Nowcasting firms' fundamentals: Evidence from the cloud*

Ran Chang† and Zhi Da‡

Current version: January 2022

Abstract

Cloud computing has been widely adopted by businesses around the world. Using a proprietary data set on firm-level cloud data records from 2013 to 2021 from China, we find that year-on-year quarterly cloud data growth (CDG) contains value-relevant information for firm fundamentals, earnings surprises, and innovation performance. Specifically, CDG positively predicts assets growth, sales growth, ROA, standardized unexpected earnings (SUE), and patent outcomes. CDG also forecasts stock returns, especially around future earnings announcements. A long-short portfolio by buying (selling) stocks with the high (low) CDGs generates a 9.0% risk-adjusted return annually. The predictive power of CDG is robust to various controls and subsample cuts, superior compared to other nowcasters, and holds in other Asian countries as well. Finally,we find cloud data to have the unintended consequence of facilitating insider trading.

JEL Classification: G10; G11; G12; G14.

Keywords: Alternative data, Cloud computing, Data growth, Expected returns, China

*We thank Samuel Hartzmark for his insightful and constructive comments. We also benefited from discussions with participants at the 3rd Shanghai Financial Forefront Symposium and seminar participants at Shanghai Jiao Tong University. Chang acknowledges financial support from the Shanghai Institute of International Finance and Economics and the Shanghai Pujiang Program.

†Antai College of Economics & Management, Shanghai Jiao Tong University, China, 200030. Phone: (86) 6293-3671, Email: r.zhang@sjtu.edu.cn.

‡Mendoza College of Business, University of Notre Dame, United States, 46556. Phone: (574) 631-0354, Email: zda@nd.edu.

Electronic copy available at: https://ssrn.com/abstract=3893300

1 Introduction

Data is ubiquitous in business. Companies collect, process and generate data on daily basis. In recent years, the exponential growth in data forced many companies to move them to the cloud. According to a 2020 cloud computing study by the International Data Group, 92% of organization's IT environment is at least somewhat in the cloud today and 81% of organizations have at least one application or a portion of their computing infrastructure in the cloud.[1] Intuitively, the growth in a firm's cloud data correlates with the growth in the business in real time. Since information regarding the firm's fundamentals is released with a delay, cloud data growth can serve as a powerful "nowcaster." Taking advantage of a proprietary dataset on firm-level cloud data records, we are the first to study the predictive power of a firm's cloud data growth on firms' fundamentals, stock returns and insider trading.

Given the importance of forecasting a firm's accounting and stock performance, the finance and accounting literature has examined a battery of "nowcasters," as more and more "alternative" data became available. Examples include product Google search (Da et al. [2011]), website traffic (Rajgopal et al. [2003]), customer ratings (Hirshleifer et al. [2018]), employer ratings (Green et al. [2019]), satellite images (Katona et al. [2018]; Zhu [2019]), and credit card usage (Zhu [2019]; Agarwal et al. [2021]). Empirically, we find the cloud data growth to have superior forecasting power than many of these existing nowcasters. This is because cloud compute technology enables the firms to more efficiently run their organizations, better serve their customers, and dramatically increase their overall profit margins.[2] In other words, a higher cloud data growth rate signals not only stronger fundamentals contemporaneously, but also greater earning power, above and beyond the current quarter.

The cloud data records for the period from 2013 to 2021 are obtained from a leading cloud computing platform in China which operates just like Amazon Wed Services (AWS). Cloud computing is also popular in China. According to the 2020 China Academy of Information and Communications Technology's cloud computing development survey report, the proportion of companies in China that have already used cloud computing reached 66.1% in 2019.[3] As of 2020, our data provider covers more than 50% of listed firms, 80% technology firms and 99% of cities in China. Not surpris-

[1] See https://www.idg.com/tools-for-marketers/2020-cloud-computing-study/

[2] The 2015 Dell study reveals that companies that invest in big data, cloud, mobility, and security enjoy up to 53% faster revenue growth than their competitors. See https://www.delltechnologies.com/en-us/blog/what-companies-growing-more-than-50-percent-faster-are-investing-in/

[3] See http://www.caict.ac.cn/kxyj/qwfb/bps/202007/t20200729_287361.htm

Electronic copy available at: https://ssrn.com/abstract=3893300

ingly, our sample firms are bigger, more profitable and have higher past returns, compared to the average Chinese firm. While information technology sector has the most number of firms (25%), our sample covers all major industries in China. The number of firms covered in our sample grew from around 300 in 2014 to more than 2000 in 2020, as evident in Panel A of Figure 1. Our sample also includes firms from other countries, which we examine as part of our international tests.

We calculate a quarterly cloud data growth (CDG) as the annual growth rate in the amount of cloud data of a firm in a quarter, relative to that in the same quarter last year. The year-on-year growth rate alleviates potential within-year seasonality in the cloud data size. While CDG of quarter $q+1$ can be observed at the end of that quarter, fundamental variables in that quarter are only released in the next quarter. The nowcasting power of the cloud data is evident even at the aggregate level. For example, Panel B of Figure 1 plots the quarterly cloud data size (in terabytes) for an average firm in our sample. The dips during the first two quarters of 2020 clearly demonstrate the impact of Covid-related lockdown in China. Indeed, Panel C plots the CDG, aggregated across all firms in our sample, against the corresponding year-on-year quarterly GDP growth rates in China (released with a delay). The correlation between the two series is more than 65%.

We confirm that CDG's nowcasting power holds up at the firm-level. We first use CDG in $q+1$ to predict (or nowcast) fundamental variables in $q+1$, after controlling for their lags and other stock characteristics and quarter-q accounting variables (which are only observable in $q+1$). We find CDG to have significant incremental nowcasting power of a firm's outputs. For example, a 10% increase in CDG predicts an increase of 7.65% in the return on assets (ROA), 3.02% in asset growth (AG), 0.71% in sales growth (SG), 3.08% in the growth rate of patents applied (PA), and 1.64% in the growth rate of patents granted (PG) during the same quarter. The fundamental predictive power of CDG goes beyond nowcasting. A 10% increase in $q+1$ CDG also predicts an increase of 5.22% in ROA, 2.31% in AG, 0.47% in SG, 1.92% in PA, 1.64% in PG during the next quarter ($q+2$), consistent with the notion that CDG contains information regarding the firm's earnings power in the long run.

We horserace CDG against a battery of alternative nowcasters. They include the year-on-year quarterly growth rates of search volume for firms' products (SEAG); firms' App visiting volume (APPG); firms' customer product ratings (CUSG), firms' employer ratings (EMPG), number of cars in firms' parking lots (CARG), and credit card spending on firms' products and services (SPEG). Compared to CDG, these alternative nowcasters cover less firms. Importantly, even after

Electronic copy available at: https://ssrn.com/abstract=3893300

controlling for them simultaneously, CDG remains significant in forecasting ROA, AG, SG, PA and PG in both the current and the next quarter. Compared to CDG, the forecasting power of the other nowcasters is more sporadic. For example, SEAG and CARG only predict ROA, AG and SG, while SPEG only predict PA and PG.

We then examine CDG's predictive power of earnings surprises and market reaction during earnings announcements in the next two quarters. We find CDG to have significant incremental predictive power on contemporaneous quarter's earnings which is released in the next quarter. A 10% increase in CDG predicts a standardized unexpected earnings (SUE) that is 2.14% higher. CDG also predicts the stock earnings announcement return in the next quarter. A 1% increase in CDG predicts earnings announcement window abnormal return (CAR) that is 2.13% higher, suggesting that CDG contains new information not fully processed by the market before the announcement. Again, the predictive power of CDG on SUE and CAR goes beyond the next quarter and remains positive and significant in quarter $q + 2$. In addition, it remains significant after controlling for other nowcasters.

Finally, we show that CDG has strong return predictive power. In a quarterly-rebalanced quintile portfolio sorting exercise, a long-short strategy that buys (sells) stocks in the top (bottom) CDG-quintile generates a monthly profit of 0.85% (value-weighted) or 1.20% (equal-weighted). The risk-adjusted value-weighted returns (alphas) are 0.75%, 0.70%, 0.68%, and 0.62% per month, respectively, for the Chinese q-factor model based on Hou et al. [2015], Chinese five-factor model based on Fama and French (2015), the LSY3 factor model and the LSY4 factor model of Liu et al. [2019], and remain highly significant. Two-third of the profit accrues during the earnings announcement month even though such a month accounts for only one third of a quarter. This result supports the view that CDG contains novel information about a firm's fundamentals and such information is incorporated into the price when they are released to the public during the earnings announcement. The return predictive power of CDG is long lasting. The long-short strategy continues to deliver positive returns up to a year and we do not observe long-run reversals beyond a year. The cumulative return patterns suggests that CDG's return predictability is unlikely driven by a persistent price pressure which eventually should be reverted.

Fama-MacBeth (1973) cross-sectional regression approach allows us to tease out the incremental return predictive power of CDG. CDG shows a robust, positive and statistically significant predictive power on future excess, industry-adjusted, and geographic-adjusted returns in multivariate Fama–MacBeth regressions when we control for a number of firm characteristics and risk

Electronic copy available at: https://ssrn.com/abstract=3893300

factors, including firm's size (SIZE), book-to-market ratio (BM), return on assets (ROA), leverage (LEV), PPE growth (PG), intangible growth (IG), one-month lagged return (STR), price momentum (MOM), earnings surprise (SUE), Amihud [2002] illiquidity measure (ILLIQ), idiosyncratic volatility (IVOL), turnover ratio (TO), analyst coverage (ANA), and institutional ownership (IO). The slope coefficient on CDG drops only slightly from 0.505 in the univariate case to 0.469 in the multivariate case when all other controls are included. Similarly, CDG remains significant when controlling for other nowcasters, either in an one-on-one horse race or in a multivariate setting where all other nowcasters are included simultaneously.

A battery of international tests and subsample analyses confirm the robustness of CDG's fundamental and return predictive power. For example, we find CDG to have strong fundamental and return predictive power among other Asian countries such as Indonesia, Japan, Malaysia, and Singapore. The predictive power is also present in various subsamples: (1) firms in manufacturing industries vs. firms in other industries; (2) firms headquartered in Top 5 provinces in China vs. other firms; (3) State-owned enterprises vs. other firms; and (4) pre- vs. post-Covid sample period. CDG's predictive power increases post-Covid, consistent with the notion that travel disruptions force many businesses to rely more on cloud services.

Since firm-level cloud data records are not available to the general public during our sample period, they can be viewed as private signals. CDG's predictive power therefore reflects the value of such private signals. In the cross section, we expect the value of CDG to be smaller among large firms, firms with higher institutional ownership and analyst coverage. This is because larger firms generally enjoy a more transparent information environment and the information production efforts by institutions and analysts also diminish the incremental value of CDG. Consistent with this notion, we find that the fundamental and return predictive power of CDG is indeed lower among large firms, firms with higher institutional ownership and analyst coverage.

We also explore different types of cloud computing service. Software-as-a-service (SaaS) involves the licensing of a software application to customers. Licenses are typically provided through a pay-as-you-go model or on-demand. This type of system can be found in Microsoft Office's 365. Platform-as-a-service (PaaS) shares some similarities with SaaS, but instead of delivering software online, it is actually a platform for creating software that is delivered via the Internet. This model includes platforms like Salesforce.com and Heroku. Finally, Infrastructure-as-a-service (IaaS) involves a method for delivering everything from operating systems to servers and storage through IP-based connectivity as part of an on-demand service. Clients can avoid the need to purchase

Electronic copy available at: https://ssrn.com/abstract=3893300

software or servers, and instead procure these resources in an outsourced, on-demand service. Popular examples of the IaaS system include IBM Cloud and Microsoft Azure. Compared to SaaS, IaaS is more integrated into the business of a firm. Thus cloud data size under the IaaS category should paint a more complete picture of a firm's fundamentals. We confirm this by computing three CDG measures, each corresponding to one type of cloud services. Indeed, the fundamental and return predictive power is strongest for CDG under IaaS, followed by CDG under PaaS, and then CDG under SaaS.

Finally, we find cloud data to facilitate insider trading. CDG predicts both the intensity and profitability of insider trades among our sample firms. More importantly, we compare insider trading outcomes of a firm before and after it uses cloud computing, benchmarked against a control group of peer firms in the same industry with similar characteristics that do not use cloud computing. Compared to its peers, the amount of (both buys and sells) and the return to insider trading of a firm increase only after its cloud computing adoption. The results are again stronger among firms using IaaS system. The Diff-in-Diff analyses provide causal evidence that cloud computing, by making available a powerful private signal of firms' fundamentals, has the unintended consequence of facilitating insider trading,

The rest of the paper is organized as follows. Section 2 describes the data and variables. Section 3 presents the main empirical results on firm fundamentals, earnings surprises, and innovation performance. Section 4 tests whether the CDG is a significant predictor to cross-sectional stock returns. Section 5 performs additional analyses in different subsamples, other countries, and using different CDG measures. Section 6 studies the impact of cloud data on insider trading. Section 7 concludes.

2 Data, Variables, and Summary Statistics

Cloud computing is a term that has gained widespread use over the last few years. With the exponential increase in data, it becomes more and more difficult for individuals and organizations to keep all of their vital information, programs, and systems up and running on in-house computer servers. The solution to this problem is one that has been around for nearly as long as the internet, but that has only recently gained widespread application for businesses. Cloud computing operates on a similar principle as web-based email clients, allowing users to access all of the features and files of the system without having to keep the bulk of that system on their own computers. In fact,

Electronic copy available at: https://ssrn.com/abstract=3893300

most people already use a variety of cloud computing services without even realizing it. Gmail, Google Drive, TurboTax, and even Facebook and Instagram are all cloud-based applications. For all of these services, users are sending their personal data to a cloud-hosted server that stores the information for later access. And as useful as these applications are for personal use, they are even more valuable for businesses that need to be able to access large amounts of data over a secure, online network connection. For example, employees can access customer information via cloud-based CRM software from their smartphone or tablet at home or while traveling, and can quickly share that information with other authorized parties anywhere in the world.

We obtain the proprietary cloud data from the leading cloud computing platform in China. Similar to the business model of Amazon Web Services (AWS), this platform offers three types cloud computing services (PaaS, IaaS, and SaaS) to enterprises in China. By 2020, the company's cloud computing services cover more than 3 million customers in the world, nearly 40% global fortune 500 firms, more than 50% listed firms in China, 80% technology firms in China, and 99% cities in China.

Our study only focuses on publicly listed firms. In China, each registered business entity has a Unified Social Credit Code (USCC) issued by the Chinese government. To identify listed firms, we extract the USCC information about the cloud computing platform's firm clients and match our platform data with the China Stock Market and Accounting Research Database (CSMAR). CSMAR provides comprehensive information about stock prices, financial statements, corporate governance, and ownership structure for all publicly listed firms in Shanghai and Shenzhen stock exchanges.

We apply several filters in constructing our main sample. First, we require firms to have at least 100TB cloud data during a quarter. Second, we exclude stocks having less than 15 days of trading records during the most recent month. Third, we exclude financial, real estates, and utility firms based on the CSRC industry classification to mitigate the influence from their different regulation and financial reporting standards. Fourth, we remove firm-quarter observations with missing financial information. While we have the firm-level cloud data since 2013, the need to compute year-on-year quarterly growth rates require us to begin our analyses in 2014. Our final sample includes 30,309 firm-quarter observations, which cover 2,298 unique firms. The sample period includes 29 quarters in total, beginning in 2014/Q2 and ending in the 2021/Q2.

