23)	(-0.80)	(-0.52)	(-0.82)	(-1.06)
TobinQ	-0.0008***	-0.0004***	-0.0006***	-0.0009***	-0.0013***	-0.0023***	-0.0038***
	(-8.02)	(-5.59)	(-6.19)	(-6.57)	(-5.97)	(-6.43)	(-5.74)
ROA	0.0005	0.0011	0.0040	0.0025	0.0079	0.0131	0.0381
	(0.12)	(0.23)	(0.59)	(0.30)	(0.68)	(0.74)	(1.45)
FCF	-0.0039	-0.0023	-0.0047	-0.0027	-0.0088	-0.0142	-0.0386
	(-0.92)	(-0.47)	(-0.69)	(-0.32)	(-0.75)	(-0.80)	(-1.50)
Size	-0.0038***	-0.0025***	-0.0039***	-0.0051***	-0.0073***	-0.0127***	-0.0218***
	(-13.82)	(-11.79)	(-12.24)	(-12.95)	(-11.68)	(-12.44)	(-12.62)
FixedAssets	0.0011	0.0004	0.0011	0.0012	0.0010	-0.0003	0.0004
	(1.55)	(0.68)	(1.22)	(0.98)	(0.59)	(-0.10)	(0.09)
Capex	-0.0080***	-0.0052**	-0.0061*	-0.0069*	-0.0115**	-0.0177**	-0.0263*
	(-3.50)	(-2.55)	(-1.97)	(-1.86)	(-2.25)	(-2.13)	(-1.94)
Leverage	0.0034***	0.0025***	0.0039***	0.0050***	0.0059***	0.0123***	0.0242***
	(3.87)	(3.00)	(3.28)	(3.39)	(2.88)	(3.76)	(4.60)
Lagged 6-m cumulative return	-0.0016***	-0.0011***	-0.0019***	-0.0027***	-0.0036***	-0.0055***	-0.0082***
	(-5.08)	(-5.16)	(-4.60)	(-4.75)	(-4.10)	(-3.90)	(-3.63)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	945363	944724	944158	943752	943194	940267	933779
R sq	0.040	0.016	0.017	0.018	0.021	0.027	0.040

Table 7: Complementarity induced by inattention among other specialists

This table tests the complementarity induced by inattention among other specialists. Panel A shows how inattention among other specialists affects market reactions to news in a category when the inattention of specialists is low. $DGTW_{i,c,t:t+x}$ refers to the DGTW-adjusted abnormal return of stock i accumulated over the period from days t to $t + x$ following news in category *c*, released on day *t* by firm *i*; $ESS_{i,c,t}$ *is* the sentiment of firm *i*'s news in category *c* from *RavenPack*; and $Inattention_{-c,t}$ is the inattention of other specialists who did not cover news category *c* due to abnormal stock returns in their portfolio holdings in the previous month. The sample only includes the news events of nonsalient stocks, and we further restrict the sample to events when the focal specialists experience no inattention shock (i.e., their inattention is in the bottom-decile). Panel B reports how the inattention of other specialists affects the trading of focal specialists when the focal specialists experience no inattention shock. The dependent variables are the total trades of focal specialists on stock *i* over days $t + x$ following stock *i*'s news in category *c* in columns (1) to (4) and the total trades of focal specialists on other stocks that have no category *c* news over days $t + x$ following stock *i*'s news in category *c* in columns (5) to (8). $Inattention_{-c,t}$ is the inattention of *other* specialists who did not cover news category *c* in the previous month. T statistics are reported in parentheses and are based on standard errors clustered by firm. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Panel A: Stock returns after news

Dependent variables:	DGTW-adjusted return$_{i,c,t:t+x}$						
	t	t + 1	[t+1, t+2]	[t+1, t+3]	[t+1, t+5]	[t+1, t+10]	[t+1, t+20]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ESS$_{i,c,t}$	0.0118***	0.0028***	0.0032***	0.0035***	0.0027***	0.0029**	0.0034**
	(6.12)	(4.86)	(4.32)	(4.38)	(3.08)	(2.53)	(2.33)
Inattention$_{-c,t}$	0.0012**	0.0003	0.0001	-0.0000	-0.0012	-0.0011	-0.0009
	(2.04)	(0.59)	(0.22)	(-0.00)	(-1.21)	(-0.75)	(-0.36)
ESS$_{i,c,t}$ * Inattention$_{-c,t}$	-0.0013*	-0.0005**	-0.0005**	-0.0006**	-0.0003	-0.0003	-0.0005
	(-1.90)	(-2.52)	(-2.08)	(-2.36)	(-0.80)	(-0.59)	(-0.97)
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	117863	117867	117867	117867	117867	117867	117867
R sq	0.099	0.058	0.060	0.060	0.064	0.075	0.093

