Breaking the Language Barriers?

Machine Translation Technology and Analysts' Forecasts for Multinational Firms*

Bingxu Fang
Singapore Management University
bxfang@smu.edu.sg

Pengkai Lin
Singapore Management University
pklin@smu.edu.sg

October 3, 2023

* We appreciate the useful comments from Philip Berger, Qiang Cheng, Gus De Franco, Ole-Kristian Hope, Sterling Huang, Leonard Li, Yun Lou, Minjae Koo, Shaphan Ng, Dushyant Vyas (discussant), Rencheng Wang, Holly Yang, Amanda Aw Yong, Bohui Zhang (discussant), Wuyang Zhao, and Christina Zhu. We thank seminar participants at Singapore Management University, University of Melbourne, University of New South Wales, and Xi'an Jiaotong University, and conference participants at the 4th Analyst Research conference and the 2023 Journal of Finance and Data Science Conference for valuable feedback. We also thank Panny Du, Sophie Liu, and several anonymous current or prior analysts for helpful discussions regarding the work process of analysts and translators in investment banks. We thank Ruilin Wang and Yue Zhao for excellent research assistance. Bingxu Fang gratefully acknowledges the financial support of the Della Suantio Fellowship. All errors are our own.

Breaking the Language Barriers?

Machine Translation Technology and Analysts' Forecasts for Multinational Firms

ABSTRACT

We study the impact of machine translation technology on analysts' forecasts for multinational firms. Exploiting the staggered rollout of Google Translate's support to translate foreign languages into English, we find that U.S. analysts improve their forecast accuracy for firms with substantial business exposure in the corresponding foreign countries. The improvement is greater for analysts with limited language skills or brokerage resources to process foreign information. We further find that analysts raise more questions about firms' foreign exposure during conference calls and incorporate more foreign economic information into their forecasts after the rollout of Google Translate. Our findings highlight the complementary role of publicly accessible technology in supplementing analysts' skills and resources.

Keywords: Technology, Sell-Side Analysts, Machine Translation, Labor and Finance, Complementarity

1. Introduction

Globalization enables firms in the U.S. to access new markets, establish local production, and leverage business incentives from foreign countries. However, an expanded geographic footprint also complicates firms' information environment with more exposure to foreign factors, including local competition, macroeconomic conditions, and political risks (e.g., Li, Richardson, and Tuna 2014; Shroff, Verdi, and Yu 2014; Huang 2015; Lin, Mihov, Sanz, and Stoyanova 2019). As a result, information from foreign sources, such as foreign firms' corporate disclosures, local media news, and government policies, can become relevant for assessing and projecting multinational firms' performance. Such information is often available in foreign languages, creating substantial language barriers for capital market participants in the U.S. (e.g., Brochet, Naranjo, and Yu 2016; Lundholm, Rahman, and Rogo 2018). Several studies demonstrate that firms' global exposure increases the complexity of information processing activities for investors and analysts (Thomas 1999; Callen, Hope, and Segal 2005; Duru and Reeb 2002).

Traditionally, certain analysts can leverage their language skills or collaborate with international colleagues within the global network of their brokerage firm to keep track of relevant foreign information that's not easily available in English (Groysberg and Healy 2013). With the advent of machine translation technology as pioneered by Google Translate, an instant and reliable translation service has been made publicly available to all. The technology may help reduce analysts' costs of processing foreign information, especially for those with limited language skills or brokerage resources. In this study, we investigate the effect of machine translation technology on sell-side analysts' forecasts for multinational firms in the U.S. By studying translation technology, we also aim to deepen our understanding toward the role of technology in the financial research industry.

Economic theory suggests that technology may have both displacement and productivity effects on human labor, depending on the nature of the technology and the type of the labor (e.g., Acemoglu and Restrepo 2020). Unlike the innovations of financial social media (e.g., Seeking Alpha) and financial technology (FinTech) platforms (e.g., robo-analysts) studied in prior literature (e.g., Chen, De, Hu, and Hwang 2014; Coleman, Merkley, and Pacelli 2022), machine translation technology, by itself, does not produce financial information that directly competes with human analysts. Instead, translation technology can assist analysts with their research on multinational firms by allowing them to expand their scope of information search and to process relevant foreign information in a timely and cost-effective way. Anecdotal evidence demonstrates that financial professionals utilize translation tools to process information in foreign languages (Johnson 2016; Sultana 2021). Nonetheless, the value of translation technology for an analyst is likely a function of both the technology's effectiveness and the analyst's pre-existing skills and resources.

We choose the rollout of Google Translate as our primary setting to study machine translation technology for two main reasons. To begin with, Google was the first to launch a free, machine-learning-based translation tool with an initial support of nine foreign languages, achieving significant improvements in translation performance over previous linguistic rule-based systems (Giles 2006). Google Translate has since become a widely used translation tool with more than 500 million users per day in 2016 (Turovsky 2016). Google's embrace of machine learning also brought enormous progress to the translation industry, inciting the development of similar translation tools by other industry players (Kakaes 2011). Second, following the initial launch, Google Translate has continued to provide support to more than 100 languages in a *staggered* fashion and such support is generally more comprehensive and timelier than other competitors.