On average, our sample covers around 1,045 firms per quarter (more than 73% of the market by market capitalization).This sample coverage is much larger than those in the prior studies exploiting

Electronic copy available at: https://ssrn.com/abstract=3893300

alternative data in the U.S. market. For example, using the customers' review data, Hirshleifer et al. [2018] covers 150 firms each month on average. Using employers' review data, Goldstein and Yang [2019] covers 508 firms each quarter on average. Figure 1 Panel A shows that our sample coverage increases over time, from 321 firms in 2014/Q2 to 2189 firms in 2021/Q2. In addition, Panel B shows that for an average firm in our sample, the cloud data size also grows over time, from 488 terabyte in 2014/Q2 to 2387 terabyte in 2021/Q2. The chart reveals cloud data's potential in tracking economic output in real time. For example, the dips during the first two quarters of 2020 clearly demonstrate the impact of Covid-related lockdown in China. Panel C plots the year-on-year quarterly cloud data growth, aggregated across all firms in our sample, against the corresponding GDP growth rates in China (released with a delay). The correlation between the two series is more than 65%.

Figure 2 presents our sample coverage by province and industry. Panel A shows average number of firms by province. Our sample covers firms headquartered in 29 provinces (out of a total of 31). The top five provinces with the most number of firms are Zhejiang (191 firms), Guangdong (123 firms), Jiangsu (105 firms), Shanghai (93 firms), and Beijing (90 firms). Panel B shows average number of firms by industry. Black bars represent manufacturing industries and grey bars represent non-manufacturing industries. The top 3 largest industries in manufacturing industries are machinery, equipment, and instrument industry (145 firms), petroleum, chemical, plastic, and rubber industry (133 firms), and metal and non-metal industry (115 firms). The top 3 largest industries in non-manufacturing industries are information technology industry (261 firms), social service industry (45 firms), and wholesale and retail trade industry (39 firms).

We construct our main variable of interest as the year-on-year cloud data growth rate for the firm in a quarter (CDG).[4] Specifically, $CDG_{i,q}$ is defined as the natural logarithm of the amount of cloud data of firm i in quarter q (# of $CD_{i,q}$) minus the natural logarithm of the amount of cloud data of the firm in the same quarter last year q-4 (# of $CD_{i,q-4}$),

$$CDG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } CD_{i,q}}{\#\text{ of } CD_{i,q-4}}\right)$$

A larger value of CDG means more cloud data growth and also indicate more cloud computing services used by the firm.

[4] In our main analysis, cloud data are aggregated at the quarterly level to reduce noise and better match the quarterly financial reports. If the cloud data are aggregated at the monthly level and we compute year-on-year growth rate in monthly cloud data, our main results still hold.

Electronic copy available at: https://ssrn.com/abstract=3893300

In the literature, many nowcasters have been proposed to forecast firm fundamentals and earnings surprise. Da et al. [2011] find that google search volume for firms' products can predict revenue surprises, earnings surprises, and earnings announcement returns. Rajgopal et al. [2003] find that website traffic has substantial explanatory power for stock prices and can forecast earnings and book value of equity. Hirshleifer et al. [2018] find abnormal customer ratings positively predict revenues and earnings surprises. The consumer opinions contain information about firms' fundamentals and stock pricing. Green et al. [2019] find firms experiencing improvements in crowdsourced employer ratings significantly outperform firms with declines. Employer rating changes are associated with growth in sales and profitability and help forecast one-quarter-ahead earnings announcement surprises. Katona et al. [2018] and Zhu [2019] use satellite images to count the number of cars in parking lots to construct abnormal changes in parking lot fill rates that can positively forecast revenue, earnings, and earnings announcement returns. Zhu [2019] and Agarwal et al. [2021] find credit card spending can forecast earnings surprise, sales surprise, and earnings announcement returns.

Above studies use US sample data. In our sample, we construct similar variables using Chinese data. First, we construct the year-over-year quarterly growth of search volume for firms' products ($SEAG_{i,q}$). Specifically, $SEAG_{i,q}$ is defined as the natural logarithm of the search volume of products of firm i in quarter q (# of $SEA_{i,q}$) minus the natural logarithm of the search volume of products of the firm in the same quarter last year q-4 (# of $SEA_{i,q-4}$),

$$SEAG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } SEA_{i,q}}{\#\text{ of } SEA_{i,q-4}}\right)$$

A larger value of SEAG means more growth of search volume for firms' products and also indicate high attention to firms' products. Our firm's product search data is from Baidu index. [5]

Second, we construct the year-over-year quarterly growth of firms' App visiting volume ($APPG_{i,q}$. Specifically, $APPG_{i,q}$ is defined as the natural logarithm of the visiting volume of App of firm i in quarter q ($\#ofAPP_{i,q}$) minus the natural logarithm of the visiting volume of App of the firm in the same quarter last year q-4 ($\#ofAPP_{i,q-4}$),

$$APPG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } APP_{i,q}}{\#\text{ of } APP_{i,q-4}}\right)$$

A larger value of APPG means more growth of visiting volume for firms' App and also indicate

[5] https://index.baidu.com/

Electronic copy available at: https://ssrn.com/abstract=3893300

high attention to firms' information. The App visiting volume is from Qianfan. [6]

Third, we construct the year-over-year quarterly growth of customer product ratings of firms ($CUSG_{i,q}$). Specifically, $CUSG_{i,q}$ is defined as the natural logarithm of the customer product ratings of firm i in quarter q ($\#ofCUS_{i,q}$) minus the natural logarithm of the customer product ratings of the firm in the same quarter last year q-4 ($\#ofCUS_{i,q}$),

$$CUSG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } CUS_{i,q}}{\#\text{ of } CUS_{i,q-4}}\right)$$

A larger value of CUSG means more growth of customer product ratings of firms and also indicate high customers' satisfaction. The customer product ratings are from ECdataway.[7]

Fourth, we construct the year-over-year quarterly growth of employer ratings of firms ($EMPG_{i,q}$). Specifically, $EMPG_{i,q}$ is defined as the natural logarithm of the employer ratings of firm i in quarter q ($\#ofEMP_{i,q}$) minus the natural logarithm of the employer ratings of the firm in the same quarter last year q-4 ($\#ofEMP_{i,q-4}$),

$$EMPG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } EMP_{i,q}}{\#\text{ of } EMP_{i,q-4}}\right)$$

A larger value of EMPG means more growth of employer ratings of firms and also indicate high employees' satisfaction. The employer ratings are from Kanzhun. [8]

Fifth, we construct the year-over-year quarterly growth of number of cars in parking lots of firms ($CAGG_{i,q}$). Specifically, $CAGG_{i,q}$ is defined as the natural logarithm of number of cars in parking lots of firm i in quarter q ($\#ofCAR_{i,q}$) minus the natural logarithm of number of cars in parking lots of the firm in the same quarter last year q-4 ($\#ofCAR_{i,q-4}$),

$$CARG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } CAR_{i,q}}{\#\text{ of } CAR_{i,q-4}}\right)$$

A larger value of CARG means more growth of number of cars in parking lots of firms and also indicate high working time. The number of cars in parking lots of firms is from Wywxdata.[9]

Sixth, we construct the year-over-year quarterly growth of credit card spending to firms' products and services ($SPEG_{i,q}$). Specifically, $SPEG_{i,q}$ is defined as the natural logarithm of credit

[6] https://qianfan.analysys.cn/
[7] https://www.ecdataway.com/
[8] https://www.kanzhun.com/
[9] https://www.wywxdata.cn/

Electronic copy available at: https://ssrn.com/abstract=3893300

card spending to the products and services of firm i in quarter q ($\#ofSPE_{i,q}$) minus the natural logarithm of credit card spending to the products and services of the firm in the same quarter last year q-4 ($\#ofSPE_{i,q-4}$),

$$SPEG_{i,q} = \text{Ln}\left(\frac{\#\text{ of } SPE_{i,q}}{\#\text{ of } SPE_{i,q-4}}\right)$$

A larger value of SPEG means more growth of credit card spending to the products and services of firms and also indicate high popularity of products and services of firms. The credit card spending data is from one of largest commercial banks in China.

In our main analysis, we include the following control variables. Specifically, SIZE is the firm's market capitalization computed as the logarithm of the market value of the firm's outstanding equity at the end of quarter q-1. BM is the logarithm of the firm's book value of equity divided by its market capitalization, where the BM ratio is computed following Fama and French [2008]. Firms with negative book values are excluded from the analysis. ROA is the quarterly operating income scaled by lagged assets. LEV is the quarterly sum of long-term debt and short-term borrowing scaled by total assets. Short-term reversal (STR) is the stock's lagged-one monthly return. MOM is the stock's cumulative return from the start of lagged-twelve month to the end of lagged-two month (skipping the STR month), following Jegadeesh and Titman [1993]. PPE Growth (PG) is the year-over-year quarterly growth in property, plant, and equipment scaled by total assets. Intangible Growth (IG) is the year-over-year quarterly growth in intangible assets scaled by total assets. TO is the quarterly turnover computed as the number of shares traded divided by the total number of shares outstanding in quarter q-1. ILLIQ is the quarterly illiquidity measure computed as the absolute daily return divided by daily dollar trading volume, averaged in quarter q-1. IVOL is the idiosyncratic volatility defined as the standard deviation of daily residuals estimated from the regression of daily excess stock returns on the daily market, size, and value factors of Fama and French [1993] in quarter q-1. SUE is the standardized unexpected earnings defined as actual earnings in the current quarter minus earnings 4 quarters ago, scaled by stock price in the current quarter following Livnat and Mendenhall [2006]. ANA is defined as the number of analysts following the firm in quarter q-1, and IO is the percentage of tradable shares held by institutional investors in quarter q-1. We winsorize control variables at the 1st and 99th cross-sectional percentiles.

To measure the innovation performance, we use two measures that represent innovation activities, i.e., the log of one plus number of patents applied (PA) and the log of one plus number of patents granted (PG). PA is the log of one plus quarterly number of patents applied of the firm.

Electronic copy available at: https://ssrn.com/abstract=3893300

PG is the log of one plus quarterly number of patents granted of the firm. The firm's patent data are from Datayes.[10]

Table 1 presents descriptive statistics for the main variables. Panel A reports the firm characteristics. The cloud data (CD), on average, has 1240.152 terabytes. Other statistics in Panel A suggests that firms in our sample, on average, have quarterly return on assets of 1.412%, market capitalization of RMB 5.23 billion RMB, the book-to-market ratio of 0.463, book leverage of 0.182, percentage ownership by institutional investors of 6.482%, and 7.612 analysts. Comparing with all A shares, our sample firms have larger ROA, larger market capitalization, lower book-to-market ratio, larger book leverage, larger institutional ownership and analyst coverage. Panel B reports the characteristics of firm fundamentals, earnings surprise, and innovation performance. The average growth of total assets (AG), growth of sales (SG), return on assets (ROA), standardized unexpected earnings (SUE), earnings announcement abnormal returns (CAR), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). are 0.182, 0.183, 1.412, 0.129, 0.309, 1.194, 1.030, respectively. Comparing with all A shares, our sample firms have larger fundamental growth, earnings surprise, and better innovation performance. Panel C reports our main variable of interest in this paper, the CDG. We note that the mean (median) value of this measure is 0.182 (0.149). The variation of this measure is also large, with the 5th and 95th percentiles being -29.6% and 73.4%, respectively. Panel C also report other nowcasters. The mean of these nowcasters ranges from 0.063 to 0.132. Their sample coverage of the A-shares market tends to be much smaller compared to that of CDG.

3 Nowcasting and forecasting firm fundamentals and earnings surprises

3.1 Nowcasting and forecasting firm fundamentals

In this section, we examine whether the CDG actually contains valuable information about firm fundamentals. If the CDG reflect the company's fundamentals, we should expect companies with higher cloud data growth to perform well both contemporaneously and in the future. Thus, we conduct quarterly panel data regressions of the measures of fundamentals on the CDG as well as the control variables used in Panel A of Table 1. Specifically, we run the following panel data regressions:

[10] https://www.datayes.com/

Electronic copy available at: https://ssrn.com/abstract=3893300

$$FF_{i,q+n} = \alpha_d + \beta_1 * CDG_{i,q+1} + \beta_2 * FF_{i,q} + \gamma_{i,q+n} + \text{controls}_{i,q} + e_{i,q+n} \tag{1}$$

where $FF_{i,q+n}$ is the firm i's fundamentals in quarter q + n (n=1 or 2), α_d is industry fixed effect, $CDG_{i,q}$ is the firm i's quarterly cloud data growth in quarter q+1, $\gamma_{i,q+n}$ is year-quarter fixed effect. We include the past firm fundamentals in the model to account for persistence in firm fundamentals. We also include control variables used in Panel A of Table 1 in the regressions. To reduce the influence of outliers, we winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a mean of zero and standard deviation of one. Standard errors are double clustered by industry and by year-quarter.

To measure the firm fundamentals, we use three proxies that are prevalent in the literature, namely return-on-asset (ROA), assets growth (AG), and sales growth (SG). ROA is quarterly operating income scaled by lagged assets. AG is quarterly growth in total assets. SG is quarterly growth in sales. These three measures all reflect the real operating performance of a company (Hirshleifer et al. [2013]; Hirshleifer et al. [2018]).

Table 2 presents the average slope coefficients and the corresponding t-statistics from the quarterly panel data regressions. The results show a significantly positive relationship between the CDG and the proxies of firm fundamentals in quarter q+1 or quarter q+2. Specifically, we regress ROA, assets growth, or sales growth, in quarter q+1 or quarter q+2 on the CDG in quarter q+1 as well as the assets growth, sales growth, or ROA in quarter q. For quarter q+1, the coefficients between the CDG and firm fundamentals are significant at the 1% level after accounting for the control variables and the industry and year-quarter fixed effects. The coefficient between the CDG in quarter q+1 and ROA (assets growth, sales growth, PA, PG) in quarter q+1 is 0.765 (0.302, 0.071, 0.308, 0.164). For quarter q+2, the coefficient between the CDG in quarter q+1 and ROA (assets growth, sales growth, PA, PG) in quarter q+2 decreases to 0.522 (0.231, 0.047, 0.192, 0.126), at 1% significance level.

Last four columns of Table 2 show the results of CDG and future innovation performance. The CDG can nowcast and forecast PA and PG of the firm. The coefficient of CDG decreases from 0.308 (0.164) to 0.192 (0.126) from nowcasting to forecasting, in the case of PA (PG).

Table 3 compares CDG to other nowcasters. In each column, we add all six nowcasters as additional controls in the regression. We find these nowcasters cannot significantly change our CDG ability to nowcast and forecast firm fundamentals. To nowcast and forecast firm's ROA,

Electronic copy available at: https://ssrn.com/abstract=3893300

the coefficients of CDG are 0.470 and 0.371 and t-statistics are 3.60 and 2.94. To nowcast and forecast firm's total asset growth, the coefficients of CDG are 0.187 and 0.148 and t-statistics are 3.47 and 2.86. To nowcast and forecast firm's sales growth, the coefficients of CDG are 0.043 and 0.035 and t-statistics are 2.90 and 2.41. Compared to CDG, the forecasting power of the other nowcasters is more sporadic. For example, SEAG and CARG only predict ROA, asset growth and sales growth, while SPEG only predict patent outcomes. The requirement to having all the seven nowcasters available for the firm significantly reduces the sample size in Table 3. We also horse race CDG against the alternative nowcaster, one at a time, in larger samples, and reach very similar conclusion. CDG's predictive power is never subsumed by the other nowcasters.