Panel B: Specialists' trading after news

Dependent variables:	Trades of focal specialists on stock *i*				Trades of focal specialists on other stocks			
	day [0, 0]	day [0, 3]	day [0, 5]	day [0, 10]	day [0, 0]	day [0, 3]	day [0, 5]	day [0, 10]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inattention$_{-c,t-1}$	-0.0039**	-0.0065*	-0.0092	-0.0044	-0.0168	-0.0104	-0.0098	-0.0453
	(-1.98)	(-1.71)	(-1.17)	(-0.34)	(-0.97)	(-0.62)	(-0.58)	(-0.59)
Firm-level controls	Yes	Yes	Yes	Yes	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes	No	No	No	No
News category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	117660	117660	117660	117660	117660	117660	117660	117660
R sq	0.078	0.108	0.118	0.146	0.376	0.387	0.273	0.054

Table 8: Specialists inattention and anomaly returns

This table examines how specialists' inattention affects anomaly returns on news days by estimating the following regression:

$$Return_{i,c,t} = \alpha + \beta_1 \times Net_{i,t} + \beta_2 \times Inattention_{c,t} + \gamma_1 \times Net_{i,t} \times Inattention_{c,t} + \delta \times M_{i,t} + \epsilon_{i,c,t},$$

where $Return_{i,c,t}$ refers to the return of stock i on day t following news in category c. The news-day DGTW-adjusted returns returns are used in Columns (1) to (4), and the news-day unadjusted-adjusted returns are used in Columns (5) to (8). $Net_{i,t}$ *is* constructed following Engelberg, McLean, and Pontiff (2018). For each stock-month observation, we first create a Long (Short) measure by summing the number of long-side (short-side) anomaly portfolios that the stock belongs to, from a list of 212 characteristics. Net is equal to Long minus Short and is measured at the beginning of each month. $Inattention_{c,t}$ refers to the inattention of the specialists who typically process the news in category c that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. In all regressions, we include control variables of lagged 1-month stock return and lagged 1-month volatility. Year-month fixed effects, news category fixed effects, and date fixed effects are included in different specifications. T statistics are reported in parentheses and are based on standard errors clustered by time. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent	News-day DGTW-adjusted $return_{i,c,t}$				News-day unadjusted $return_{i,c,t}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Net_{i,t}$	0.0010***	0.0010***	0.0021***	0.0020***	0.0010***	0.0010***	0.0024***	0.0023***
	(7.01)	(7.10)	(4.88)	(4.82)	(6.54)	(6.92)	(4.89)	(4.78)
$Inattention_{c,t}$			-0.0001	0.0002			-0.0000	0.0002
			(-0.61)	(0.60)			(-0.18)	(0.51)
$Net_{i,c,t}$ * $Inattention_{c,t}$			-0.0004***	-0.0004***			-0.0005***	-0.0004***
			(-3.12)	(-3.02)			(-3.41)	(-3.14)
Lagged return	-0.0008	-0.0007	-0.0009	-0.0008	-0.0007	-0.0006	-0.0007	-0.0006
	(-0.56)	(-0.49)	(-0.58)	(-0.52)	(-0.43)	(-0.36)	(-0.46)	(-0.38)
Lagged volatility	0.0010	0.0010	0.0010	0.0010	0.0011	0.0011	0.0011	0.0011
	(1.30)	(1.27)	(1.30)	(1.28)	(1.34)	(1.36)	(1.35)	(1.36)
Year–month FE	Yes	No	Yes	No	Yes	No	Yes	No
News category FE	No	Yes	No	Yes	No	Yes	No	Yes
Date FE	No	Yes	No	Yes	No	Yes	No	Yes
N	945238	945238	945238	945238	945238	945238	945238	945238
R sq	0.005	0.086	0.005	0.086	0.001	0.010	0.002	0.010

Table 9: Alternative explanations and characteristics-adjusted data specialization