Therefore, we rely on the staggered rollout of Google Translate to capture the availability of advanced translation tools to analysts for translation of specific foreign languages. This variation enables us to separate out the effect of translation technology from general technological or market trends.[1]

We collect the initial launch date for each language covered by Google Translate using its historical web pages archived on the Wayback Machine supplemented with Google's official announcements. We capture U.S. firms' foreign exposure using their granular disclosure of subsidiaries operated across countries. We adopt a generalized difference-in-differences (DID) design to study the effect of Google Translate on the accuracy of analysts' annual earnings forecasts. We define a firm-year as treated if the firm operates at least 10% of its subsidiaries in countries where their major languages are supported by Google Translate in that year.[2]

We find that the rollout of Google Translate leads to a significant reduction in the errors of analysts' consensus forecasts, with the treatment firms experiencing a decrease of 6.5% in forecast errors relative to the sample standard deviation after controlling for a set of common firm characteristics as well as firm and year fixed effects. Our findings are robust to alternative definitions of treatment firms as well as several additional DID estimators (de Chaisemartin and d'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2021; Callaway and Sant'Anna 2021; Baker, Larcker, and Wang 2022). We also confirm that our findings are not only driven by several major languages that received the earliest support from Google Translate. Importantly, we estimate the

---

[1] We provide a more detailed discussion of our setting in Section 2. We confirm that Google Translate leads its primary competitor, Microsoft Translator, in both the scope and timeliness of language coverage. We also explore factors that can explain Google's language rollout decisions to strengthen our identifying assumption that Google's decisions are plausibly exogenous to an individual firm's information environment. We further describe some other translation supports available to analysts and explain how analysts can benefit from using free machine translation tools based on our discussions with industry practitioners as well as anecdotal evidence.
[2] Our control group thus includes the same group of firms in years before the rollout of Google Translate as well as firms that operate less than 10% of their subsidiaries in countries with their major languages ever supported by Google Translate.

dynamic effect of Google Translate on forecast errors and find no significant pre-trends. Overall, we provide robust evidence that the advancement in machine translation technology enhances analysts' forecasting performance for multinational firms.

We further examine forecast errors at the analyst level to explore the heterogeneous effects of machine translation technology across analysts with different innate abilities and external resources. First, we focus on an analyst's potential language skills. Exploiting the feature that U.S. analysts possess diverse cultural backgrounds (Merkley, Michaely, and Pacelli 2020), we rely on an analyst's ethnicity to infer a language that she is more likely to understand than other analysts of different ethnicities. We find that the effect of Google Translate in reducing forecast errors is more pronounced for analysts who are *less* likely to understand relevant foreign languages. Second, we examine an analyst's brokerage resources, with a specific focus on the foreign expertise of her colleagues within the same brokerage firm (e.g., Do and Zhang 2020; Hope and Su 2021; Huang, Lin, and Zang 2022). We identify an analyst as having fewer foreign coverage resources if her brokerage firm employs no analysts covering firms from relevant foreign countries, and find stronger effects of Google Translate for these analysts. Overall, we find greater benefit of machine translation technology for analysts who are subject to higher costs of processing foreign information, supporting the complementary role of translation technology in supplementing analysts' skills and resources.

Next, we conduct additional analyses to explore how translation technology may assist analysts' information-related activity. Because we cannot directly observe the specific pieces of information that analysts process with the help of translation technology, we employ related empirical proxies to provide suggestive evidence. First, we study analysts' questions during conference calls to gauge the set of information that analysts possess (Mayew, Sharp, Venkatachalam 2013; Cen, Han, and Harford 2022). We focus on questions that mention at least one of the foreign countries where their major languages are ever supported by Google Translate in our sample period (hereafter Supported countries). We find that the rollout of Google Translate leads to a larger proportion of these questions being raised by analysts for the treatment firms relative to the control firms. Our textual analysis further reveals that the increase is mainly driven by questions regarding a firm's related foreign entities (e.g., suppliers and business partners) and macro-level factors (e.g., currency risk, exports, tariffs, and regional market conditions). Our findings are consistent with the idea that translation technology facilitates analysts' access to certain foreign information, allowing them to raise related questions during conference calls.

Second, we study analysts' incorporation of foreign information, specifically firms' exposure to foreign economic conditions, into their forecasts. Following Li et al. (2014), we aggregate a firm's exposure to country-level economic factors using weighted-average GDP growth, separately for Supported and Non-Supported countries. We find that the rollout of Google Translate enhances analysts' incorporation of economic information from Supported countries, but not information from Non-Supported countries.

Finally, we investigate the market implications of analysts' improved forecast accuracy following the rollout of Google Translate. It is possible that other market participants can also utilize translation technology to incorporate more foreign information into their decisions. As a result, more accurate forecasts from analysts may not convey any incremental information to the market.[3] To explore this possibility, we examine the market reaction to analysts' forecast revisions.

[3] We also explore the effects of Google Translate on a firm's other information sources, such as the quantity of 8-K filings, the length of 10-K filings, the precision of management earnings guidance, and the quantity of media coverage. We do not find strong evidence of any significant changes. A possible explanation is that managers and media outlets are less subject to cross-border information processing frictions than analysts are. Managers, for instance, can access a rich set of foreign information through internal accounting systems or local agents (e.g., Liu and Lu 2023). Media outlets are also known to operate local branches in foreign countries, staffed with journalists who understand local languages.