Overall, the results indicate that CDG indeed contains valuable information about the firm fundamentals.

3.2 Nowcasting and forecasting earnings surprises

While information regarding the cloud data is not available to the public in real time, part of it may be released via future earnings announcements. In this subsection, we examine whether the CDG can nowcast and forecast future earnings surprises. We use standardized unexpected earnings (SUE), defined as actual earnings in the current quarter minus earnings 4 quarters ago, scaled by stock price in the current quarter, following Livnat and Mendenhall [2006], to proxy for earnings surprise. We conduct panel data regressions of the quarterly SUE (for fiscal quarters q+1 and q+2 which are annouced in quarters q+2 and q+3, respectively) on the CDG in quarter q+1 and control variables of Panel A of Table 1 in quarter q. We also examine whether CDG can nowcast and forecast earnings announcement abnormal returns (CAR). CAR is the cumulative abnormal returns over the three-day window surrounding the earnings announcement. Abnormal return is calculated as the raw daily return minus the daily return on size and market-to-book matched portfolio as in Livnat and Mendenhall [2006]. We conduct panel data regressions of the quarterly CAR (corresponding to announcements of quarter q+1 and q+2 earnings) on CDG in quarter q+1 and control variables of Panel A of Table 1 in quarter q. For panel data regressions, we also control for the industry and year-quarter fixed effects. We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a mean of zero and standard deviation of one to reduce the effect of outliers. Standard errors are double clustered by industry and by year-quarter. If CDG contains nowcasting and forecasting information about SUE or CAR, we should expect the slope coefficient to be positive and significant.

Electronic copy available at: https://ssrn.com/abstract=3893300

Consistent with our expectation, for quarter q+1 SUE, Table 4 shows that the coefficient on the CDG is 0.214 with a t-statistic of 3.55 accounting for past SUE, control variables, and the industry and year-quarter fixed effects. For quarter q+2 SUE, the coefficient on the CDG is 0.161 with a t-statistic of 2.73 after controls. Moreover, consistent with Bernard and Thomas (1989), the lagged SUE at quarter q is strongly positively correlated with the future SUE. In Column 3 and 4, we find that CDG can forecast CAR in the next two quarters. The coefficients on the CDG are 2.130 (t-statistic = 3.18) and 1.364 (t-statistic = 2.28), respectively.

We also examine CDG's SUE nowcasting and forecasting after controlling other nowcasters in Table 5. In each column, we add all six nowcasters as additional controls in the regression. We find that these nowcasters cannot significantly change our CDG nowcasting and forecasting power to earnings surprises. To nowcast and forecast firm's SUE, the coefficients of CDG are 0.133 and 0.104 and t-statistics are 2.44 and 1.90. To forecast firms' next two quarter CARs, the coefficients of CDG are 1.418 and 1.120 and t-statistics are 2.28 and 1.77.

Overall, the results confirm CDG's predictive power of earnings surprises and market reaction during earnings announcements in the next two quarters.

4 Forecasting stock returns

In this section, we test whether the CDG predicts the cross-section of future stock returns using portfolio-sort and firm-level cross-sectional regression analyses.

4.1 Univariate portfolio sorts

To construct the long-short portfolio, at the end of each quarter from 2013/Q2 to 2021/Q2, individual stocks are sorted into quintile portfolios based on their CDGs in that quarter and are held for the next quarter. We then compute the value-weighted and equal-weighted average monthly excess return of each quintile portfolio. To examine the cross-sectional relation between the CDG and the future stock returns, we form a long-short portfolio that takes a long position in the highest quintile of CDG and a short position in the lowest quintile of CDG.

In Table 6, we report the average monthly excess returns of each quintile portfolio and the long-short portfolio (in excess of the one-month deposit interest rate). We also report the abnormal returns (alphas) estimated with various factor models, including the China q-factor model based on Hou et al. [2015], China five-factor model based on Fama and French [2015], the LSY3 factor

Electronic copy available at: https://ssrn.com/abstract=3893300

model of Liu et al. [2019], and the LSY4 factor model of Liu et al. [2019]. Controlling for these factors helps to ensure that the CDG indeed contains incremental predictive power beyond these well-known factor models.We also report average excess returns in earnings announcement months and average excess returns in non-earnings announcement months.

In general, the excess returns and alphas of five quintile portfolios increase monotonically from quintile 1 to quintile 5. The long-short portfolio that buys 20% of the stocks with the highest CDG (quintile 5) and short-sells 20% of the stocks with the lowest CDG (quintile 1) earns a value-weighted (equal-weighted) average return of 0.851% (1.202%) per month with a t-statistic of 4.17 (5.11), translating into an annualized return of 10.212% (14.424%). [11] Controlling for the factors does not change the magnitude and statistical significance of the return spreads on the CDG-sorted portfolios for most of the factor models. The alpha is from 0.751% (HXZ) to 0.623% (LSY4) per month and the corresponding t-statistic is from 3.99 to 2.91 for the value-weighted portfolio. Finally, the significant relation between CDG and future returns is largely coming from the short leg of the arbitrage portfolio as the economic magnitude and statistical significance are larger among the stocks in the short leg than those in the long leg. This implies that high CDG firms are overvalued relative to firms with lower CDG, perhaps due to the short selling limitation in China.In earnings announcement months, the value-weighted (equal-weighted) long-short excess returns are 0.559% (0.898%). In non-earnings announcement months, the value-weighted (equal-weighted) long-short excess returns are 0.291% (0.304%). The excess returns in earnings announcement months are about 2-3 times larger than the excess returns in non-earnings announcement months.

We investigate the long-term predictive power of CDG by calculating the LSY4-factor alphas of the CDG long-short portfolio from first to twenty-fourth month after portfolio formation. The results are presented in Figure 3. The predictive power of CDG on future returns decreases after first month. The alpha drops from 62.3 basis points in the first month to 38.3 and 32.4 basis points in the second and third month, respectively. The alpha becomes even smaller beyond the first quarter but never switches to be negative. The lack of long-term return reversal suggests that CDG's return predictability is unlikely driven by a persistent price pressure which eventually should be reverted.

[11] The t-statistics reported in our portfolio and regression analyses are Newey and West [1987] adjusted with three lags to control for heteroskedasticity and autocorrelation.

Electronic copy available at: https://ssrn.com/abstract=3893300

4.2 Fama-MacBeth cross-sectional regressions

In this section, we conduct firm-level Fama-MacBeth cross-sectional regressions to test if CDG predicts the cross-section of monthly returns in the next quarter. The test allows us to examine the incremental predictive power of CDG by controlling for other known return predictors. Each month, we run a cross-sectional regression of stock returns in that month on the last quarter CDG as well as a number of control variables, including lagged size, book-to-market, ROA, leverage, PPE growth, intangible growth, earnings surprise, short-term return reversal, price momentum, idiosyncratic volatility, illiquidity, turnover ratio, analyst coverage, and institutional ownership. To minimize the effect of outliers, all independent variables are winsorized at the 1st and 99th percentiles. We also control for the industry and geography fixed effects following the CSRC industry classification and China province classification. The stock-level cross-sectional regressions are run each month and the standard errors of the average slope coefficients are corrected for heteroskedasticity and autocorrelation following Newey and West [1987].

Panel A of Table 7 reports the Fama-MacBeth cross-sectional regressions' results. In column 1, we include only CDG in the cross-sectional regressions. We control the industry and geography fixed effects using CSRC industry classification and China province classification. Consistent with the portfolio results, we find a positive and significant relation between the CDG and one-month-ahead returns. The average slope coefficient on the CDG ratio is 0.505 with a t-statistic of 4.08. In column 2, we further control other well-known return predictors in the cross-sectional regressions. We find a positive and significant relation between the CDG and one-month-ahead returns controlling for a large number of predictors. The CDG retains significant predictive power, and the magnitude of the average slope coefficient decreases only slightly to 0.469, suggesting that the information embedded in CDG is almost orthogonal to that in other known return predictors. The slope coefficients on the control variables are consistent with prior literature: market capitalization (SIZE), short term reversal (STR), and idiosyncratic volatility (IVOL) are negatively correlated with the future return, and ROA, earnings surprise (SUE), and institutional ownership (IO) are positively related to the next month's return.

In column 3, we include INDRET, which is computed as the value-weighted CSRC industry portfolio returns, as a control variable in our main regression to further control for the industry effect. Specifically, we adjust the dependent variable, by subtracting the firm's value-weighted CSRC industry return INDRET from the firm's current month return. Doing so allows us to tease

Electronic copy available at: https://ssrn.com/abstract=3893300

out the return predictive power from the CDG rather than the one-month industry momentum effect. The coefficient of the CDG remains similar controlling for the industry return directly. In column 4, we further control for the geographic momentum that are shown to affect stock returns systematically. Specifically, we use RET-GEORET, which is the difference between the firm's return and the corresponding province portfolio returns. We replace the firm's raw return with this geographic-adjusted return as the dependent variable and run the same monthly cross-sectional regressions. Again, the magnitude of the slope coefficient on CDG becomes slightly weaker, but remains highly significant.

Panel B of Table 7 reports Fama-MacBeth cross-sectional regressions of CDG and other nowcasters. The nowcasters cannot significantly change the predictive power of CDG. After adding all six nowcasters, the predictive coefficient of CDG becomes 0.291 and corresponding t-statistic is 2.72. The sample size is much smaller due to the requirement that both CDG and other nowcasters have to be non-missing.

Overall, these results indicate that the CDG provides incrementally value-relevant information. The predictive power of the CDG is distinct and robust to the inclusion of other well-known return predictors, nowcasters, and regression specifications.

5 Additional Analyses

5.1 International evidence

We further examine CDG nowcasting and forecasting power to firm fundamentals, earnings surprise, and long-short excess returns and alphas in other countries. Table 8 reports the international evidence. Panel A examines CDG nowcasting and forecasting power to firm fundamentals in other countries. We obtain CDG of firms in Indonesia, Japan, Malaysia, and Singapore, respectively. Panel A of Table 8 shows results of predicting firms' ROA, total asset growth, and sales growth. The coefficients between the CDG in quarter q+1 and firm fundamentals (ROA, asset growth, sales growth) in quarter q+1 are significant after accounting for the control variables and the industry and year-quarter fixed effects in four countries. As for firms in Indonesia, Japan, and Singapore, the coefficients between the CDG in quarter q+1 and firm fundamentals (ROA, asset growth, and sales growth) in quarter q+2 is significant. But as for firms in Malaysia, the coefficients between the CDG in quarter q+1 and firm fundamentals (ROA, asset growth, and sales growth) in quarter

Electronic copy available at: https://ssrn.com/abstract=3893300

q+2 become insignificant.[12]

Panel B examine CDG nowcasting and forecasting power to earnings surprises in other countries, including Indonesia, Japan, Malaysia, and Singapore. Panel B shows results of nowcasting and forecasting firms' SUE and CAR. The coefficients between the CDG in quarter q+1 and earnings surprise (SUE and CAR) in quarter q+1 are significant after accounting for the control variables and the industry and year-quarter fixed effects in four countries. As for firms in Indonesia, Japan, and Singapore, the coefficients between the CDG in quarter q+1 and earnings surprise (SUE and CAR) in quarter q+2 is significant. But as for firms in Malaysia, the coefficients between the CDG in quarter q+1 and earnings surprise (SUE and CAR) in quarter q+2 become insignificant.

Panel C reports long-short value-weighted excess returns and alphas of CDG in four countries. The long-short value-weighted excess returns of Indonesia, Japan, Malaysia, and Singapore are 0.398%, 0.575%, 0.328%, and 0.439%, respectively. The Fama and French (2018) six-factor alphas of Indonesia, Japan, and Singapore are significant, but Fama and French (2018) six-factor alpha of Malaysia is insignificant.[13]

Overall, we find strong evidence for CDG's nowcasting and forecasting power in international markets. The only exception is regarding forecasting fundamentals in Malaysia, where CDG's predictive power becomes marginally insignificant.

5.2 Robustness in subsamples

In this section, we conduct different robustness tests. First, we study CDG's nowcasting and forecasting power within the manufacturing industries' sample firms vs. the non-manufacturing industries' sample firms. The manufacturing industries account for 61% of all sample firms and the non-manufacturing industries account for the remaining 39%. Second, we study CDG's nowcasting and forecasting within 5 largest provinces' sample firms vs. other provinces' sample firms. The 5 largest provinces are Zhejiang, Guangdong, Jiangsu, Shanghai, and Beijing, accounting for 55% of all sample firms in total. Third, we study CDG's nowcasting and forecasting power within state-owned enterprises vs. non-state-owned enterprises. We obtain the enterprise type (state-owned and non-state-owned) from the CSMAR. Non-state-owned firms use more cloud computing services and have more cloud data than state-owned firms, as state-owned enterprises have more financial

[12] The patent data of the four countries are not obtainable.

[13] For Japan and Singapore, we use the Fama and French developed markets six-factor model. For Indonesia and Malaysia, we use the Fama and French emerging markets six-factor model. The factors are available in Ken French data library.

Electronic copy available at: https://ssrn.com/abstract=3893300

strength to build their own private clouds rather than using public cloud platform. Finally, we study CDG's nowcasting and forecasting power during the pre-Covid (from April 2014 to December 2019) and during the post-Covid (from January 2020 to June 2021).

Table 9 shows these subsample results. In Panel A, we find that the CDG in different subsamples can nowcast and forecast firm fundamentals. Panel B shows that CDG can nowcast and forecast earnings surprise and CAR in different subsamples. In both panels, the coefficients on CDG are larger among firms in manufacturing industries, in Top5 provinces, among private enterprises, and during the post-Covid sample period. Panel C shows that CDG can generate economically and statistically significant returns and alphas in different subsamples. Again, the excess returns and alphas are larger among firms in manufacturing industries, in Top5 provinces, among private enterprises, and during the post-Covid sample period.

It is intuitive why CDG's predictive power increases post-Covid. The travel disruptions forced many businesses to rely more on cloud services, and as a result, cloud data size becomes an even better proxy for their outputs. In addition, Covid changed the competitive landscape across industries in China. For example, zooming into the cloud data growth during the first two quarters of 2020, we find that many pharmaceutical companies enjoyed the highest CDGs while companies in consumer goods sectors suffered the lowest CDGs. While the different is not surprising ex-post, cloud data quantify the impact of Covid on a real time basis.

5.3 CDG predictability across different firms

During our sample period, firm-level cloud data records are not available to the public. CDG can therefore be viewed as a private signal. Its predictive power thus reflects the signal value. In the cross section, we expect the value of CDG to be smaller among large firms, firms with higher institutional ownership and analyst coverage. This is because larger firms generally enjoy a more transparent information environment and the information production efforts by institutions and analysts also diminish the incremental value of CDG.

Results in Table 10 are consistent with our conjecture. Indeed, the fundamental and return predictive power of CDG is lower among large firms, firms with higher institutional ownership and analyst coverage.

Electronic copy available at: https://ssrn.com/abstract=3893300

5.4 CDG measures for different service types

In this section, we study whether cloud data in three different types of cloud computing services (IaaS, PaaS, and SaaS) reveal different information about the fundamentals of a firm. As discussed early, IaaS is more integrated into the business of a firm, compared to PaaS and SaaS. Thus cloud data size under the IaaS category should paint a more complete picture of a firm's fundamentals. We confirm this by computing three CDG measures, each corresponding to one type of cloud services. We then repeat the fundamental and return predictive exercises using each measure. The results are reported in Table 11. Indeed, the fundamental and return predictive power is strongest for CDG under IaaS, followed by CDG under PaaS, and then CDG under SaaS.