This table examines the performance of characteristics-adjusted data specialization and its implications for market efficiency. We regress the news category concentration index of each investor each quarter on a set of their contemporaneous characteristics, calculate the characteristics-adjusted index as the residual from the regression, and use it to reclassify investors. Panel A reports the characteristics of the monthly portfolio performance for each quintile. Panel B Columns (1) to (3) report the effects of focal specialists' inattention on market reactions to news. $Inattention_{c,t}$ refers to inattention of the specialists who typically process news in category c that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the news events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. Panel B Columns (4) to (6) report the effects of other specialists' inattention on market reactions to news. $Inattention_{-c,t}$ is the inattention of other specialists who did not cover news category c in the previous month, and we further restrict the sample to events when the focal specialists experience no inattention shock (i.e., their inattention is in the bottom-decile). T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Panel A: Portfolio performance of generalists vs. specialists (characteristics-adjusted classification)

	Quintile 1 (Generalists)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Specialists)	Difference (5 minus 1)
Average monthly return	1.44%***	1.45%***	1.57%***	1.71%***	1.90%***	0.46%***
	(3.11)	(3.16)	(3.43)	(3.66)	(3.87)	(4.84)
Average monthly return (DGTW-adjusted)	1.05%***	1.00%***	1.06%***	1.14%***	1.29%***	0.24%***
	12.67)	(12.64)	(13.32)	(13.56)	(14.64)	(4.17)
σ	0.05	0.05	0.05	0.05	0.06	0.01
α	1.26%***	1.27%***	1.39%***	1.53%***	1.72%***	0.46%***
	(11.10)	(11.43)	(11.59)	(10.92)	(11.02)	(4.90)
β	1.04***	1.03***	1.02***	1.03***	1.08***	0.04*
	(44.94)	(45.49)	(41.73)	(36.06)	(33.82)	(1.87)
SR	0.24	0.24	0.26	0.28	0.30	0.24
Skewness	-0.26	-0.18	-0.24	-0.25	-0.20	1.57
Kurtosis	3.37	3.29	3.12	3.08	3.15	9.26

Panel B: Stock returns after news

Dependent variables:	DGTW-adjusted $return_{i,c,t:t+x}$				DGTW-adjusted $return_{i,c,t:t+x}$		
	t	[t+1, t+3]	[t+1, t+20]		t	[t+1, t+3]	[t+1, t+20]
	(1)	(2)	(3)		(4)	(5)	(6)
$ESS_{i,c,t}$	0.0095***	0.0010**	0.0013	$ESS_{i,c,t}$	0.0125***	0.0036***	0.0037***
	(13.99)	(2.58)	(1.33)		(8.91)	(6.05)	(2.91)
$Inattention_{c,t}$ (focal)	-0.0001	0.0002	0.0015	$Inattention_{-c,t}$ (other)	0.0006	-0.0003	-0.0025
	(-0.69)	(0.98)	(1.41)		(1.02)	(-0.51)	(-1.10)
$ESS_{i,c,t}$*$Inattention_{c,t}$	-0.0009***	0.0004***	0.0006**	$ESS_{i,c,t}$ * $Inattention_{-c,t}$	-0.0020***	-0.0008***	-0.0007
	(-4.94)	(3.47)	(2.09)		(-3.09)	(-3.23)	(-1.16)
Firm-level controls	Yes	Yes	Yes	Firm-level controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Firm FE	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	News category FE	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Year–month FE	Yes	Yes	Yes
N	945363	943752	933779	N	115449	115456	115456
R sq	0.041	0.018	0.040	R sq	0.098	0.059	0.090