We find stronger market reactions to analysts' forecast revisions for the treatment firms following the rollout of Google Translate, suggesting that translation technology enhances the informativeness of analysts' forecasts. We further document lower dispersion of analysts' forecasts and higher stock liquidity for the treatment firms after the rollout of Google Translate, corroborating the positive effect of translation technology.

Our study contributes to a growing literature on the impacts of technology in the investment research industry. Prior work documents the competition effects of financial social media and FinTech platforms over sell-side analysts' outputs (e.g., Grennan and Michaely 2021; Jame, Markov, and Wolfe 2022). In contrast, we focus on a non-financial type of technology—machine translation technology—and examine how such technology aids analysts in their forecasting tasks. Our evidence suggests that machine translation technology enhances analysts' forecast accuracy for multinational firms, and such technology may especially empower analysts who lack certain language skills or brokerage resources. Our findings are consistent with recent experimental and field studies on the complementary role of publicly accessible technologies in enhancing productivity and closing performance gaps among professional workers (Noy and Zhang 2023; Brynjolfsson, Li, and Raymond 2023; Peng, Kalliamvakou, Cihon, and Demierer 2023).

Relatedly, concurrent studies (e.g., Cao, Jiang, Wang, Yang 2022; Chi, Hwang, and Zheng 2022; Grennan and Michaely 2020) explore the potential benefits of artificial intelligence and big data to analyst research. We complement their work by focusing on free machine translation tools that are publicly available to all analysts but not just a selective set of analysts. Moreover, the staggered rollout of translation support for different foreign languages over time facilitates robust research designs to provide plausibly causal evidence of the technology's benefits.

Our paper also adds to research on the role of information processing costs for analysts (O'Brien and Tan 2015; Jennings, Lee, and Matsumoto 2017; Chen, Ma, Martin, and Michaely 2022). We exploit a novel setting to study the effects of technology on the processing costs of foreign information, which is increasingly relevant to capital market participants amid global economic integration and geopolitical risks. A related stream of research examines Google's withdrawal from the Chinese search engine market and documents its detrimental impacts on foreign information flows within China's capital markets (Xu, Xuan, and Zheng 2021; Kong, Lin, Wei, and Zhang 2022; Wang, Yu, and Zhang 2022).

Finally, but as importantly, our work relates to the broad literature on the information and investment frictions associated with language barriers in capital markets (e.g., Grinblatt and Keloharju 2001; Chan, Covrig, and Ng 2005; Brochet et al. 2016; Du, Yu, and Yu 2017; Lundholm et al. 2018; Zhang 2022; Lang, Stice-Lawrence, Wong, and Wong 2023;). Prior research primarily focuses on the existence and effects of inherent language differences across countries or among market participants from different origins. Our findings extend this literature by highlighting the potential of machine translation technology in addressing these language-related frictions.

2. Institutional Background and Setting

2.1. Analysts' Processing of Foreign Information

Financial analysts synthesize a wide range of information sources into their forecasts, such as corporate disclosures (Berger and Hann 2003; Hope 2003; Gibbons, Iliev, and Kalodimos 2021), media news (Bradshaw, Lock, Wang, and Zhou 2021), industry and macroeconomics factors (Hutton, Lee, and Shu 2012; Hugon, Kumar, and Lin 2016), as well as direct conversations with managers (Brown, Call, Clement, and Sharp 2015). The efficiency of these information processing activities varies with analysts' innate characteristics (e.g., experience) and external factors (e.g., brokerage firms' resources) (e.g., Clement 1999; De Franco and Zhou 2009; Cohen, Frazzini, and Malloy 2010; Bradley, Gokkaya, and Liu 2017).[4] Direct costs of information acquisition, such as geographic proximity to firms (O'Brien and Tan 2015; Jennings et al. 2017) and travel costs (Chen et al. 2022), can also impact analysts' information acquisition and their research performance.

The global footprint of multinational firms greatly expands the set of information that's relevant for analysts' forecasting tasks. As a result, analysts' costs of processing the above information sources are substantially higher given some of them may only be available in foreign languages (Duru and Reeb 2002).[5] In response to the challenge of processing foreign information, leading brokerage firms employ analysts with foreign language skills or encourage internal collaboration across geographic regions. For instance, the research department at the Bank of America started to promote globalized industry research that involved analysts from different regions working together since the 2000s (Groysberg and Healy 2013, p. 121). Some analysts may even incur significant costs to travel to foreign countries to visit a firm's related foreign entities (e.g., competitors) and to understand local market dynamics (Groysberg and Healy 2013, p. 148). However, such practices are unlikely to be universal across brokerage firms due to their high-cost nature and the variation in brokers' resources (Clement 1999).

We further corroborate these institutional details by reviewing a random set of analyst reports on multinational firms and identifying examples of foreign information and their corresponding sources, which we present in Section B of the Online Appendix. Our review suggests that analysts discuss a variety of foreign information pertaining to multinational firms in

[4] We follow Blankspoor, deHaan, and Marinovic (2020) to use "information processing" to refer to a set of analysts' information activities, which can include awareness, acquisition, and integration. Because we are interested in the overall costs of information processing, we do not intend to separate out one activity from another.

[5] Admittedly, certain foreign information can be available in English but is not necessarily complete. For example, U.S. media outlets predominantly report foreign events that are of substantial relevance to their domestic audience (Wu 2000), which are unlikely to include events only related to a firm's foreign operations.

their research reports, such as foreign policy implications, local industry and competition dynamics, as well as specific events related to firms' foreign operations. Analysts' sources of information can include media news, research reports, corporate disclosure, discussions with colleagues, and even field trips to foreign countries.