6 Insider trading

So far, we have established that cloud data contain value-relevant information about a firm. While the data are not available to the general public, they are available to the insiders in the firm and allow them to track the firm's fundamentals in real time. We therefore examine how cloud data impact insider trading.

We define insider trading as trading conducted by the firm's board directors and executive officers. We exclude the changes in shareholdings due to stock dividends or exercising stock options, and only consider the changes in shareholdings due to board directors and executive officers' trades in the secondary market. To gauge insider trading activities, we construct the three different measures. InsiderBuy is the total number of shares purchased by insiders during the quarter, scaled by the number of total tradable shares. InsiderSell is the total number of shares sold by insiders during the quarter, scaled by the number of tradable shares. InsiderNet is the net insider trading, calculated as InsiderBuy minus InsiderSell.

In Table 12, we examine if CDG predicts the direction and profitability of insider trading. Panel A shows that CDG is positively (negatively) related to InsiderBuy (InsiderSell). It is interesting to note the asymmetry between insiders' buying and selling behavior. It seems that insider sales are much more responsive to cloud data growth than insider purchases. This result is consistent with the fact that insiders usually have already allocated a substantial amount of their wealth on the underlying stock, and therefore it is easier for them to make the selling decision than buying even more for diversification and risk control purposes (e.g., see Aboody and Lev [2000]; Huddart et al. [2007]; Marin and Olivier [2008]). Moreover, CDG is also positively related to the net insider

Electronic copy available at: https://ssrn.com/abstract=3893300

trading and can forecast one-month and three-month LSY4 abnormal returns of insider trading.

In Panel B of Table 12, we replace the total insider trading measures with opportunistic insider trading as a robustness check. Following Cohen et al. [2012], we define OppInsiderBuy (OppInsiderSell) as the percentage of shares opportunistically purchased (sold) during the quarter. Specifically, we classify an insider's trades on a stock in a particular month as either opportunistic or routine trades according to whether she/he traded in the same month in the past two years. If the insider has traded consecutively in a particular month over the past two years, the trade in the current month is classified as a routine trade; otherwise, it is classified as an opportunistic trade. We exclude routine trades from this analysis as they are documented as not informative about firms' futures (Cohen et al. [2012]). OppInsiderNet is the net opportunistic insider trading, calculated as OppInsiderBuy minus OppInsiderSell. The results using opportunistic insider trading in Panel B are consistent with those in Panel A.

Of course, the fact that CDG predicts both the intensity and profitability of insider trading does not mean that insiders actually use cloud data to conduct trading. They could have access to other correlated private signals and such signals are available regardless whether the firm uses cloud computing or not. In order to examine the causal impact of cloud data on insider trading, we conduct additional diff-in-diff tests. Specifically, we compare insider trading outcomes of a firm before and after it uses cloud computing, benchmarked against a control group of peer firms in the same industry with similar characteristics that do not use cloud computing.

We examine three years before and three years after a firm's adoption of cloud computing, where event year zero is the year when the firm first uses the clouding computing services. The total insider trading shares are the total number of shares purchased and sold by insiders during the quarter, scaled by the number of shares outstanding. The insider trading buying shares are the total number of shares purchased by insiders during the quarter, scaled by the number of shares outstanding. The insider trading selling shares are the total number of shares sold by insiders during the quarter, scaled by the number of shares outstanding. The dummy variable Treat equals one when the firm uses cloud service, otherwise zero. The control firms do not use cloud services from our sample. For each treatment firm, we match control firms in the same industry using propensity score matching method based on the characteristics of size, value, and turnover. The dummy variable Post equals one when the firm begin to use cloud services, otherwise zero. we test the diff-in-diff tests in the full, IaaS, PaaS, and SaaS samples.

Table 13 reports the diff-in-diff results of insider trading shares. In full sample of Panel A, the

Electronic copy available at: https://ssrn.com/abstract=3893300

coefficient of interaction term Treat * Post is 0.175 at 1% significance level. The coefficient of Treat and the coefficient of Post are not statistically significant. Figure 4 provides a graphic illustration of the result. It is clear that the intensities of insider trading for the treatment and the control groups are similar during the pre-period. They start to diverge only after event year zero. While the insider trading intensity does not change for the control group, it increases significantly for firms using cloud computing. Since insider trading is unlikely to be the reason for a firm to adopt cloud computing, the causality is more like to go from cloud computing to insider trading.

In Table 13, we further break down total insider trading to insider purchases (Panel B) and insider sales (Panel C), and find similar results. One missing variable concern is related to managerial optimism which may drive both the adoption of cloud computing and future shares purchase. The fact that we also observe increased insider sales during the post-period helps to rule out such a concern. The availability of cloud data facilitates opportunistic insider trading. Cloud data allow insiders to better assess if the stock is currently overvalued or undervalued and trade accordingly. Table 13 also conduct the diff-in-diff tests for treatment firms using IaaS, PaaS, and SaaS separately. In general, we find stronger results among the IaaS subsample, consistent with the notion that cloud data better reveal the fundamentals of a firm whose cloud computing is fully integrated with its business.

In Table 14, we examine the insider trading returns. The insider trading returns are one-month or three-month LSY4 abnormal returns of insider trading. In Panel A, we report the one-month abnormal returns. The coefficient of interaction term Treat * Post of the full sample is 0.005 at 1% significance level. The coefficient of Treat and the coefficient of Post are not statistically significant. Figure 5 shows the time-varying firms' insider trading returns before and after using cloud services. We find that treatment firms and control firms have similar insider trading returns before using cloud services (year zero), but treatment firms have dramatically larger insider trading returns than control firms after using cloud services (year zero). The results suggest that cloud data make insider trading more profitable. Again, the effect is bigger among the IaaS subsample.

7 Conclusions

In this paper, we examine a very intuitive nowcaster of a firm's fundamentals. As more and more businesses are moving their data to the cloud, the growth in a firm's cloud data signals the growth in the business in real time. Since information regarding the firm's fundamentals is

Electronic copy available at: https://ssrn.com/abstract=3893300

released with a delay, cloud data growth serves as a powerful "nowcaster." We find that the year-on-year quarterly cloud data growth (CDG) indeed contains value-relevant information for firm fundamentals, earnings surprises, and innovation performance. Specifically, CDG positively predicts assets growth, sales growth, ROA, standardized unexpected earnings (SUE), and patent outcomes. CDG also forecasts stock returns, especially around future earnings announcements. A long-short portfolio by buying (selling) stocks with the high (low) CDGs generates a 9.0% risk-adjusted return annually. In other words, the investment value embedded in CDG is highly significant economically.

We also find CDG to have superior forecasting power than many existing nowcasters based on online product search, App usage, credit card spending, parking lot fill rates, customer and employee ratings. One reason is that cloud computing technology enables the firms to more efficiently run their organizations, better serve their customers, and dramatically increase their overall profit margins. In other words, a higher cloud data growth rate signals not only stronger fundamentals contemporaneously, but also greater earning power going forward. In addition, while many existing nowcasters are available or relevant only for a small subset of firms, CDG has a much wider sample coverage as firms in all sectors are using cloud computing as part of their business.

Finally, we document an unintended consequence of cloud data: it actually facilitates opportunistic insider trading. As cloud computing becomes more efficient and adopted by more firms, the value-relevance of CDG should only increase as well. We leave it to future research to analyze additional benefits and costs associated with cloud computing.

Electronic copy available at: https://ssrn.com/abstract=3893300

References

Aboody, D. and Lev, B. (2000). Information asymmetry, r&d, and insider gains. *The journal of Finance*, 55(6):2747–2766.

Agarwal, S., Qian, W., and Zou, X. (2021). Disaggregated sales and stock returns. *Management Science*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5:31–56.

Cohen, L., Malloy, C., and Pomorski, L. (2012). Decoding inside information. *The Journal of Finance*, 67(3):1009–1043.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5):1461–1499.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63:1653–1678.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636.

Goldstein, I. and Yang, L. (2019). Good disclosure, bad disclosure. *Journal of Financial Economics*, 131(1):118–138.

Green, T. C., Huang, R., Wen, Q., and Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, 134(1):236–251.

Hirshleifer, D., Hsu, P.-H., and Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics*, 107:632–654.

Hirshleifer, D., Hsu, P.-H., and Li, D. (2018). Innovative originality, profitability, and stock returns. *Review of Financial Studies*, 31:2553–2605.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28:650–705.

Huddart, S., Ke, B., and Shi, C. (2007). Jeopardy, non-public information, and insider trading around sec 10-k and 10-q filings. *Journal of Accounting and Economics*, 43(1):3–36.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications

Electronic copy available at: https://ssrn.com/abstract=3893300

for stock market efficiency. *Journal of Finance*, 48:65–91.

Katona, Z., Painter, M., Patatoukas, P. N., and Zeng, J. (2018). On the capital market consequences of alternative data: Evidence from outer space. In *9th Miami Behavioral Finance Conference*.

Liu, J., Stambaugh, R. F., and Yuan, Y. (2019). Size and value in china. *Journal of Financial Economics*, 134(1):48–69.

Livnat, J. and Mendenhall, R. R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, 44:177–205.

Marin, J. M. and Olivier, J. P. (2008). The dog that did not bark: Insider trading and crashes. *The Journal of Finance*, 63(5):2429–2476.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

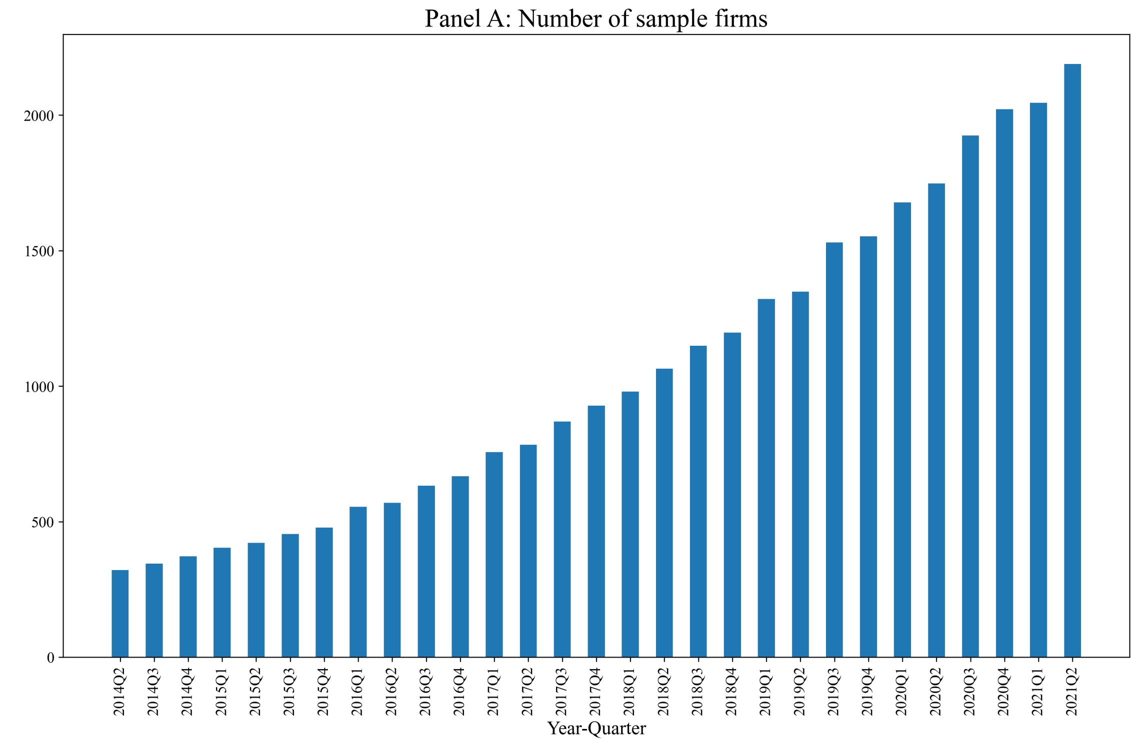
Rajgopal, S., Venkatachalam, M., and Kotha, S. (2003). The value relevance of network advantages: The case of e–commerce firms. *Journal of Accounting Research*, 41(1):135–162.

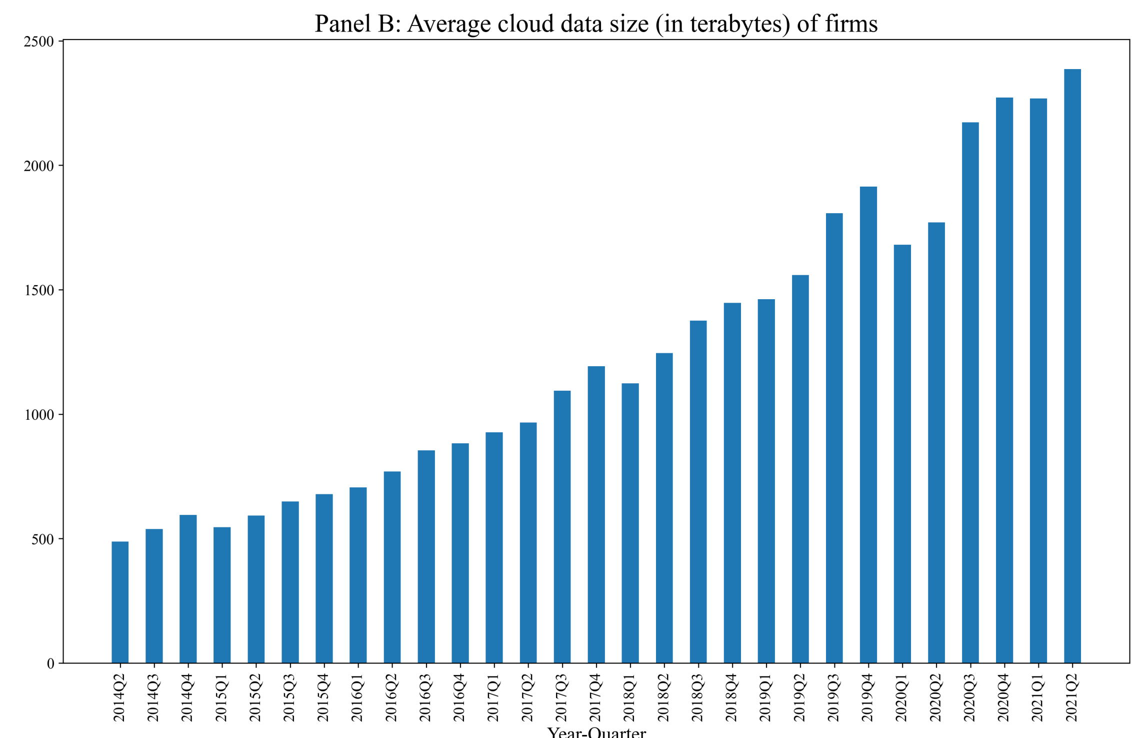
Zhu, C. (2019). Big data as a governance mechanism. *The Review of Financial Studies*, 32(5):2021–2061.