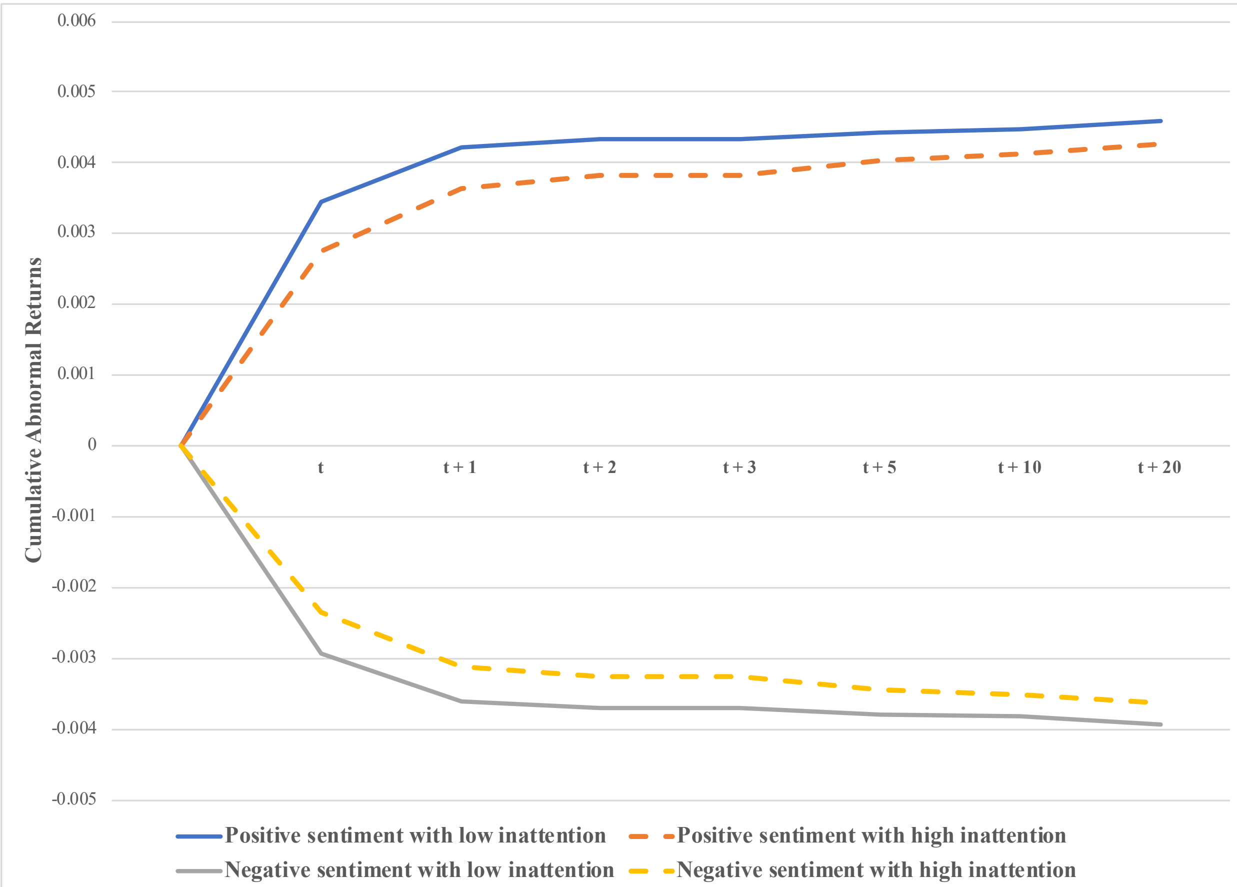
Table 10: A market timing test

This table presents the results of the joint tests that examine the market timing ability of generalists and specialists. The regression specifications follow Daniel and Moskowitz (2016). The dependent variables are the real monthly portfolio returns of generalists (in columns (1) and (2)) and specialists (in columns (3) and (4)) from 2000:04 to 2010:12. The independent variables are a constant; an indicator for bear markets, $I_{B,t-1}$, which equals one if the cumulative past two-year return on the market is negative; the excess market return, $R^e_{m,t}$; and a contemporaneous up-market indicator, $I_{U,t}$, which equals one if $R^e_{m,t}$ is positive. T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

		Portfolio return of generalists		Portfolio return of specialists	
Coefficients	Variables	(1)	(2)	(3)	(4)
α_0	1	0.0055***	0.0055***	0.0111***	0.0111***
		(6.78)	(7.11)	(7.03)	(7.00)
α_B	$I_{B,t-1}$	0.0005	-0.0046**	0.0017	0.0015
		(0.43)	(-2.51)	(0.68)	(0.39)
β_0	$R^e_{m,t}$	0.6520***	0.6520***	0.7053***	0.7053***
		(30.62)	(32.14)	(16.90)	(16.84)
β_B	$I_{B,t-1}$*$R^e_{m,t}$	0.0189	-0.0710**	0.1243**	0.1212*
		(0.71)	(-2.03)	(2.40)	(1.68)
$\beta_{B,U}$	$I_{B,t-1}$* $I_{U,t}$*$R^e_{m,t}$		0.2025***		0.0070
			(3.70)		(0.06)
N		129	129	129	129
R sq		0.957	0.961	0.891	0.891

Figure 1: Specialists' inattention and postnews price drift

This figure demonstrates the effect of specialists' inattention on market reactions to news. We plot the cumulative abnormal returns for four groups separately from days *t* to *t + 20*, where the four groups are 1) news with positive sentiment and low specialist inattention (at the 25% quantile value of the distribution); 2) news with positive sentiment and high specialist inattention (at the 75% quantile value); 3) news with negative sentiment and low specialist inattention; and 4) news with negative sentiment and high specialist inattention. To estimate the conditional returns, we use the estimated regression coefficients from Table 5, realized sentiment values, and the respective inattention quantile values.