2.2. The Setting

Google Translate was officially launched as the first free, statistically based translation tool in April of 2006 with the initial support of translation between English and nine foreign languages.[6] Different from other translation tools that were still adopting the older rule-based systems calibrated by linguists, Google Translate applied statistical learning techniques to train a machine translation model using large amounts of texts without the need of human interventions. The automatic algorithm helped identify a broader and more nuanced set of matching pairs between two languages than linguists did, enhancing both the accuracy and efficiency of machine translation tasks. As a pioneering shift in approach to automatic translation, Google Translate began to outperform others in industry translation evaluations (Giles 2006). Google's push also catalyzed technological advancements in the machine translation industry, leading several competitors to introduce similar translation tools (Kakaes 2011).

As of June 2020, Google Translate has expanded its support to 108 foreign languages, many of which couldn't be translated by any available systems before. Appendix B lists these languages by the year when they were initially supported. Besides the comprehensive language coverage, Google Translate supports the translation of multiple forms of texts and media, including websites, documents, and speech. Its web interface allows users to type in texts, upload documents, or enter a website link for instant translation. Google Translate also offers a browser extension,

---

[6] These nine foreign languages include Arabic, Chinese (simplified), French, German, Italian, Japanese, Korean, Portuguese, and Spanish.

which enables users to access foreign websites in their preferred language directly from their browsers.

Despite there being alternative machine translation tools (e.g., Microsoft Translator), Google Translate has become one of the most, if not the most, popular and widely used translation products, with more than 500 million users per day according to their most recent report in 2016 (Turovsky 2016). Google Translate also provides the most extensive and timely coverage of foreign languages among industry rivals.[7] In terms of translation quality, Google Translate led other 30 products according to a recent independent evaluation conducted by a translation solutions provider (Intento 2022).[8] Because of these reasons, we rely on Google Translate's initial support for a given foreign language to proxy for the availability of advanced machine translation tools for that specific language.[9]

The public availability of free, advanced translation tools may enable analysts to expand their information search and to process foreign information in a timely and cost-effective way. Consistent with this view, anecdotal evidence suggests that financial professionals rely on translation tools to read financial reports and corporate announcements from foreign companies, as well as media posts or blogs in foreign languages. These anecdotes are presented in Section C

[7] To put this into perspective, we contrast Google Translate with its primary competitor, Microsoft Translator, in terms of their coverage breadth and timeliness of the initial support across languages in Figures OA1 and OA2 of the Online Appendix, respectively. As of June 2020, Microsoft Translator supported a total of 68 foreign languages, which are fewer than two-thirds of the languages supported by Google Translate. For the set of languages supported by both, Microsoft Translator consistently lagged behind Google Translate in its initial rollout, with a delay ranging from a few months to several years.
[8] In July 2022, Intento evaluated the performance of 31 machine translation engines across 11 language pairs (between English and 11 foreign languages) in 9 content domains including financial content. The translation output from these engines were compared against a reference human translation to identify top performing translation providers. Google Translate received the "best" score in the highest number of language pairs across content domains.
[9] Relatedly, several studies investigate the effects of Google's withdrawal from the Chinese market in 2010 on the flows of foreign information within China's capital markets (Xu et al. 2021; Kong et al. 2022; Wang et al. 2022). Although Google stopped its search engine and other products in China starting in 2010, it continued to operate Google Translate in the country until 2022 (Strumpf 2022). Therefore, the effects documented in those studies are unlikely to be driven by a change in the availability of Google Translate in China.

of the Online Appendix. Our discussions with industry practitioners also offer consistent insights. Analysts who cover firms with significant foreign exposure tend to keep track of local news by visiting major foreign media outlets using translation tools. As another illustration, when analysts delve into research on a multinational firm, they often go through corporate disclosures of foreign industry competitors or partners along the supply chain, as well as policy documents issued by local governments. Analysts can quickly translate such documents with the help of translation tools.

We recognize that analysts may also rely on human translators to translate foreign information. It is possible that analysts find machine translation tools of limited use given that machine translation can still not match human translators in translation quality (e.g., Wu et al. 2016). However, the cost of human translation is significant in practice and the turnaround time can range from a few hours to a few days.[10] Although financial institutions typically maintain a small team of professional translators, our discussion with one translator from a leading investment bank suggests that these internal translators are primarily tasked with translating details of investment products and confidential documents. Only more recently, certain investment banks have established their own internal machine translation tools. For example, BNP Paribas developed an internal translation engine in 2022 (Stasimitoti 2022). However, this engine only supported translations in 15 foreign languages and was costly to develop according to their researchers.

[10] The average translation rate ranges from $180 to $250 per 1,000 words translated in 2020 according to the General Services Administration (Blackwood 2020). Although Google Translate cannot match professional human translators, especially when used for formal communications (Patil and Davies 2014; Chen, Acosta, and Barry 2016), it can handle a large volume of translation tasks in a fast manner while preserving a fair amount of the underlying information (De Vries, Schoonvelde, and Schumacher 2018).