Electronic copy available at: https://ssrn.com/abstract=3893300

Figure 1: Average cloud data and number of sample firms per quarter

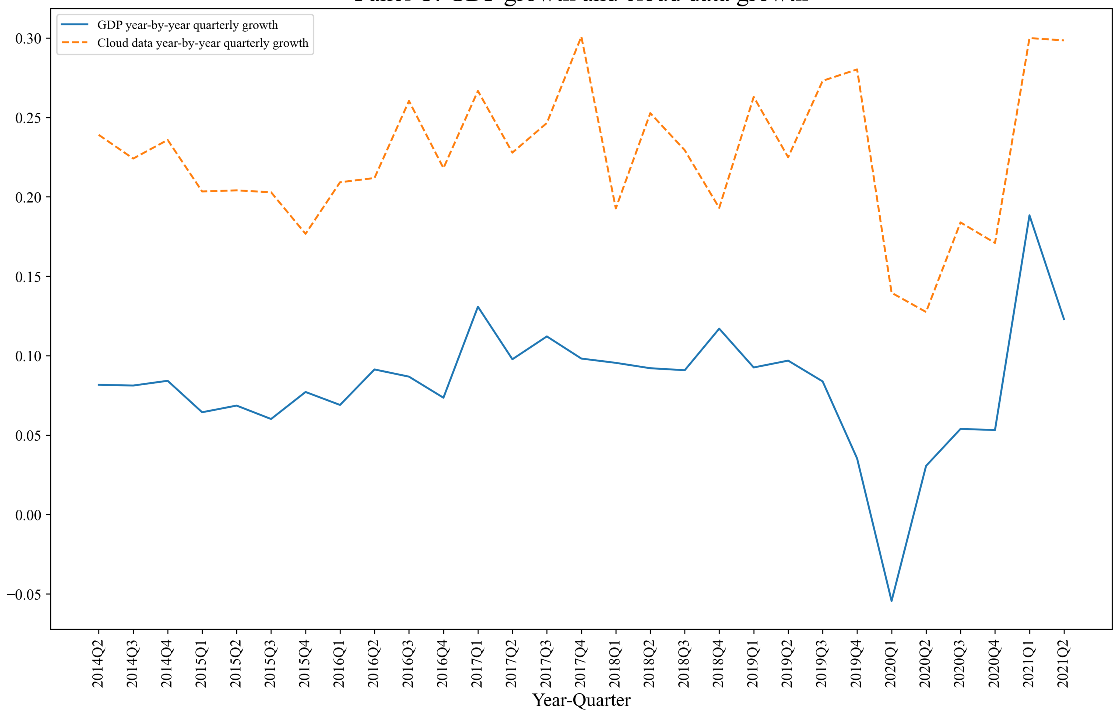
Panel A shows the number of firms in our sample from 2014 to 2021. The vertical axis represents the number of firms in our sample each quarter. The horizontal axis represents each quarter included in our sample. Panel B shows the average cloud data size (in terabytes) of firms in our sample from 2014 to 2021. The vertical axis represents the average cloud data of firms in our sample each quarter. The horizontal axis represents each quarter included in our sample. Panel C shows the GDP year-by-year quarterly growth and the cloud data year-by-year quarterly growth. The correlation between the GDP growth and the cloud data growth is 65.19%.





Electronic copy available at: https://ssrn.com/abstract=3893300

Panel C: GDP growth and cloud data growth

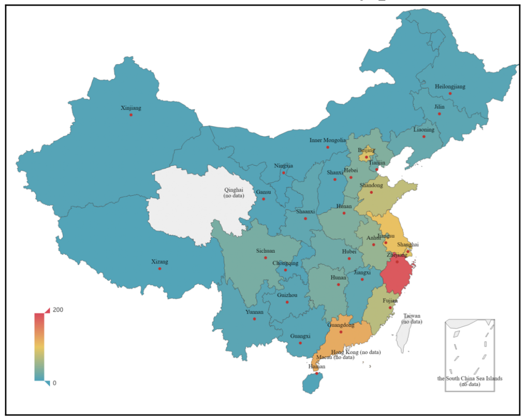


Electronic copy available at: https://ssrn.com/abstract=3893300

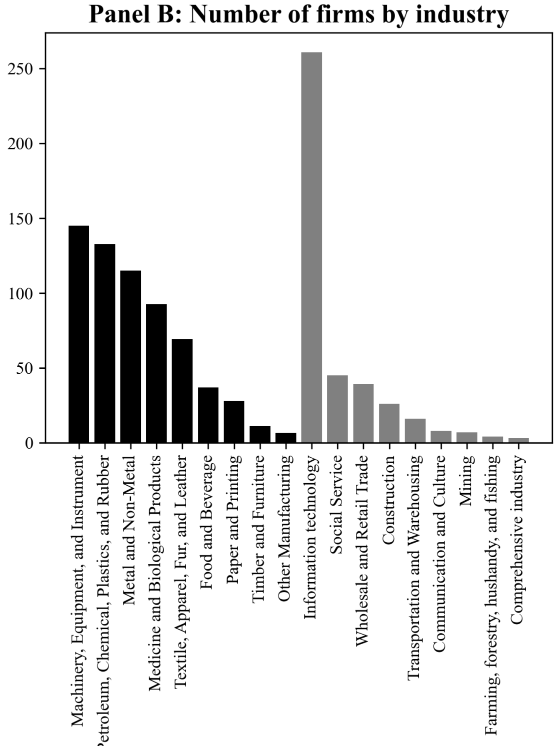
Figure 2: Average number of firms by province and industry

Panel A shows the geographic distribution of the average number of sample firms by province from 2014 to 2021. A province with darker color indicates a higher number of firms in this province. Panel B shows the industry distribution of the average number of sample firms by industry from 2014 to 2021. Black color bars represent manufacturing industries and Gray color bars represent non-manufacturing industries.

Panel A: Number of firms by province



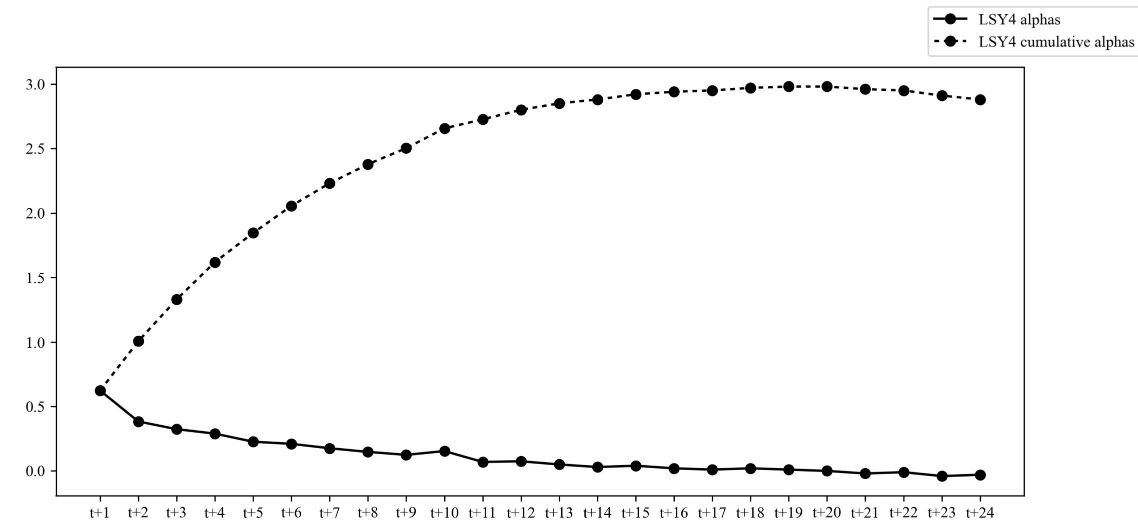
Panel B: Number of firms by industry



Electronic copy available at: https://ssrn.com/abstract=3893300

Figure 3: Long-term alpha

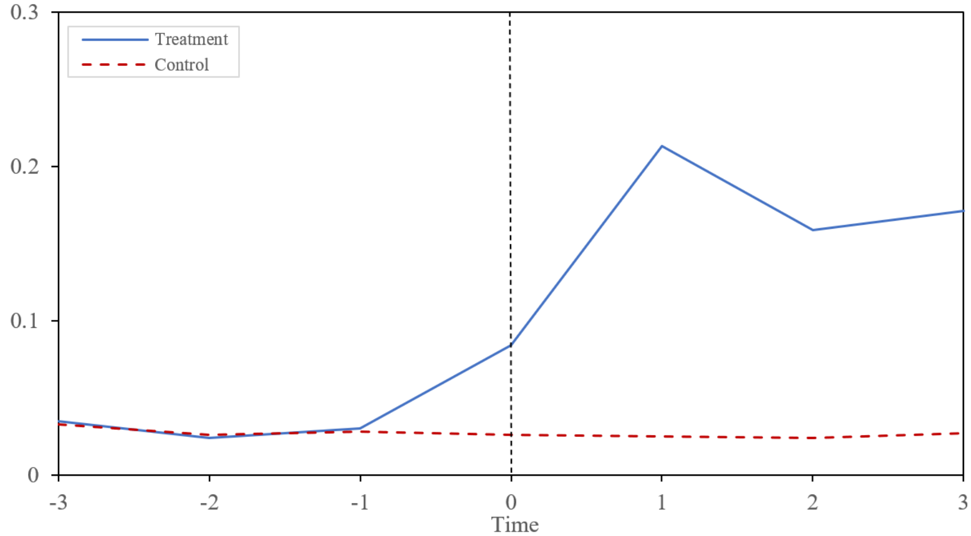
Figure 3 shows Liu et al. [2019] China four-factor alphas in 24 months after portfolio formation. All stocks are value-weighted within each portfolio, and the portfolios are rebalanced every calendar month to maintain value weights. The hedge portfolio is a zero-cost portfolio that buys the top quintile and sells short the bottom quintile. The vertical axis represents the cumulative hedge-portfolio alphas. The horizontal axis represents each month.



Electronic copy available at: https://ssrn.com/abstract=3893300

Figure 4: Insider Trading Shares before and after Using Cloud Services

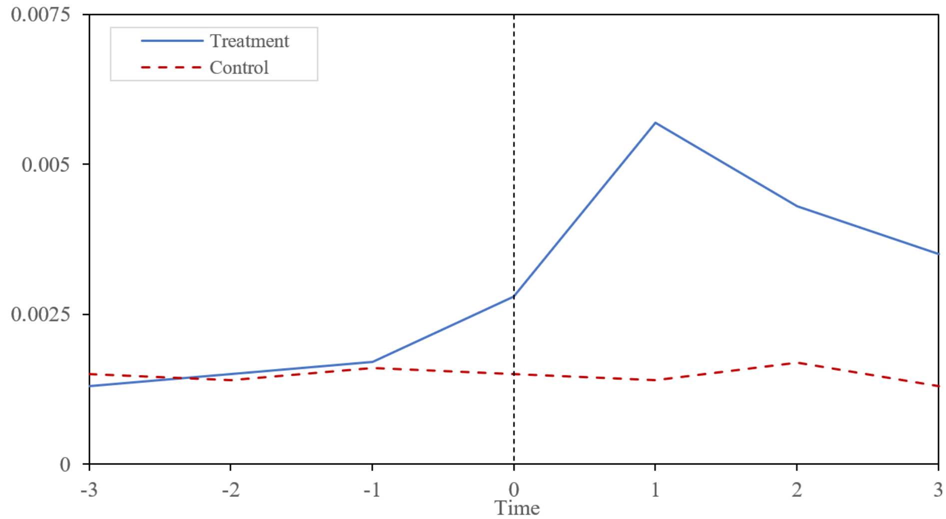
Figure 4 shows the insider trading shares before and after using cloud services. The sample window is 6 years. The first three years are when firms do not use cloud services. The second three years are when firms use cloud services. The total insider trading shares are the total number of shares purchased and sold by insiders during the quarter, scaled by the number of shares outstanding. The treatment firms use cloud services after time zero. The control firms do not use cloud services before and after time zero. For each treatment firm, we match control firms using propensity score matching method based on the characteristics of size, value, and turnover.



Electronic copy available at: https://ssrn.com/abstract=3893300

Figure 5: Insider Trading Returns before and after Using Cloud Services

Figure 5 shows the insider trading returns before and after using cloud services. The sample window is 6 months. The first three months are when firms do not use cloud services. The second three months are when firms use cloud services. The insider trading returns are monthly LSY4 abnormal returns of insider trading. The treatment firms use cloud services after time zero. The control firms do not use cloud services before and after time zero. For each treatment firm, we match control firms using propensity score matching method based on the characteristics of size, value, and turnover.



Electronic copy available at: https://ssrn.com/abstract=3893300

Table 1: Summary Statistics

This table reports the summary statistics of the dependent and independent variables in our main analysis. The sample consists of all publicly listed firms in Shanghai and Shenzhen stock exchanges. Stocks that have become public within the past 12 months and stocks having less than 15 days of trading records during the most recent month are excluded. Financial, real estate, and utility firms are also excluded from the analysis. The sample is further restricted to firms that have at least 100TB cloud data during a quarter. Panel A shows firm characteristics. CD is the amount of cloud data of a firm in a quarter. RET_{t+1} is the one-month-ahead return. SIZE is the firm's market capitalization computed as the logarithm of the market value of the firm's outstanding equity. BM is the logarithm of the firm's book value of equity divided by its market capitalization, where the BM ratio is computed following Fama and French [2008]. Firms with negative book values are excluded from the analysis. ROA is the quarterly operating income scaled by lagged assets. LEV is the quarterly sum of long-term debt and short-term borrowing scaled by total assets. Short-term reversal (STR) is the stock's lagged-one monthly return. MOM is the stock's cumulative return from the start of lagged-twelve month to the end of lagged-two month (skipping the STR month), following Jegadeesh and Titman [1993]. PPE Growth (PG) is the year-over-year quarterly growth in property, plant, and equipment scaled by total assets. Intangible Growth (IG) is the year-over-year quarterly growth in intangible assets scaled by total assets. TO is the quarterly turnover computed as the number of shares traded divided by the total number of shares outstanding in a quarter. ILLIQ is the quarterly illiquidity measure computed as the absolute daily return divided by daily dollar trading volume, averaged in a quarter. IVOL is the idiosyncratic volatility defined as the standard deviation of daily residuals estimated from the regression of daily excess stock returns on the daily market, size, and value factors of Fama and French [1993] in a quarter. SUE is the standardized unexpected earnings defined as actual earnings in the current quarter minus earnings 4 quarters ago, scaled by stock price in the current quarter following Livnat and Mendenhall [2006]. ANA is defined as the number of analysts following the firm in a quarter, and IO is the percentage of tradable shares held by institutional investors in a quarter. Panel B shows characteristics of firm fundamentals, earnings surprise, innovation performance. AG is the quarterly growth of total assets. SG is the quarterly growth of sales. ROA is the quarterly operating income scaled by lagged assets. SUE is the standardized unexpected earnings defined as actual earnings in the current quarter minus earnings 4 quarters ago, scaled by stock price in the current quarter following Livnat and Mendenhall [2006]. CAR is the cumulative abnormal returns over the three-day window surrounding the earnings announcement. Abnormal return is calculated as the raw daily return minus the daily return on size and market-to-book matched portfolio as in Livnat and Mendenhall [2006]. PA is the log of one plus quarterly number of patents applied of the firm. PG is the log of one plus quarterly number of patents granted of the firm. Panel C shows statistics of CDG and other nowcasters. CDG is the annual growth of the amount of cloud data of a firm in a quarter relative to that in the same quarter last year. SEAG is the year-over-year quarterly growth of search volume for firms' products. APPG is the year-over-year quarterly growth of firms' App visiting volume. CUSG is the year-over-year quarterly growth of customer product ratings of firms. EMPG is the year-over-year quarterly growth of employer ratings of firms. CARG is the year-over-year quarterly growth of number of cars in parking lots of firms. SPEG is the year-over-year quarterly growth of credit card spending to products and services of firmsAll variables are winsorized at the 1st and 99th percentiles to mitigate the effect of outliers. The mean, standard deviation (SD), 5% percentile, median, and 95% percentile of each variable are shown in the Table. The mean of each characteristic of all A shares and differences between sample firms' characteristic and all A shares' characteristic are shown in Panel A and B. The percentage of sample firms of CDG and other nowcasters of the market capitalization coverage and the observation of sample firms are shown in Panel C. The sample period is from 2014 to 2021.