Online Appendix

Data Specialists and Market Efficiency

Table A1: News-related trading in each news category

This table presents the frequency and intensity of news-related trading of generalists and specialists in each news category. The categorization of each news is based on the *RavenPack* database, which assigns a category to each news of every stock. We label a trade of institutional investors from the *ANcerno* database as *news-related* if there is news reported for the stock (firm) either on the day of the trade or the day before, and classify each news-related trade into one of the 15 news categories accordingly. Then, we construct a news category concentration index for each investor each quarter as the sum of the squared percentages of news-related trades in each news category. We refer to investors in the highest (lowest) concentration quintile as data *specialists* (*generalists*). In this table, Column 2 presents the number of stock-day with news-related trading by generalists in each news category. Column 3 presents the average news-related trading among generalists that trade on the stock, expressed as a percentage of each generalist's monthly trading dollars. Similarly, Column 4 presents the number of stock-day with news-related trading by specialists in each news category, and Column 5 presents the average news-related trading among specialists that trade on the stock, expressed as a percentage of each specialist's monthly trading dollars.

	Generalists		Specialists	
News categories	# of stock-day with news-related trading	Average news-related trading (% of monthly trading dollar)	# of stock-day with news-related trading	Average news-related trading (% of monthly trading dollar)
Insider trading	76,885	0.35%	78,298	0.71%
Order imbalances	149,115	0.43%	44,406	1.28%
Earnings	40,889	0.54%	39,877	0.79%
Products services	57,816	0.50%	32,104	0.91%
Analyst ratings	57,197	0.53%	34,233	0.94%
Revenues	31,917	0.56%	28,629	0.81%
Labor issues	27,580	0.52%	17,978	0.82%
Equity actions	15,164	0.63%	11,895	0.91%
Acquisitions mergers	15,107	0.63%	10,485	0.96%
Investor relations	7,528	0.43%	8,568	0.74%
Dividends	9,639	0.49%	7,377	0.76%
Credit ratings	13,722	0.52%	6,004	0.95%
Stock prices	12,711	0.75%	7,395	1.16%
Price targets	7,691	0.45%	1,590	0.86%

Table A2: Daily trades initiated by specialists/generalist and stock returns

This table examines if the signals from specialists and generalists' trades can predict daily stock daily returns. More explicitly, we estimate the following regression using daily stock-level observations:

$$DGTW_{i,t} = \alpha + \beta_1 \times Specialists\ buy_{i,[t-5,t-1]} + \beta_2 \times Specialists\ sell_{i,[t-5,t-1]} + \beta_3 \times Generalists\ buy_{i,[t-5,t-1]} + \beta_4 \times Generalists\ sell_{i,[t-5,t-1]} + \delta \times M_{i,t-1} + \epsilon_{i,t},$$