In summary, to the extent that machine translation tools such as Google Translate can fairly capture the underlying content that's relevant for analysts' research, the benefits of using such tools (i.e., flexibility, timeliness, and cost-effectiveness) likely outweigh its compromise in translation quality relative to human translation. Therefore, the advancement in machine translation technology likely lowers analysts' costs of processing foreign information, thereby enhancing their forecasting performance for firms with substantial foreign exposure.

3. Data and Research Design

3.1. Data and Sample Selection

We collect analysts' annual earnings forecasts (FPI=1) for U.S. firms from the I/B/E/S detail history database. For each firm-year, we retrieve all forecasts with a horizon between one and 12 months from each analyst covering the firm. If an analyst issues more than one forecast for a firm-year, we retain only the last forecast (Bradley et al. 2017). Following prior studies (e.g., Desai, Foley, and Hines 2008; Dyreng and Lindsey 2009; Lin et al. 2019), we capture a firm's foreign exposure using its subsidiaries disclosed in Exhibit 21 of Form 10-K and obtain the data from WRDS.[11] We collect the data starting from 1997 when the electronic filings on EDGAR were fully phased in (Asthana, Balsam, and Sankaraguruswamy 2004). Because we rely on one-year-lagged subsidiary data to identify a firm's treatment status, our main sample starts from 1998 and ends in 2020. We further obtain firms' accounting information from Compustat, stock prices from CRSP, institutional holdings from Thomson 13F filings, detailed transcripts of earnings conference calls from Capital IQ, and country-level economic data from World Bank.

[11] Regulation S-K requires firms to disclose both the name and jurisdiction of incorporation of all significant subsidiaries, representing the most granular public disclosure of a firm's geographic footprint (Dyreng, Hoopes, Langetieg, and Wilde 2020). According to 17 CFR 210.1-02, a subsidiary is defined as an affiliate controlled by the firm. Dyreng et al. (2020) validate the quality of Exhibit 21 disclosure of firms' foreign subsidiaries and find low incidence of nondisclosure.

We rely on two primary data sources to identify the initial launch date for each language covered by Google Translate. Our first data source is from the Wayback Machine, a digital archive of the World Wide Web established by a non-profit organization, the Internet Archive.[12] We obtain a series of Google Translate's historical web pages from the Wayback Machine, and by comparing these web pages over time, we identify the initial date at which Google Translate began supporting a specific foreign language. Figure OA3 of the Online Appendix provides two examples of the historical web pages to illustrates the process. The Wayback Machine generally provides at least one snapshot of Google Translate's historical page each day since July 2010, but the coverage is relatively sparse before that. We hence supplement and validate the identified dates using Google's own official announcements and weekly updates, which are our second source of data.[13] We are able to collect a set of announcement dates and a list of new languages discussed in each announcement. If a language's initial launch date is not clearly specified in the announcement, we verify that the launch date identified using historical web pages is within a short period before the announcement date. We further conduct an extensive search of other online sources to cross-check our collected dates. Appendix B provides the list of languages supported by Google Translate by their initial launch year.

Next, we map each country with its major language using the World Factbook compiled by the Central Intelligence Agency (CIA). The World Factbook covers basic information for countries in the world and is available for public use online.[14] The Factbook lists top languages used in a country based on their relative size of users over the country's total population. Although multiple languages can be used in a given country, we rely on the most commonly-used language

[12] https://web.archive.org.
[13] https://ai.googleblog.com.
[14] https://www.cia.gov/the-world-factbook/, accessed on October 10, 2022.

to capture the major medium through which information in this country is communicated. The most commonly-used language also tends to be the official language or one of the official languages designated by the country.[15] Figure 1 labels each country based on the first year since which Google Translate began supporting its most commonly-used language.

Panel A of Table 1 presents the construction of our main sample at the firm-year level. We start with firm-year observations from Compustat with available analysts' annual earnings forecasts from I/B/E/S. Because we are interested in U.S. firms, we exclude firms that are either not headquartered or incorporated in the U.S. We further exclude observations with missing lagged subsidiary data and observations with missing control variables. Last, we drop firms with only one observation in our sample because we include firm fixed effects in our firm-year regressions. Our final sample consists of 49,450 firm-year observations with 5,520 unique firms.

For our analyses of individual forecast errors at the analyst level, we obtain a sample of 524,078 analyst-firm-year observations following the same procedures above. For each analyst in our sample, we further collect their detailed demographic information. We start with an analyst's last name, first initial, and brokerage firm from the I/B/E/S recommendation file and construct their full career history using I/B/E/S/ broker information. With this information, we search Capital IQ to identify the analyst and obtain her full name. For analysts that can't be identified from Capital IQ, we perform additional manual searches using other platforms, such as LinkedIn, BrokerCheck, and websites of brokerage firms. We are able to identify the full names of 9,060 analysts, representing 79% of the unique analysts in our sample. We then determine the language that an

[15] It is possible that a country may designate more than one language as official. For example, India has designated both Hindi and English as their official languages with Hindi spoken by 57.1 percent of the country's population and English spoken by only 10.6 percent (Kawoosa 2018). Although English is used for national and political communication, a large number of news articles and local government policies are still written and communicated in Hindi and other local languages. We note that mapping a country with all of its languages with proper weights is challenging given the data limitation. We provide an alternative research design that incorporates the possibility of multiple official or business languages (including English) used in a country, which we elaborate in Section 4.2.