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel A: Firm characteristics							
	Mean	SD	P5	P50	P95	Mean of all A shares	Differences
CD	1240.152	11.836	538.681	1124.259	2269.470	N/A	N/A
SIZE	22.377	0.917	20.984	22.907	24.133	21.948	0.429**
BM	0.463	0.249	0.098	0.342	0.940	0.561	-0.098***
ROA	1.412	1.688	-1.006	1.325	5.083	1.105	0.307***
LEV	0.182	0.170	0.000	0.178	0.524	0.167	0.015
STR	0.010	0.115	-0.505	0.010	0.908	0.008	0.002***
MOM	0.071	0.565	-0.433	0.080	1.007	0.065	0.006***
PG	0.020	0.013	0.002	0.015	0.044	0.015	0.005*
IG	0.041	0.034	0.004	0.034	0.105	0.033	0.008*
TO	0.474	0.991	0.056	0.197	8.514	0.352	0.122**
ILLIQ	0.153	0.457	0.009	0.040	13.618	0.182	-0.029*
IVOL	0.021	0.009	0.011	0.052	0.086	0.030	-0.009***
SUE	0.129	2.298	-11.352	0.123	5.509	0.088	0.041**
ANA	7.612	8.686	0.000	5.000	29.000	6.131	1.481**
IO	6.482	9.233	0.001	2.714	26.934	5.715	0.767*

Panel B: Characteristics of firm fundamentals, earnings surprise, innovation performance							
	Mean	SD	P5	P50	P95	Mean of all A shares	Differences
AG	0.182	0.333	-0.239	0.117	0.776	0.141	0.041**
SG	0.183	0.291	-0.094	0.104	0.746	0.143	0.040**
ROA	1.412	1.688	-1.006	1.325	5.083	1.105	0.307***
SUE	0.129	2.298	-11.352	0.123	5.509	0.088	0.041**
CAR	0.309	6.983	-10.407	-0.383	14.058	0.202	0.107**
PA	1.194	1.706	0.000	0.000	2.015	0.428	0.766***
PG	1.030	1.695	0.000	0.000	1.504	0.341	0.689***

Panel C: CDG and other nowcasters							
	Mean	SD	P5	P50	P95	% of mktcap coverage	Observations
CDG	0.182	0.468	-0.296	0.149	0.734	73.189%	30309
SEAG	0.132	0.525	-0.498	0.101	1.133	43.148%	18411
APPG	0.079	0.431	-0.350	0.062	0.800	59.031%	23578
EMPG	0.063	0.699	-0.471	0.041	1.194	37.485%	15016
CUSG	0.075	0.821	-0.698	0.098	1.327	41.624%	16838
CARG	0.108	0.470	-0.474	0.093	1.082	38.987%	16316
SPEG	0.071	0.526	-0.678	0.065	1.008	60.975%	25615

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 2: Nowcasting and forecasting firm fundamentals

This table reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q. The dependent variables include return on assets (ROA), growth of total assets (AG), growth of sales (SG), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from second quarter of 2014 to second quarter of 2021.

	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}	PA_{q+1}	PA_{q+2}	PG_{q+1}	PG_{q+2}
CDG_{q+1}	0.765***	0.522***	0.302***	0.231***	0.071***	0.047***	0.308***	0.192***	0.164***	0.126***
	(5.12)	(3.15)	(4.89)	(3.49)	(4.12)	(2.60)	(4.70)	(3.33)	(4.07)	(3.21)
BM_q	-0.703***	-0.499*	-0.070**	-0.136***	-0.020	-0.066**	-0.051	-0.090**	-0.013	-0.043
	(-2.60)	(-1.93)	(-2.55)	(-2.97)	(-0.96)	(-2.28)	(-1.58)	(-1.98)	(-0.63)	(-1.45)
ROA_q	4.189	4.107**	0.574***	1.132***	0.405***	0.497**	0.869***	1.538***	0.553***	0.704***
	(1.40)	(2.39)	(5.12)	(4.52)	(2.75)	(2.41)	(6.88)	(6.95)	(3.74)	(3.15)
LEV_q	-0.913***	-1.074***	-0.024	-0.072	-0.032	-0.096**	-0.017	-0.047	-0.022	-0.055
	(-4.01)	(-4.57)	(-0.83)	(-1.56)	(-1.14)	(-2.07)	(-0.55)	(-1.05)	(-0.78)	(-1.30)
PG_q	-0.472*	-0.276	0.009	0.052	-0.016	0.025	0.014	0.069	-0.021	0.030
	(-1.84)	(-0.93)	(0.32)	(0.94)	(-0.39)	(0.32)	(0.42)	(1.16)	(-0.53)	(0.45)
IG_q	-0.051	0.561	0.098	0.271	0.059	0.091	0.070	0.190	0.041	0.060
	(-0.06)	(0.65)	(0.82)	(1.46)	(0.44)	(0.44)	(0.47)	(0.87)	(0.32)	(0.27)
SUE_q	-0.008	0.099***	0.017***	0.051***	0.010**	0.016**	0.011**	0.033***	0.007	0.010
	(-0.26)	(3.01)	(3.46)	(6.67)	(1.98)	(2.22)	(2.28)	(4.20)	(1.27)	(1.41)
$SIZE$	0.098	0.016	-0.007	-0.032*	-0.014	-0.048***	-0.009	-0.042**	-0.016**	-0.059***
	(1.27)	(0.23)	(-0.59)	(-1.73)	(-1.54)	(-2.77)	(-0.79)	(-2.31)	(-1.98)	(-3.69)
STR	0.329	0.055	-0.023	-0.099**	-0.042*	-0.154***	-0.030	-0.127***	-0.054***	-0.189***
	(1.64)	(0.28)	(-0.60)	(-2.20)	(-1.87)	(-3.17)	(-1.06)	(-2.74)	(-2.70)	(-4.62)
MOM	0.176***	0.150***	0.019***	0.033***	0.026***	0.040***	0.012**	0.022*	0.018**	0.024*
	(3.94)	(3.22)	(3.29)	(3.06)	(3.28)	(3.22)	(2.13)	(1.83)	(2.29)	(1.91)
TO	0.061	0.046	0.007	0.011	0.009	0.013	0.009	0.015	0.011	0.017
	(1.01)	(0.86)	(0.80)	(0.75)	(0.86)	(0.78)	(1.03)	(1.04)	(1.18)	(1.01)
$ILLIQ$	2.862	2.868	0.456	0.783	0.286	0.358	0.253	0.487	0.210	0.214
	(0.12)	(0.20)	(0.45)	(0.42)	(0.23)	(0.21)	(0.27)	(0.27)	(0.16)	(0.15)
$IVOL$	-3.427***	-1.860***	0.069	0.386***	-0.104	0.164	0.088*	0.499***	-0.147**	0.199
	(-7.32)	(-4.10)	(1.22)	(3.63)	(-1.40)	(1.12)	(1.79)	(4.34)	(-2.09)	(1.62)
ANA	0.002	0.003	0.000	0.000	-0.001**	-0.001*	0.000	0.000	-0.001***	-0.001**
	(0.75)	(0.86)	(-0.79)	(-0.41)	(-2.49)	(-1.67)	(-1.14)	(-0.58)	(-3.78)	(-2.19)
IO	0.011***	0.010***	0.001***	0.002***	0.001***	0.002***	0.001**	0.001**	0.001*	0.001**
	(3.39)	(3.32)	(3.84)	(3.41)	(2.79)	(3.24)	(2.45)	(2.13)	(1.74)	(2.13)
AG_q			0.280***	0.187***						
			(5.11)	(3.55)						
SG_q					0.509***	0.401***				
					(9.26)	(5.97)				
PA_q							0.280***	0.198***		
							(4.73)	(2.97)		
PG_q									0.191***	0.135***
									(3.55)	(3.51)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	29703	29388	28490	28188	29096	28788	21216	20991	21216	20991
Adj. R2	0.55	0.48	0.42	0.38	0.33	0.29	0.22	0.18	0.17	0.13

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 3: Nowcasting and forecasting firm fundamentals after controlling nowcasters

This table reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG and nowcasters in quarter q+1 and other control variables in quarter q. The dependent variables include return on assets (ROA), growth of total assets (AG), growth of sales (SG), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). The nowcasters include the year-over-year quarterly growth of search volume for firms' products ($SEAG$), the year-over-year quarterly growth of firms' App visiting volume ($APPG$), the year-over-year quarterly growth of customer product ratings of firms ($CUSG$), the year-over-year quarterly growth of employer ratings of firms ($EMPG$), the year-over-year quarterly growth of number of cars in parking lots of firms ($CARG$), and the year-over-year quarterly growth of credit card spending to products and services of firms ($SPEG$). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the firm and year-quarter levels are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from second quarter of 2014 to second quarter of 2021.

	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}	PA_{q+1}	PA_{q+2}	PG_{q+1}	PG_{q+2}
CDG_{q+1}	0.470***	0.371***	0.187***	0.148***	0.043***	0.035**	0.226***	0.174**	0.099***	0.082**
	(3.60)	(2.94)	(3.47)	(2.86)	(2.90)	(2.41)	(3.19)	(2.56)	(2.82)	(2.34)
$SEAG_{q+1}$	0.238***	0.186**	0.106***	0.081**	0.026**	0.020*	0.007	0.005	0.013	0.010
	(2.68)	(2.19)	(2.75)	(2.19)	(2.25)	(1.84)	(0.04)	(0.03)	(0.25)	(0.20)
$APPG_{q+1}$	0.128	0.105	0.060	0.049	0.014	0.011	0.060	0.047	0.036	0.029
	(1.01)	(0.81)	(1.34)	(1.07)	(1.09)	(0.85)	(1.06)	(0.88)	(0.73)	(0.57)
$EMPG_{q+1}$	0.181*	0.146	0.072**	0.057	0.018	0.015	0.123*	0.095	0.068	0.053
	(1.78)	(1.37)	(2.03)	(1.60)	(1.57)	(1.29)	(1.92)	(1.48)	(1.48)	(1.22)
$CUSG_{q+1}$	0.145	0.117	0.058	0.045	0.015	0.012	0.041	0.034	0.016	0.013
	(1.07)	(0.87)	(1.60)	(1.27)	(1.20)	(0.93)	(0.75)	(0.61)	(0.54)	(0.41)
$CARG_{q+1}$	0.260***	0.213**	0.114***	0.089**	0.025**	0.020**	0.084	0.069	0.042	0.034
	(2.74)	(2.17)	(3.04)	(2.46)	(2.47)	(2.02)	(1.39)	(1.11)	(1.37)	(1.06)
$SPEG_{q+1}$	0.069	0.054	0.025	0.020	0.006	0.005	0.136**	0.112*	0.071*	0.058
	(0.59)	(0.48)	(0.69)	(0.55)	(0.60)	(0.47)	(2.21)	(1.76)	(1.91)	(1.57)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	9605	9274	8721	8421	10162	9812	6861	6624	6861	6624
Adj. R2	0.71	0.61	0.53	0.47	0.41	0.35	0.30	0.25	0.24	0.21

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 4: Nowcasting and forecasting earnings surprise

This table reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q. The dependent variables include standardized unexpected earnings (SUE(and earnings announcement abnormal returns (CAR). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from second quarter of 2014 to second quarter of 2021.

	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
CDG_{q+1}	0.214***	0.161***	2.130***	1.364**
	(3.55)	(2.73)	(3.18)	(2.28)
BM_q	-0.173	-0.029	0.022	0.132
	(-0.45)	(-0.11)	(0.07)	(0.41)
ROA_q	4.901**	5.941***	-6.528**	1.064
	(2.05)	(3.45)	(-2.46)	(0.41)
LEV_q	-0.022	-0.054	-0.636	-0.745**
	(-0.10)	(-0.29)	(-1.43)	(-2.16)
PG_q	-0.349	-0.002	-0.082	0.298
	(-1.21)	(-0.01)	(-0.14)	(0.53)
IG_q	1.425*	1.182	-0.606	0.413
	(1.69)	(1.20)	(-0.30)	(0.25)
SUE_q	0.362***	0.287***	-0.105	0.072
	(3.96)	(3.70)	(-1.55)	(1.26)
SIZE	-0.022	-0.139	-0.292**	-0.318*
	(-0.35)	(-1.61)	(-2.21)	-(1.94)
STR	-0.067	-0.485*	-1.003**	-1.140***
	(-0.40)	(-1.79)	(-2.57)	(-2.58)
MOM	0.173***	0.159***	-0.034	0.004
	(2.82)	(2.66)	(-0.39)	(0.06)
TO	0.058	0.052	-0.012	0.001
	(0.77)	(0.68)	(-0.09)	(0.01)
$ILLIQ$	3.565	3.954	-4.393	0.791
	(0.18)	(0.31)	(-0.21)	(0.04)
$IVOL$	-2.498***	-0.017	-0.525	2.174*
	(-4.59)	(-0.02)	(-0.59)	(1.91)
ANA	-0.008***	-0.007***	0.001	0.001
	(-3.62)	(-2.92)	(0.05)	(0.08)
IO	0.004*	0.006**	0.012**	0.002
	(1.70)	(2.02)	(2.15)	(0.44)
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
N	28793	28488	29399	29088
Adj. R2	0.40	0.32	0.09	0.07

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 5: Nowcasting and forecasting earnings surprise after controlling nowcasters

This table reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG and nowcasters in quarter q+1 and other control variables in quarter q. The dependent variables include standardized unexpected earnings (SUE) and earnings announcement abnormal returns (CAR). The nowcasters are defined in Table 3. We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from second quarter of 2014 to second quarter of 2021

	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
CDG_{q+1}	0.133**	0.104*	1.418**	1.120*
	(2.44)	(1.90)	(2.28)	(1.77)
$SEAG_{q+1}$	0.074**	0.060*	0.700*	0.566
	(2.01)	(1.67)	(1.71)	(1.39)
$APPG_{q+1}$	0.040	0.033	0.420	0.325
	(0.95)	(0.76)	(0.80)	(0.65)
$EMPG_{q+1}$	0.057	0.045	0.575	0.455
	(1.59)	(1.26)	(1.33)	(1.03)
$CUSG_{q+1}$	0.040	0.031	0.372	0.310
	(1.07)	(0.82)	(0.89)	(0.74)
$CARG_{q+1}$	0.089**	0.069*	0.854*	0.687
	(2.21)	(1.79)	(1.85)	(1.52)
$SPEG_{q+1}$	0.016	0.013	0.184	0.153
	(0.49)	(0.39)	(0.39)	(0.31)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
N	9396	9072	10033	9687
Adj. R2	0.56	0.51	0.12	0.11

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 6: Univariate Portfolio Analysis

Panel A reports the average monthly excess returns and alphas on the value-weighted portfolios of stocks sorted by the CDG. Panel B reports the average monthly excess returns and alphas on the equal-weighted portfolios of stocks sorted by the CDG. At each month t from April 2014 to June 2021, individual stocks of companies are sorted into quintiles based on CDG at quarter q-1, and are held for the next one quarter. P1 is the portfolio of stocks with the lowest CDG and P5 is the portfolio of stocks with the highest CDG. L/S is a zero-cost portfolio that buys stocks in quintile 5 (highest CDG) and sells stocks in quintile 1 (lowest CDG). All returns and alphas are expressed in percentage. Excess return is the raw return of the portfolio over the risk-free rate. Alpha is the intercept from a time-series regression of monthly excess returns on the factors of alternative models: China q-factor model (HXZ) based on Hou et al. [2015], China five-factor model (FF5) based on Fama and French [2015], Liu et al. [2019] China three-factor model (LSY3), and Liu et al. [2019] China four-factor model (LSY4).EA represents average excess returns in earnings announcement months. Non-EA represents average excess returns in non-earnings announcement months. Newey and West [1987] adjusted t-statistics are given in parentheses.Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from April 2014 to June 2021.