where $DGTW_{i,t+x}$ refers to the DGTW-adjusted abnormal return of stock *i* on day *t*. For both specialists and generalists, we aggregate their trades for stock *i* within a rolling window of five trading days. The dummy variable $Specialists\ buy_{i,[t-5,t-1]}$ is equal to one if the aggregate trading dollars of specialists on stock *i* are positive over the period from day *t – 5* and *t – 1*. The dummy variable $Specialists\ sell_{i,[t-5,t-1]}$ is equal to one if the aggregate trading dollars of specialists on stock *i* are negative over the period from day *t – 5* and *t – 1*. Both $Specialists\ buy_{i,[t-5,t-1]}$ and $Specialists\ sell_{i,[t-5,t-1]}$ are equal to zero if specialists have no trades on stock *i* over the period from day *t – 5* and *t – 1*. Similarily, $Generalists\ buy_{i,[t-5,t-1]}$ and $Generalists\ sell_{i,[t-5,t-1]}$ are constructed based on the aggregate trading dollars of generalist. The control variables ($M_{i,t-1}$) include lagged 5-day returns, volatility and CRSP trading volume. Date fixed effects and stock fixed effects are included in all the specifications. T statistics are reported in parentheses and are based on standard errors clustered by firm. Superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent variable	DGTW-adjusted return$_{i,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Specialists buy$_{i,[t-5,t-1]}$	0.0001***	0.0001***			0.0001***	0.0001***
	(3.94)	(4.66)			(4.01)	(4.73)
Specialists sell$_{,[t-5,t-1]}$	-0.0003***	-0.0004***			-0.0003***	-0.0003***
	(-9.05)	(-12.08)			(-8.87)	(-11.86)
Generalists buy$_{i,[t-5,t-1]}$			-0.0001	0.0000	-0.0000	0.0000
			(-1.38)	(0.04)	(-1.27)	(0.20)
Generalists sell$_{,[t-5,t-1]}$			-0.0004***	-0.0005***	-0.0004***	-0.0005***
			(-10.11)	(-14.01)	(-9.91)	(-13.74)
Lagged returns		-0.1179***		-0.1179***		-0.1181***
		(-99.51)		(-99.50)		(-99.61)
Lagged volatility		0.0076***		0.0076***		0.0077***
		(9.38)		(9.38)		(9.38)
Lagged CRSP trading volume		0.0025		0.0028		0.0027
		(1.19)		(1.30)		(1.27)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9287792	9281699	9287792	9281699	9287792	9281699
R sq	0.002	0.006	0.002	0.006	0.002	0.006

Table A3: Characteristics-adjusted data specialization

This table reports the first step of constructing the characteristics-adjusted news category concentration index. We regress the news category concentration index on contemporaneous characteristics of each institutional investor each quarter, including portfolio size, quarterly trading volume, quarter-end holding rebalance, the number of days with news-related trading within the quarter, the trading reliance on news, the industry concentration (Kacperczyk, Sialm, and Zheng 2005), the R squared measure of Amihud and Goyenko (2013), and the active share (Cremers and Petajisto, 2009). The regression results are reported in Columns (1) to (9). T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Dependent variables:	News category concentration index								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolio size	-0.0104***								0.0065***
	(-31.60)								(5.23)
Dollar trading volume		-0.0107***							-0.0046***
		(-30.67)							(-3.58)
% of holding rebalance			-0.0106**						0.0965***
			(-1.97)						(6.29)
# of days with news-related trading				-0.0005***					-0.0005***
				(-16.14)					(-13.24)
Trading reliance on news					-2.0325***				-1.7669***
					(-46.40)				(-34.52)
Industry concentration						0.2199***			0.0681***
						(38.66)			(10.56)
R_squared							-0.2421***		-0.0572***
							(-33.31)		(-8.17)
Active share								0.4485***	0.1846***
								(34.78)	(17.30)
N	47846	47848	47848	47848	47848	47817	42359	47848	42335
R sq	0.023	0.025	0.000	0.007	0.213	0.090	0.064	0.036	0.247

Table A4: DGTW-adjusted returns of double-sorted portfolios

This table presents the average DGTW-adjusted portfolio returns of 25 double-sorted groups of investors. In Panel A, we double-sort all investors into 25 groups in each quarter according to quintiles of their news concentration index values and quintiles based on their trading reliance on news from low to high. We divide an investor's news-related trading by her total trades (including news- and nonnews-related trades) as a proxy for her trading reliance on firm news. In Panel B, we independently double-sort all investors into 25 groups in each quarter according to quintiles of their news category concentration index values and quintiles based on the industry concentration of their portfolio from low to high. The industry concentration of each investor's portfolio is calculated following Kacperczyk, Sialm and Zheng (2005). T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Panel A: Double-sorting based on trading reliance on news

		Quintile by trading reliance on news				
		1	2	3	4	5
Quintile by news category concentration	1 (Generalist)	0.93%***	1.05%***	0.93%***	0.86%***	0.89%***
		(3.93)	(9.41)	(11.51)	(9.98)	(9.10)
	2	1.36%***	1.19%***	1.11%***	0.85%***	0.77%***
		(7.07)	(11.15)	(10.60)	(11.38)	(9.26)
	3	1.61%***	1.33%***	1.07%***	0.92%***	0.89%***
		(11.85)	(12.57)	(12.33)	(11.11)	(9.23)
	4	1.64%***	1.45%***	1.12%***	0.96%***	0.86%***
		(13.24)	(13.51)	(12.24)	(10.09)	(8.01)
	5 (Specialist)	1.54%***	1.45%***	1.27%***	1.13%***	1.23%***
		(12.23)	(13.15)	(11.53)	(9.92)	(9.16)
	Difference	0.62%**	0.40%***	0.34%***	0.27%**	0.34%***
	(5 minus 1)	(2.51)	(3.00)	(3.72)	(2.56)	(2.94)