analyst is more likely to be proficient in using the ethnicity inferred from her full name. Following prior studies (e.g., Munz, Jung, and Alter 2020; Krishnan, Singer, and Zhang 2023), we rely on a machine-learning-based software trained on historical census data (NamSor) to identify each analyst's most likely ethnicity based on her full name.[16] We then assign the most commonly-used language associated with the ethnicity as the language that the analyst is likely to be proficient in. For instance, Fadel Gheit is an analyst in our sample and his predicted ethnicity from NamSor is Egyptian. As a result, we assign Arabic as the language that Fadel is likely to be proficient in.[17]

3.2. Research Design and Summary Statistics

3.2.1. Firm-Year Level Analyses

For our main identification strategy, we leverage the staggered rollout of Google Translate's support to foreign languages over time. We rely on a firm's foreign subsidiaries to capture its geographic exposure across countries. If a firm operates a subsidiary in a given foreign country, the information from this foreign country (e.g., local news, disclosure from industry peers or suppliers, macroeconomic information, and government policies) is likely relevant for assessing the subsidiary's sales or production, which are meaningful to earnings forecasts at the firm level. With Google Translate's support to the corresponding foreign language, the cost of processing foreign information from that country is likely reduced compared with the pre-period without the support.

---

[16] Prior research has validated the accuracy of NamSor's classification. Bursztyn, Chaney, Hassan, and Rao (2022) use a random sample of 250,000 individuals from North Carolina Voter Registration data and match NamSor's ethnicity predictions with the self-reported ethnicity information on the registrations. They find that NamSor's error rate is smaller than 1 percent. Krishnan et al. (2023) use NamSor to identify ethnic minorities among audit partners. The authors randomly select 150 auditors from their sample and find less than 5% of Type I and Type II errors in NamSor's classification.

[17] The 10 most frequent ethnicities in our sample are British, Irish, Jewish, Chinese, German, Indian, Italian, French, Hispanic, and Austrian. Their corresponding languages are English, English, Hebrew, Chinese, German, Hindi, Italian, French, Spanish, and German.

One key assumption underlying our identification strategy is that the rollout timing of specific languages by Google Translate should be arguably independent of factors that might otherwise affect analysts' forecast errors. In Table OA1 of the Online Appendix, we employ a hazard model to investigate Google's language rollout decisions. We find that these decisions are largely explained by factors pertaining to the language itself or its user base, such as the language complexity and the size of its potential users. We do not find significant evidence that the language rollout decisions correlate directly with the number of U.S. subsidiaries operating in the respective countries.

For a given firm-year, we aggregate the treatment effect of Google Translate by counting the total number of the firm's foreign subsidiaries domiciled in countries where their major languages are supported by Google Translate in that year, scaled by the firm's total number of subsidiaries. We use a firm's pre-existing subsidiaries from the prior year to alleviate the concern of Google Translate' potential real effects on a firm's foreign operations (Brynjolfsson, Hui, and Liu 2019).[18] We illustrate our calculation using the following equation:

$$Translate\ Percentage_{i,t} = \frac{\sum_{s=1}^{N_{i,t-1}} Support_{i,s,t}}{N_{i,t-1}} \times 100\%$$

where *i* index firms, *t* indexes years, and *s* index subsidiaries. For pre-existing subsidiary *s* of firm *i* in year *t-1*, $Support_{i,s,t}$ is an indicator that equals one if subsidiary *s* is domiciled in a foreign country where its major language is supported by Google Translate in year *t* (i.e., fiscal year *t* starts after the initial launch date of Google Translate's support to that language), zero otherwise.

[18] We explore the persistence of a firm's total number of foreign subsidiaries over time and tabulate the results in Table OA2 of the Online Appendix. We find that 94% of the variation in a firm's total number of foreign subsidiaries can be explained by its value from the last period. Importantly, we do not find evidence that the rollout of Google Translate affects a firm's number of foreign subsidiaries in the next period, which suggests that any documented effect of Google Translate is unlikely to be driven by the real effect of Google Translate on a firm's foreign operations.

We adopt a generalized difference-in-differences (DID) design to estimate the effect of Google Translate on analysts' forecast errors at the firm-year level using the following model:

$$Error_{i,t} = \beta_0 + \beta_1\ Translate_{i,t} + \beta_2\ Controls_{i,t} + Firm\ FE + Year\ FE + \varepsilon_{i,t}, \qquad (1)$$

where *i* index firms and *t* indexes years. The dependent variable, *Error*, is the absolute difference between the consensus analyst forecast and the actual value, scaled by stock price. Our test variable, *Translate,* is an indicator that equals one if *Translate Percentage* is 10% or greater, zero otherwise.[19] Panel B of Table 1 tabulates the number of firms by the initial treatment year (i.e., the first year when *Translate Percentage* is at least 10%). Around 16% of firms are first treated in 2007 following Google's initial support of ten major foreign languages. As Google Translate continues to rollout new languages, more firms are defined as treated with 4% in 2008 and 3% in 2009, and the percentage stays around 1-2% in the following years. The pattern supports the notion that firms are treated in a staggered fashion. There remain 60% of firms that are never treated by the end of our sample period, providing ample control firms for estimating the treatment effect that's less subject to the "bad comparisons" problem (Baker et al. 2022). We also adopt several alternative DID estimation methods as suggested by recent econometric literature, which we elaborate in Section 4.2.