Panel A: Value-weighted CDG-sorted quintile portfolios

Rank	Excess	HXZ	FF5	LSY3	LSY4	EA	Non-EA
P1	0.046	-0.603***	-0.511***	-0.481***	-0.487***	0.031	0.015
	(0.11)	(-2.70)	(-2.86)	(-2.68)	(-2.59)	(0.09)	(0.05)
P2	0.264	-0.354**	-0.440*	-0.316**	-0.401**	0.183	0.082
	(0.49)	(-2.48)	(-1.73)	(-2.45)	(-1.99)	(0.39)	(0.20)
P3	0.353	-0.208	-0.141	-0.246*	-0.283	0.236	0.118
	(1.50)	(-0.49)	(-1.34)	(-1.93)	(-0.87)	(1.15)	(0.65)
P4	0.705**	-0.105	-0.041	0.078	-0.175	0.463	0.242
	(2.06)	(-0.38)	(-0.18)	(0.40)	(-0.79)	(1.56)	(0.91)
P5	0.897***	0.148*	0.191*	0.197	0.136	0.592***	0.305
	(4.39)	(1.95)	(1.73)	(1.39)	(1.04)	(3.34)	(1.93)
L/S	0.851***	0.751***	0.703***	0.677***	0.623***	0.559***	0.291*
	(4.17)	(3.99)	(3.85)	(3.36)	(2.91)	(3.16)	(1.85)

Panel B: Equal-weighted CDG-sorted quintile portfolios

Rank	Excess	HXZ	FF5	LSY3	LSY4	EA	Non-EA
P1	0.092	-0.721***	-0.688***	-0.741***	-0.654***	0.069	0.024
	(0.27)	(-4.31)	(-3.70)	(-3.78)	(-3.26)	(0.23)	(0.10)
P2	0.395**	-0.175***	-0.630***	-0.301**	-0.558*	0.288**	0.106
	(2.50)	(-3.48)	(-2.93)	(-2.03)	(-1.66)	(2.08)	(0.93)
P3	0.481**	0.060**	-0.531*	-0.077	-0.459	0.343**	0.138
	(2.63)	(2.02)	(-1.87)	(-1.49)	(-1.06)	(2.13)	(1.02)
P4	0.582***	0.226	-0.246	0.006	-0.092	0.421***	0.161
	(3.43)	(1.16)	(-0.03)	(0.13)	(-0.82)	(2.82)	(1.29)
P5	1.294***	0.312*	0.302*	0.181	0.222	0.966***	0.328*
	(5.17)	(1.92)	(1.68)	(1.62)	(0.91)	(4.37)	(1.83)
L/S	1.202***	1.032***	0.991***	0.922***	0.876***	0.898***	0.304*
	(5.11)	(4.91)	(4.93)	(4.88)	(3.63)	(4.33)	(1.80)

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 7: Fama-MacBeth Cross-Sectional Regressions

This table reports the Fama and MacBeth [1973] cross-sectional regression results. Panel A reports the Fama and MacBeth [1973] cross-sectional regressions of CDG. The CDG and other accounting variables in quarter q-1 are matched to monthly stock returns in quarter q. The monthly price-based variables are calculated using the last non-missing observations prior to each month. The dependent variable is the firm's future raw return in the first two columns, the firm's future excess return over its value-weighted industry peers' return (Column 3), or the firm's future excess return over its value-weighted geographic peers' return (Column 4). Panel B reports the Fama and MacBeth [1973] cross-sectional regressions of CDG and other nowcasters. The dependent variable is the firm's future raw return. The nowcasters are defined in Table 3. We control for the industry and geography fixed effects following the CSRC industry classification and China province classification. All returns are expressed in percentage. The CDG and other firm-specific characteristics are defined in Panel A of Table 1. All explanatory variables are generated using the last non-missing available observation for each quarter q-1. Cross-sectional regressions are run every calendar month, and the time-series standard errors are corrected for heteroskedasticity and autocorrelation. Newey and West [1987] adjusted t-statistics are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from April 2014 to June 2021.

Panel A: Fama and MacBeth regressions of CDG

Independent Variables	RET	RET	RET-INDRET	RET-GEORET
CDG	0.505***	0.469***	0.437***	0.370***
	(4.08)	(3.90)	(3.41)	(3.29)
SIZE		-0.592*	-0.551**	-0.542**
		(-1.96)	(-1.97)	(-2.15)
BM		0.264	0.282	0.264
		(0.55)	(0.51)	(0.45)
STR		-1.992**	-1.874***	-2.020***
		(-2.43)	(-2.85)	(-2.90)
MOM		-0.200	-0.182	-0.189
		(-0.95)	(-1.14)	(-1.08)
ROA		13.317***	11.197***	12.595***
		(2.89)	(2.80)	(2.80)
LEV		-0.562	-0.510	-0.601
		(-1.27)	(-1.30)	(-1.37)
PG		-0.526	-0.538	-0.505
		(-0.51)	(-0.56)	(-0.64)
IG		0.535	0.575	0.596
		(0.71)	(0.69)	(0.70)
TO		-0.071	-0.066	-0.073
		(-0.23)	(-0.23)	(-0.21)
ILLIQ		9.148	9.010	7.553
		(0.24)	(0.27)	(0.25)
IVOL		-3.531**	-2.878**	-2.579**
		(-2.18)	(-2.25)	(-2.14)
SUE		0.089***	0.092***	0.099***
		(2.86)	(3.38)	(3.11)
ANA		-0.004	-0.004	-0.004
		(-0.32)	(-0.31)	(-0.32)
IO		0.021**	0.022**	0.021**
		(2.19)	(2.49)	(2.39)
Industry FE	Y	Y	N	Y
Geography FE	Y	Y	Y	N
N	90,926	88198	88198	88198
Adj. R2	0.06	0.09	0.07	0.07

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel B: Fama and MacBeth regressions of CDG and other nowcasters

	RET	RET	RET	RET	RET	RET	RET
CDG	0.356***	0.434***	0.377***	0.395***	0.337***	0.438***	0.291***
	(2.91)	(3.64)	(3.20)	(3.30)	(2.84)	(3.65)	(2.72)
SEAG	0.178**						0.144*
	(2.35)						(1.90)
APPG		0.098					0.074
		(0.96)					(0.85)
EMPG			0.149				0.125
			(1.52)				(1.25)
CUSG				0.111			0.090
				(1.01)			(0.91)
CARG					0.205**		0.169**
					(2.38)		(2.07)
SPEG						0.050	0.042
						(0.53)	(0.46)
Controls	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Geography FE	Y	Y	Y	Y	Y	Y	Y
N	44099	57329	48509	52919	39689	61739	27374
Adj. R2	0.10	0.09	0.10	0.09	0.10	0.09	0.13

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 8: International evidence

This table reports international evidence. Panel A reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q in Indonesia, Japan, Malaysia, and Singapore. The dependent variables include return on assets (ROA), growth of total assets (AG), and growth of sales (SG). Panel B reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q in Indonesia, Japan, Malaysia, and Singapore.The dependent variables include standardized unexpected earnings (SUE) and earnings announcement abnormal returns (CAR). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Panel C reports the average monthly long-short excess returns and long-short alphas on the value-weighted portfolios of stocks sorted by the CDG in Indonesia, Japan, Malaysia, and Singapore. Coefficients marked with *, **, and *** are significant at the 10%, the 5%, and the 1% level, respectively. The sample period is from second quarter of 2014 to second quarter of 2021

Panel A: nowcasting and forecasting firm fundamentals in four countries

	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}
Indonesia						
CDG_{q+1}	0.408**	0.285**	0.184***	0.138*	0.040**	0.030*
	(2.57)	(2.09)	(2.64)	(1.78)	(2.37)	(1.95)
Japan						
CDG_{q+1}	0.762***	0.617***	0.382***	0.276***	0.073***	0.060***
	(4.19)	(3.10)	(3.81)	(2.91)	(4.07)	(2.95)
Malaysia						
CDG_{q+1}	0.215**	0.141	0.101**	0.077	0.019**	0.014
	(2.14)	(1.50)	(2.07)	(1.55)	(2.06)	(1.54)
Singapore						
CDG_q+1	0.568***	0.447**	0.304***	0.205**	0.058***	0.045**
	(2.82)	(2.43)	(2.87)	(2.11)	(3.13)	(2.36)
Controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel B: nowcasting and forecasting earnings surprises in four countries				
	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
Indonesia				
CDG_{q+1}	0.145**	0.113*	1.382***	0.896*
	(2.48)	(1.89)	(2.65)	(1.69)
Japan				
CDG_{q+1}	0.327***	0.225***	2.743***	1.745*
	(4.12)	(3.08)	(3.83)	(1.84)
Malaysia				
CDG_{q+1}	0.099**	0.079	0.982**	0.648
	(2.17)	(1.48)	(2.03)	(1.57)
Singapore				
CDG_q+1	0.202***	0.138**	1.965***	1.220*
	(3.19)	(2.37)	(3.14)	(1.67)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Panel C: long-short excess returns and alphas in four countries					
Value-weighted	Excess	CAPM	FF3	FF5	FF6
Indonesia					
L/S	0.398***	0.351***	0.328***	0.317**	0.291*
	(2.80)	(2.68)	(2.59)	(2.26)	(1.96)
Japan					
L/S	0.575***	0.508***	0.475***	0.458***	0.421**
	(3.42)	(3.27)	(3.15)	(2.75)	(2.39)
Malaysia					
L/S	0.328**	0.290**	0.271**	0.261*	0.240
	(2.25)	(2.15)	(2.07)	(1.81)	(1.57)
Singapore					
L/S	0.439***	0.387***	0.362***	0.349**	0.321**
	(2.96)	(2.83)	(2.72)	(2.38)	(2.06)

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 9: Robustness in subsamples

This table presents results from the value-weighted portfolios in different stock subsamples. First, the stock subsamples are partitioned by manufacturing industry and Non-manufacturing industries. Second, the stock subsamples are partitioned by the Top5 provinces and Non-Top5 provinces. Third, the stock subsamples are partitioned into state enterprises and private enterprises. Fourth, the stock subsamples are partitioned into before COVID-19 period and after COVID-19 period. Panel A reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q in different stock subsamples. The dependent variables include return on assets (ROA), growth of total assets (AG), growth of sales (SG), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). Panel B reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q in different stock subsamples. The dependent variables include standardized unexpected earnings (SUE) and earnings announcement abnormal returns (CAR). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Panel C reports the average monthly long-short excess returns and long-short alphas on the value-weighted portfolios of stocks sorted by the CDG in different stock subsamples. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from April 2014 to June 2021.

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel A: nowcasting and forecasting firm fundamentals in subsamples										
	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}	PA_{q+1}	PA_{q+2}	PG_{q+1}	PG_{q+2}
Manufacturing										
CDG_{q+1}	0.863***	0.576***	0.332***	0.275***	0.080***	0.053***	0.322***	0.211***	0.191***	0.140***
	(5.34)	(3.29)	(5.14)	(3.72)	(4.35)	(2.65)	(5.00)	(3.45)	(4.23)	(3.28)
Non-Manufacturing										
CDG_{q+1}	0.659***	0.459**	0.257***	0.205***	0.056***	0.040**	0.243***	0.168***	0.139***	0.106**
	(3.89)	(2.36)	(3.75)	(2.71)	(3.17)	(1.96)	(3.63)	(2.60)	(3.12)	(2.45)
Top5 provinces										
CDG_{q+1}	0.880***	0.610***	0.366***	0.275***	0.085***	0.054**	0.352***	0.230***	0.205***	0.148***
	(4.99)	(3.14)	(4.99)	(3.51)	(4.07)	(2.53)	(4.59)	(3.37)	(3.89)	(3.24)
Non-Top5 provinces										
CDG_{q+1}	0.722***	0.521**	0.278***	0.223***	0.063***	0.044**	0.286***	0.185***	0.156***	0.123***
	(4.34)	(2.55)	(4.27)	(3.12)	(3.75)	(2.31)	(4.08)	(2.91)	(3.38)	(2.70)
State enterprises										
CDG_{q+1}	0.591***	0.424**	0.231***	0.188***	0.054***	0.036**	0.242***	0.156***	0.136***	0.102**
	(3.88)	(2.48)	(3.79)	(2.72)	(3.22)	(2.03)	(3.52)	(2.62)	(2.99)	(2.49)
Private enterprises										
CDG_{q+1}	0.984***	0.648***	0.389***	0.295***	0.090***	0.060***	0.385***	0.255***	0.209***	0.164***
	(5.96)	(3.48)	(5.68)	(4.02)	(4.93)	(2.92)	(5.38)	(3.70)	(4.48)	(3.54)
Before COVID-19										
CDG_{q+1}	0.740***	0.503***	0.292***	0.231***	0.070***	0.045**	0.304***	0.183***	0.161***	0.120***
	(5.07)	(3.03)	(4.71)	(3.37)	(4.03)	(2.55)	(4.60)	(3.20)	(4.04)	(3.05)
After COVID-19										
CDG_{q+1}	0.838***	0.573***	0.336***	0.254***	0.081***	0.051***	0.346***	0.213***	0.188***	0.137***
	(5.20)	(3.23)	(4.89)	(3.49)	(4.26)	(2.75)	(4.73)	(3.27)	(4.04)	(3.31)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel B: nowcasting and forecasting earnings surprises in subsamples				
	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
Manufacturing				
CDG_q+1	0.243***	0.184***	2.383***	1.612**
	(3.72)	(2.80)	(3.48)	(2.38)
Non-Manufacturing				
CDG_q+1	0.175 ***	0.144**	1.925***	1.228**
	(2.81)	(2.15)	(2.55)	(1.83)
Top5 provinces				
CDG_q+1	0.252***	0.196***	2.576***	1.730**
	(3.65)	(2.80)	(3.21)	(2.38)
Non-Top5 provinces				
CDG_q+1	0.209***	0.154**	1.997***	1.311**
	(3.02)	(2.31)	(2.70)	(2.00)
State enterprises				
CDG_q+1	0.166***	0.134**	1.763***	1.135*
	(2.77)	(2.17)	(2.61)	(1.86)
Private enterprises				
CDG_q+1	0.273***	0.210***	2.785***	1.711***
	(4.25)	(3.02)	(3.61)	(2.68)
Before COVID-19				
CDG_{q+1}	0.206***	0.160***	2.027***	1.341**
	(3.43)	(2.71)	(3.17)	(2.20)
After COVID-19				
CDG_{q+1}	0.222***	0.167***	2.180***	1.405**
	(3.40)	(2.63)	(3.09)	(2.15)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel C: long-short excess returns and alphas in subsamples					
Value-weighted	Excess	HXZ	FF5	LSY3	LSY4
Manufacturing					
L/S	0.933***	0.782***	0.746***	0.814***	0.720***
	(4.17)	(4.26)	(4.44)	(3.79)	(3.11)
Non-Manufacturing					
L/S	0.723***	0.596***	0.619***	0.592***	0.503**
	(3.20)	(3.30)	(2.92)	(2.73)	(2.43)
Top5 provinces					
L/S	1.011***	0.755***	0.787***	0.761***	0.694***
	(4.14)	(4.41)	(4.04)	(3.66)	(3.26)
Non-Top5 provinces					
L/S	0.804***	0.654 ***	0.615***	0.562***	0.547**
	(3.56)	(3.24)	(3.31)	(2.97)	(2.51)
State enterprises					
L/S	0.672 ***	0.591***	0.567***	0.520**	0.504**
	(3.24)	(2.99)	(2.96)	(2.48)	(2.22)
Private enterprises					
L/S	1.072***	0.856***	0.840***	0.752***	0.769***
	(4.74)	(4.60)	(4.47)	(4.00)	(3.29)
Before COVID-19					
L/S	0.843***	0.722***	0.684***	0.656***	0.622***
	(4.13)	(3.84)	(3.75)	(3.24)	(2.91)
After COVID-19					
L/S	0.936***	0.797***	0.741***	0.744***	0.675***
	(4.03)	(3.77)	(3.69)	(3.26)	(2.81)

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 10: CDG predictability across different firms

This table presents results of CDG predictability across different firms. We split the stock sample into two equal subsamples based on the market capitalization (Large/Small), the institutional ownership (High/Low), or the analyst coverage (High/Low). Panel A reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q across different firms. The dependent variables include return on assets (ROA), growth of total assets (AG), growth of sales (SG), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). Panel B reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG in quarter q+1 and other control variables in quarter q across different firms. The dependent variables include standardized unexpected earnings (SUE) and earnings announcement abnormal returns (CAR). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Panel C reports the average monthly long-short excess returns and long-short alphas on the value-weighted portfolios of stocks sorted by the CDG across different firms. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from April 2014 to June 2021.