Panel B: Double-sorting based on industry concentration

		Quintile by industry concentration				
		1	2	3	4	5
Quintile by news category concentration	1 (Generalist)	0.82%***	0.86%***	0.89%***	0.98%***	0.98%***
		(9.85)	(9.28)	(10.75)	(11.01)	(8.40)
	2	0.88%***	0.90%***	0.98%***	1.05%***	1.01%***
		(11.82)	(11.30)	(10.78)	(10.02)	(9.57)
	3	1.03%***	1.10%***	1.16%***	1.16%***	1.19%***
		(13.08)	(13.03)	(12.41)	(10.61)	(9.31)
	4	1.16%***	1.31%***	1.39%***	1.34%***	1.21%***
		(13.84)	(14.36)	(13.30)	(12.28)	(9.84)
	5 (Specialist)	1.26%***	1.32%***	1.46%***	1.46%***	1.43%***
		(12.92)	(14.29)	(13.07)	(13.05)	(10.59)
	Difference	0.44%***	0.46%***	0.57%***	0.49%***	0.45%***
	(5 minus 1)	(5.68)	(6.39)	(6.36)	(5.62)	(3.08)

Table A5: Thirteen news categories and characteristics-adjusted data specialization

In this table, we redefine the fifteen news categories by combining closely related pairs (i.e., "earnings" and "revenues", and "analysts rating" and "price targets") to create an alternative set of thirteen categories, and then reclassify investors into five quintiles following previous steps. Panel A reports the characteristics of the monthly portfolio performance for each quintile. Panel B Columns (1) to (3) report the effects of focal specialists' inattention on market reactions to news. $Inattention_{c,t}$ refers to inattention of the specialists who typically process news in category *c* that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the news events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. Panel B Columns (4) to (6) report the effects of *other* specialists' inattention on market reactions to news. $Inattention_{-c,t}$ is the inattention of *other* specialists who did not cover news category *c* in the previous month, and we further restrict the sample to events when the focal specialists experience no inattention shock (i.e., their inattention is in the bottom-decile). T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Panel A: Portfolio performance of generalists vs. specialists

	Quintile 1 (Generalists)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Specialists)	Difference (5 minus 1)
Average monthly return	1.41%***	1.42%***	1.55%***	1.74%***	1.93%***	0.52%***
	(3.07)	(3.10)	(3.41)	(3.70)	(3.93)	(5.29)
Average monthly return (DGTW-adjusted)	1.03%***	0.98%***	1.07%***	1.15%***	1.31%***	0.28%***
	(12.31)	(12.53)	(13.19)	(13.60)	(14.99)	(4.92)
σ	5.24%	5.21%	5.18%	5.32%	5.59%	1.11%
α	1.23%***	1.24%***	1.37%***	1.56%***	1.75%***	0.52%***
	(10.94)	(11.49)	(11.46)	(10.94)	(11.09)	(5.38)
β	1.04***	1.03***	1.02***	1.04***	1.08***	0.04**
	(44.94)	(46.74)	(41.56)	(35.58)	(33.42)	(2.20)
SR	0.23	0.23	0.26	0.29	0.31	0.28
Skewness	-0.26	-0.17	-0.24	-0.25	-0.20	1.01
Kurtosis	3.38	3.21	3.18	3.06	3.17	6.19

Panel B: Stock returns after news

Dependent variables:	DGTW-adjusted return$_{i,c,t:t+x}$				DGTW-adjusted return$_{i,c,t:t+x}$		
	t	[t+1, t+3]	[t+1, t+20]		t	[t+1, t+3]	[t+1, t+20]
	(1)	(2)	(3)		(4)	(5)	(6)
$ESS_{i,c,t}$	0.0104***	0.0007*	0.0009	$ESS_{i,c,t}$	0.0111***	0.0021***	0.0026*
	(14.77)	(1.96)	(1.14)		(6.61)	(2.86)	(1.86)
$Inattention_{c,t}$ (focal)	-0.0001	0.0003	0.0015*	$Inattention_{-c,t}$ (other)	0.0010*	0.0004	0.0000
	(-0.32)	(1.65)	(1.78)		(1.89)	(0.66)	(0.00)
$ESS_{i,c,t}$*$Inattention_{c,t}$	-0.0011***	0.0005***	0.0007***	$ESS_{i,c,t}$ * $Inattention_{-c,t}$	-0.0014**	-0.0003	-0.0003
	(-5.76)	(5.12)	(2.92)		(-2.26)	(-1.04)	(-0.53)
Firm-level controls	Yes	Yes	Yes	Firm-level controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Firm FE	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	News category FE	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Year–month FE	Yes	Yes	Yes
N	855474	853889	844339	N	114895	114898	114898
R sq	0.044	0.017	0.040	R sq	0.097	0.062	0.094