We control for a number of firm-year level characteristics that may affect firms' information environment and analysts' forecast errors. These controls include market capitalization (*Size*), book-to-market ratio (*BTM*), an indicator of loss (*Loss*), return on assets (*ROA*), earnings volatility (*EarnVol*), return volatility (*RetVol*), leverage ratio (*Leverage*), R&D expenses scaled by total assets (*R&D*), the logarithm of one plus the total number of foreign

[19] We choose the indicator as our main test variable to facilitate the DID design and to ease the interpretation of our results. In Section 4.2, we conduct a number of sensitivity tests to ensure our findings are robust to other specifications, such as using alternative cutoffs to define the indicator or using the continuous measure, *Translate Percentage*.

subsidiaries (*ForeignSub*), the logarithm of one plus the total number of manager's earnings guidance issued in the year (*Guidance Count*), and percentage of shares owned by institutions (*InstOwn*). We provide detailed variable definitions and data sources in Appendix A. We also include firm fixed effects to account for firm-specific and time-invariant characteristics, and year fixed effects to absorb time-varying economy-wide trends. We cluster standard errors at the firm level. If Google Translate reduces the costs processing foreign information for analysts and facilitates their information production activities, we expect β_1 to load negatively.

Table 2 presents the summary statistics of our firm-year sample. The mean value of *Translate Percentage* is 13.40, indicating that, on average, a sample firm operates 13.40% of its foreign subsidiaries in countries where the major language is supported by Google Translate. The mean of our treatment variable, *Translate*, is 0.30, suggesting that 30% of our observations have a value of *Translate Percentage* that is at least 10%. The mean (median) value of consensus analyst forecast *Error* is 2.14% (0.3%).

3.2.2. Analyst-Firm-Year Level Analyses

To explore the heterogenous effects of Google Translate across analysts, we conduct analyses at the analyst-firm-year level using the following model:

$$\begin{aligned} Error_{i,j,t} = {} & \beta_0 + \beta_1\, Translate_{i,j,t} + \beta_2\, Firm\ Level\ Controls_{i,t} + \beta_3\, Analyst\ Level\ Controls_{j,t} \\ & + Fixed\ Effects + \varepsilon_{i,j,t}, \end{aligned} \tag{2}$$

where *i* indexes firms, *j* indexes analysts, and t indexes years. The dependent variable is individual analyst forecast error, calculated as the absolute difference between the individual forecast and the actual value, scaled by stock price. The test variables include the firm-year treatment indicator (*Translate*) as defined in Equation (1) and two other sets of treatment indicators at the analyst-firm-year level. In addition to the firm-year level characteristics included in Equation (1), we further control for several analyst- or forecast-level features, such as brokerage size (*Brokerage Size*), general work experience in years *(General Experience)*, firm-specific experience in years (*Firm Experience)*, the number of firms covered in the portfolio (*Portfolio Size*), the number of unique industries covered in the portfolio (*Industry Count*), the total number of forecasts issued (*Effort*), and the number of days between the forecast issuance date and the earnings announcement date (*Horizon*). We adopt the logarithm transformation of these variables in our regressions. We use two sets of relatively strict fixed effects specifications for these analyses. In the first set, we include analyst fixed effects in addition to firm and year fixed effects to control for analyst-specific and time-invariant characteristics that can explain analysts' forecasting performance difference (e.g., innate ability). In the second set, we further strengthen the design by including analyst-firm and year fixed effects. The inclusion of analyst-firm fixed effects ensures that our identification comes from the change in forecast accuracy for the same analyst-firm pair. Standard errors are clustered at the firm and analyst levels in both specifications.

We define the two additional sets of treatment indicators at the analyst-firm-year level by decomposing $Translate\ Percentage_{i,t}$ into two components based on an analyst's potential foreign language proficiency and brokerage resources. First, we separately aggregate the affected subsidiaries in countries where the major language is either likely or unlikely understood by the analyst. We then construct indicators for the two groups based on whether the affected subsidiaries account for at least 10% of the firm's total subsidiaries, and define *Translate (Proficient)* and *Translate (Non-Proficient)*, respectively. Second, we similarly construct another set of indicators based on whether the affected subsidiaries are domiciled in a country from which the analyst's brokerage firm covers at least one local firm and obtain *Translate (Covered)* and *Translate (Non-Covered)*, respectively. We consider the industry practice that resourceful brokers tend to involve analysts from various regions to conduct more globalized research (Groysberg and Healy 2013). Consistently, recent research shows that colleagues within the same brokerage firm can be a useful source of information for analysts (Do and Zhang 2020; Hope and Su 2021; Huang et al. 2022). If an analyst works in a brokerage firm with her colleagues covering firms from certain foreign countries, this analyst may be able to receive additional support from their colleagues to process foreign information, reducing the benefit of using Google Translate.

4. Empirical Results

4.1. Main Results

Table 3 presents the estimates of the effect of Google Translate on analysts' forecast errors at the firm-year level using Equation (1). In Column 1, we estimate the difference between the treatment and control groups without including any control variables or fixed effects. The coefficient on *Translate* is negative and statistically significant at the 1% level, suggesting that analysts' forecast errors are significantly lower for firms with at least 10% of subsidiaries domiciled in countries where the major language is supported by Google Translate. The results remain similar after we further include firm and year fixed effects in Column 2, and the full list of control variables in Column 3. We rely on the coefficient in Column 3 to gauge the economic significance of Google Translate's impact on analysts' forecast errors. The forecast errors of the treatment firms decrease by 0.48 relative to the control firms, representing 6.5% of its standard deviation (0.48/7.40), which is economically meaningful.[20] The findings in Table 3 support our prediction that machine translation technology improves analysts' forecasting performance for firms with substantial foreign exposure.