Panel A: nowcasting and forecasting firm fundamentals in subsamples

	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}	PA_{q+1}	PA_{q+2}	PG_{q+1}	PG_{q+2}
Large										
CDG_{q+1}	0.597***	0.418**	0.232***	0.179***	0.055***	0.036**	0.242***	0.149***	0.124***	0.096**
	(3.91)	(2.42)	(3.69)	(2.73)	(3.09)	(2.00)	(3.63)	(2.53)	(3.21)	(2.45)
Small										
CDG_{q+1}	0.926***	0.640***	0.364 ***	0.287***	0.088***	0.057***	0.375***	0.237***	0.205***	0.154***
	(6.15)	(3.89)	(5.94)	(4.23)	(5.04)	(3.25)	(5.81)	(4.00)	(4.98)	(4.00)
High IO										
CDG_{q+1}	0.679***	0.465***	0.261***	0.207***	0.063***	0.042**	0.266***	0.170***	0.145***	0.107***
	(4.46)	(2.79)	(4.35)	(3.12)	(3.67)	(2.26)	(4.17)	(2.87)	(3.56)	(2.79)
Low IO										
CDG_{q+1}	0.878***	0.581***	0.342***	0.255***	0.080***	0.052***	0.354***	0.219***	0.182***	0.140***
	(5.77)	(3.60)	(5.55)	(3.88)	(4.57)	(2.89)	(5.21)	(3.68)	(4.65)	(3.56)
High coverage										
CDG_{q+1}	0.614***	0.439***	0.248***	0.192***	0.058***	0.038**	0.249***	0.161***	0.***	0.102***
	(4.33)	(2.64)	(3.97)	(2.91)	(3.31)	(2.21)	(3.87)	(2.79)	(3.40)	(2.69)
Low coverage										
CDG_{q+1}	0.900***	0.607***	0.359***	0.275***	0.082***	0.055***	0.363***	0.226***	0.197***	0.147***
	(5.96)	(3.68)	(5.85)	(4.05)	(4.88)	(3.04)	(5.49)	(3.91)	(4.72)	(3.82)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel B: nowcasting and forecasting earnings surprises in subsamples				
	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
Large				
CDG_{q+1}	0.173***	0.131**	1.708***	1.149*
	(2.92)	(2.26)	(2.66)	(1.91)
Small				
CDG_{q+1}	0.246***	0.188***	2.462***	1.577***
	(4.12)	(3.14)	(3.72)	(2.65)
High IO				
CDG_{q+1}	0.194***	0.146**	1.922***	1.259**
	(3.30)	(2.46)	(2.96)	(2.08)
Low IO				
CDG_{q+1}	0.226***	0.173***	2.267***	1.449**
	(3.84)	(2.97)	(3.34)	(2.45)
High coverage				
CDG_{q+1}	0.192***	0.143**	1.902***	1.200**
	(3.14)	(2.43)	(2.80)	(1.98)
Low coverage				
CDG_{q+1}	0.236***	0.182***	2.430***	1.554***
	(4.02)	(3.04)	(3.64)	(2.58)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel C: long-short excess returns and alphas in subsamples					
Value-weighted	Excess	HXZ	FF5	LSY3	LSY4
Large					
L/S	0.527**	0.390**	0.377**	0.367*	0.381*
	(2.27)	(2.25)	(2.31)	(1.75)	(1.72)
Small					
L/S	1.046***	0.908***	0.813***	0.813***	0.744***
	(4.89)	(4.69)	(4.45)	(3.93)	(3.49)
High IO					
L/S	0.629***	0.538***	0.495***	0.500***	0.476**
	(3.01)	(3.19)	(3.04)	(2.84)	(2.24)
Low IO					
L/S	0.980***	0.767***	0.755***	0.733***	0.681***
	(4.61)	(4.17))	(4.14)	(3.48)	(3.26)
High coverage					
L/S	0.572***	0.429***	0.449**	0.419**	0.350*
	(2.72)	(2.61)	(2.38)	(2.18)	(1.76)
Low coverage					
L/S	1.007***	0.815***	0.871***	0.775***	0.696***
	(4.94)	(4.73)	(4.79)	(3.93)	(3.63)

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 11: CDG measures for different service types

This table presents results using CDG on IaaS, PaaS, or SaaS. Panel A reports the results on the regressions of firm fundamentals measured in quarter q+1 or quarter q+2 on the CDG on IaaS, PaaS, or SaaS in quarter q+1 and other control variables in quarter q. The dependent variables include return on assets (ROA), growth of total assets (AG), growth of sales (SG), the log of one plus quarterly number of patents applied (PA), and the log of one plus quarterly number of patents granted (PG). Panel B reports the results on the regressions of earnings surprise measured in quarter q+1 or quarter q+2 on the CDG on IaaS, PaaS, or SaaS in quarter q+1 and other control variables in quarter q. The dependent variables include standardized unexpected earnings (SUE) and earnings announcement abnormal returns (CAR). We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Panel C reports the average monthly long-short excess returns and long-short alphas on the value-weighted portfolios of stocks sorted by the CDG on IaaS, PaaS, or SaaS. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from April 2014 to June 2021.

Panel A: nowcasting and forecasting firm fundamentals using CDG on IaaS, PaaS, or SaaS

	ROA_{q+1}	ROA_{q+2}	AG_{q+1}	AG_{q+2}	SG_{q+1}	SG_{q+2}	PA_{q+1}	PA_{q+2}	PG_{q+1}	PG_{q+2}
IaaS										
CDG_{q+1}	0.720***	0.489***	0.278***	0.224***	0.061***	0.043**	0.263***	0.185***	0.159***	0.111***
	(4.90)	(3.00)	(4.84)	(3.36)	(4.08)	(2.56)	(4.35)	(3.07)	(3.77)	(3.14)
PaaS										
CDG_{q+1}	0.613***	0.427**	0.234***	0.199***	0.057***	0.039**	0.238***	0.160***	0.132***	0.104***
	(4.06)	(2.42)	(4.09)	(2.92)	(3.40)	(2.03)	(3.64)	(2.63)	(3.04)	(2.58)
SaaS										
CDG_{q+1}	0.529***	0.373**	0.200***	0.160***	0.049***	0.032***	0.213***	0.136**	0.115***	0.087**
	(3.45)	(2.18)	(3.52)	(2.62)	(2.86)	(1.92)	(3.16)	(2.37)	(2.86)	(2.32)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Panel B: nowcasting and forecasting earnings surprises using CDG on IaaS, PaaS				
	SUE_{q+1}	SUE_{q+2}	CAR_{q+1}	CAR_{q+2}
IaaS				
CDG_{q+1}	0.200***	0.157**	2.078***	1.329**
	(3.52)	(2.50)	(3.07)	(2.23)
PaaS				
CDG_{q+1}	0.173***	0.136**	1.810**	1.158*
	(2.84)	(2.06)	(2.51)	(1.83)
SaaS				
CDG_{q+1}	0.151***	0.117**	1.455**	0.999*
	(2.60)	(1.96)	(2.19)	(1.66)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Panel C: long-short excess returns and alphas using CDG on IaaS, PaaS, or SaaS					
Value-weighted	Excess	HXZ	FF5	LSY3	LSY4
IaaS					
L/S	0.778***	0.623***	0.575***	0.598***	0.542**
	(3.97)	(3.59)	(3.75)	(2.87)	(2.47)
PaaS					
L/S	0.695***	0.527***	0.517***	0.462***	0.466**
	(3.27)	(3.11)	(3.32)	(2.67)	(2.22)
SaaS					
L/S	0.587***	0.469***	0.481**	0.450**	0.419*
	(2.91)	(2.63)	(2.58)	(2.20)	(1.94)

Electronic copy available at: https://ssrn.com/abstract=3893300

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 12: Insider trading

This table reports the regression results on regressions of insider trading on cloud data growth (CDG) and other control variables. InsiderBuy is the total number of shares purchased by insiders during the quarter, scaled by the number of shares outstanding. InsiderSell is the total number of shares sold by insiders during the quarter, scaled by the number of shares outstanding. InsiderRet is the one-month or three-month LSY4 abnormal returns of insider trading. OppInsiderBuy (OppInsiderSell) is the percentage shares opportunistically purchased (sold) during the quarter. Opportunistic trades are defined as in Cohen et al. [2012]. OppInsiderNet is calculated as OppInsiderBuy minus OppInsiderSell. It is expressed in percentage points. OppInsiderRet is the one-month or three-month LSY4 abnormal returns of opportunistic insider trading. We winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and one standard deviation. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. Newey and West (1987) adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from April 2014 to June 2021.

Panel A: Insider trading

	$InsiderBuy_{q+1}$	$InsiderSell_{q+1}$	$InsiderNet_{q+1}$	$InsiderRet_{1m}$	$InsiderRet_{3m}$
CDG_{q+1}	0.002***	-0.020***	0.022***	0.335***	0.672***
	(4.49)	(-3.60)	(3.87)	(3.37)	(2.96)
Controls	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y

Panel B: Opportunistic insider trading

	$OppInsiderBuy_{q+1}$	$OppInsiderSell_{q+1}$	$OppInsiderNet_{q+1}$	$OppInsiderRet_{1m}$	$OppInsiderRet_{3m}$
CDG_{q+1}	0.001***	-0.015***	0.016***	0.288***	0.570***
	(4.08)	(-3.16)	(3.46)	(2.97)	(2.68)
Controls	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y

Table 13: Diff-in-Diff tests and Insider Trading Shares

This table reports the diff-in-diff tests of total insider trading shares. The sample window is 6 years. The first three years are when firms do not use cloud services. The second three years are when firms use cloud services. The total insider trading shares are the total number of shares purchased and sold by insiders during the quarter, scaled by the number of shares outstanding. Panel A reports the total insider trading shares. Panel B and C report the insider trading buying shares and selling shares, respectively. The dummy variable Treat equals one when the firm uses cloud service, otherwise zero. The control firms do not use cloud services from our sample. For each treatment firm, we match control firms using propensity score matching method based on the characteristics of size, value, and turnover. The dummy variable Post equals one when the firm begin to use cloud services, otherwise zero. Column 1 shows the diff-in-diff tests in full sample. Column 2-4 shows the diff-in-diff tests in IaaS, PaaS, or SaaS sample. The control variables are from Panel A of Table 1. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Total insider trading shares

	Full	IaaS	PaaS	SaaS
Treat * Post	0.175***	0.113***	0.100**	0.077**
	(4.82)	(3.08)	(2.41)	(2.06)
Treat	0.064	0.043	0.038	0.029
	(1.60)	(1.00)	(0.95)	(0.75)
Post	0.050	0.030	0.026	0.023
	(1.08)	(0.73)	(0.57)	(0.53)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Panel B: Insider trading buying shares

	Full	IaaS	PaaS	SaaS
Treat * Post	0.060***	0.038***	0.034**	0.026*
	(4.12)	(2.70)	(2.05)	(1.69)
Treat	0.022	0.014	0.013	0.010
	(1.37)	(0.90)	(0.78)	(0.62)
Post	0.017	0.010	0.009	0.008
	(0.97)	(0.59)	(0.48)	(0.44)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Panel C: Insider trading selling shares

	Full	IaaS	PaaS	SaaS
Treat * Post	0.116***	0.074***	0.066**	0.051*
	(4.55)	(3.00)	(2.26)	(1.86)
Treat	0.042	0.028	0.025	0.019
	(1.60)	(0.98)	(0.90)	(0.70)
Post	0.033	0.020	0.017	0.015
	(1.04)	(0.73)	(0.54)	(0.49)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300

Table 14: Diff-in-Diff tests and Insider Trading Returns

This table reports the diff-in-diff tests of insider trading returns. The insider trading returns are one-month or three-month LSY4 abnormal returns of insider trading. The dummy variable Treat equals one when the firm uses cloud service, otherwise zero. The control firms do not use cloud services from our sample. For each treatment firm, we match control firms using propensity score matching method based on the characteristics of size, value, and turnover. The dummy variable Post equals one when the firm begin to use cloud services, otherwise zero. Panel A shows one-month LSY4 abnormal returns of insider trading. Panel B shows three-month LSY4 abnormal returns of insider trading. Column 1 shows the diff-in-diff tests in full sample. Column 2-4 shows the diff-in-diff tests in IaaS, PaaS, or SaaS sample. The t-statistics of robust standard errors clustered at the industry and year-quarter levels are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: One-month abnormal returns

	Full	IaaS	PaaS	SaaS
Treat * Post	0.005***	0.003***	0.003**	0.002**
	(4.88)	(3.17)	(2.48)	(2.42)
Treat	0.002	0.001	0.001	0.001
	(1.43)	(0.99)	(0.80)	(0.71)
Post	0.001	0.001	0.001	0.000
	(0.84)	(0.58)	(0.43)	(0.40)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Panel B: Three-month abnormal returns

	Full	IaaS	PaaS	SaaS
Treat * Post	0.012***	0.008***	0.006***	0.005**
	(4.37)	(2.92)	(2.60)	(2.11)
Treat	0.003	0.002	0.002	0.001
	(0.97)	(0.66)	(0.56)	(0.47)
Post	0.004	0.003	0.002	0.002
	(1.25)	(0.87)	(0.74)	(0.58)
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Electronic copy available at: https://ssrn.com/abstract=3893300