Table A6: Alternative classification of data specialization

In this table, we construct an alternative news category specialization measures based on the dollar amount of news-related trading and then reclassify investors into five quintiles following previous steps. Panel A reports the characteristics of the monthly portfolio performance for each quintile. Panel B Columns (1) to (3) report the effects of focal specialists' inattention on market reactions to news. $Inattention_{c,t}$ refers to inattention of the specialists who typically process news in category c that arises due to abnormal stock returns in the portfolio holdings of the specialists. The sample only includes the news events of nonsalient stocks, that is, those that did not experience extreme stock returns in the previous month. Panel B Columns (4) to (6) report the effects of other specialists' inattention on market reactions to news. $Inattention_{-c,t}$ is the inattention of *other* specialists who did not cover news category c in the previous month, and we further restrict the sample to events when the focal specialists experience no inattention shock (i.e., their inattention is in the bottom-decile). T statistics are reported in parentheses, and superscripts of *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Panel A: Portfolio performance of generalists vs. specialists (alternative classification)

	Quintile 1 (Generalists)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Specialists)	Difference (5 minus 1)
Average monthly return	1.3%***	1.45%***	1.64%***	1.76%***	1.82%***	0.52%***
	(2.86)	(3.20)	(3.55)	(3.66)	(3.75)	(4.70)
Average monthly return (DGTW-adjusted)	0.94%***	1.00%***	1.14%***	1.23%***	1.35%***	0.41%***
	(11.81)	(13.05)	(13.42)	(13.89)	(14.12)	(6.38)
σ	5.16%	5.15%	5.24%	5.47%	5.52%	1.27%
α	1.12%***	1.27%***	1.46%***	1.58%***	1.64%***	0.52%***
	(11.30)	(11.33)	(11.26)	(10.80)	(10.33)	(4.73)
β	1.03***	1.02***	1.03***	1.06***	1.07***	0.04
	(50.90)	(44.50)	(38.79)	(35.48)	(32.78)	(1.62)
SR	0.21	0.24	0.27	0.29	0.29	0.25
Skewness	-0.20	-0.27	-0.23	-0.21	-0.20	0.76
Kurtosis	3.18	3.19	3.10	3.10	3.41	5.55

Panel B: Stock returns after news

Dependent variables:	DGTW-adjusted return$_{i,c,t:t+x}$				DGTW-adjusted return$_{i,c,t:t+x}$		
	t	[t+1, t+3]	[t+1, t+20]		t	[t+1, t+3]	[t+1, t+20]
	(1)	(2)	(3)		(4)	(5)	(6)
$ESS_{i,c,t}$	0.0100***	0.0019***	0.0022***	$ESS_{i,c,t}$	0.0131***	0.0054***	0.0057***
	(16.65)	(6.60)	(3.05)		(7.98)	(7.10)	(2.98)
$Inattention_{c,t}$ (focal)	-0.0001	0.0002	0.0020**	$Inattention_{-c,t}$ (other)	0.0024***	-0.0000	0.0021
	(-0.54)	(0.92)	(2.24)		(3.27)	(-0.03)	(0.65)
$ESS_{i,c,t}$*$Inattention_{c,t}$	-0.0014***	0.0002*	0.0004	$ESS_{i,c,t}$ * $Inattention_{-c,t}$	-0.0015**	-0.0012***	-0.0011*
	(-6.42)	(1.85)	(1.60)		(-2.36)	(-4.62)	(-1.67)
Firm-level controls	Yes	Yes	Yes	Firm-level controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Firm FE	Yes	Yes	Yes
News category FE	Yes	Yes	Yes	News category FE	Yes	Yes	Yes
Year–month FE	Yes	Yes	Yes	Year–month FE	Yes	Yes	Yes
N	945363	945363	945363	N	114756	114762	114762
R sq	0.041	0.018	0.040	R sq	0.100	0.061	0.092