[20] We also compare the economic significance of our findings with prior studies. For example, Merkley et al. (2020) show in their Panel B of Table 3 that the effect of a unit increase in their measure of cultural diversity on analysts' forecast accuracy is 16% relative to their unconditional sample mean. In our setting, the impact of *Translate* on forecast error is 22% (0.48/2.14) relative to the sample mean, which is comparable to that in Merkley et al. (2020).

We test the parallel trends assumption by examining the dynamic effect on analysts' forecast errors in the periods before and after the initial treatment year for the treatment firms, relative to the control firms. Since we define the treatment indicator based on whether a firm's affected subsidiaries account for at least 10% of the total subsidiaries, we further control for the percentage of treated subsidiaries below the 10% threshold to remove their potential impacts on the pre-trend estimates. Figure 2 plots the coefficients and their 95% confidence intervals by event year. We observe no significant difference in analysts' forecast errors between the treatment and control firms in the pre period, supporting the parallel-trend assumption. We observe significant treatment effect starting from the first year when Google Translate affects at least 10% of the firm's subsidiaries. The effect continues to persist after four years.

4.2. Robustness

We present a battery of additional analyses in the Online Appendix to demonstrate that our main findings are robust and not sensitive to our specific choices of research design.

First, we show that our results are robust to using alternative definitions of the treatment variable (*Translate*) in Table OA3. Columns 1 and 2 employ indicators defined based on two alternative thresholds, 0% and 20%, respectively. Column 3 uses a continuous treatment variable based on the percentage of treated subsidiaries (*Log(1+Translate Percentage)*). Column 4 relies on a firm's constant set of subsidiaries in the first year of its appearance in our sample period to define the treatment indicator. We adopt this specification to further alleviate the concern that a change in the mix of a firm's foreign subsidiaries contaminates our identification. Column 5 uses an indicator defined based on a firm's reported sales across geographic segments.[21] In Column 6,

[21] Compared with the subsidiary data that we use to define our main treatment variable in Table 3, segment reporting often provides less granular information because firms typically aggregate their sales from multiple countries into one single geographic segment (e.g., Europe). If a firm reports some of their geographic segments in regions, we follow

we further consider the possibility that multiple official or business languages including English can be used in a given country in addition to the major foreign language that we identify. We first treat foreign countries with English as one of their official or business languages (e.g., India) as equivalent to English-speaking countries and exclude any subsidiaries located in those countries from the calculation of treated subsidiaries. For the remaining countries with multiple non-English official or business languages, we consider a subsidiary in the country as treated if *any* of the official or business languages are supported by Google Translate in a given year and define the treatment indicator accordingly.

Second, we evaluate the soundness of our staggered DID design in Table OA4 by decomposing our continuous treatment variable into two components: one driven by the first batch of nine languages supported by Google Translate and the other by the remaining supported languages. Panel A suggests that the first component has an average of 8.19% and the second component has an average of 5.22%. The difference is not surprising as the first batch covers nine important foreign languages that are spoken in many countries as shown in Figure 1. The regression results in Panel B show significant treatment effects for both components, reducing the concern that our results are only driven by the first batch of nine important languages.

Third, we consider the potential estimation bias under the generalized DID design with staggered treatment effects and show that our inferences are robust to using several alternative estimation methods as suggested by recent econometric literature. We present the results in Table OA5. In Column 1, we employ the method proposed by de Chaisemartin and d'Haultfoeuille (2020), which is suitable in our context where the treatment status possibly varies between on and off over time. In Columns 2 and 3, we use the imputation estimator introduced in Borusyak et al.

Li et al. (2014) to allocate the aggregated sales into specific countries in the segment, using each country' gross domestic product (GDP) as the weight.

(2021) and the Callaway and Sant'Anna (2021) estimator, respectively. In the last column, we test the robustness of our findings using the stacked-regression analysis suggested by Baker et al. (2022).

Last, we test whether our inferences are robust to using the staggered rollout of Microsoft Translator as an alternative proxy for the availability of advanced machine translation tools. As discussed in Section 2, we choose Google Translate as our main research setting as it is most widely-used in the industry, offers the most comprehensive coverage with high translation quality, and leads Microsoft Translator in rolling out almost every foreign languages. Nonetheless, we test the effect of Microsoft Translator as well as its incremental effect over Google Translate in Table OA6. Columns 1 and 2 of Panel B suggest that our main results are robust to using the staggered rollout of Microsoft Translator to define the treatment variables. More importantly, Columns 3 and 4 further reveal that the effect of Google Translate subsumes the effect of Microsoft Translator, supporting our choice of Google Translate as the main proxy for the availability of translation technology for analysts.

4.3. Analyst-Firm-Year Level Analyses

To further explore the heterogenous effects of Google Translate across analysts, we conduct analyses using forecast errors at the analyst-firm-year level. Table 4, Panel A presents the summary statistics of this sample. The mean (median) value of individual analysts' forecast errors is 1.24% (0.2%). We use the unlogged value of analyst- or forecast-level controls to describe our sample. A brokerage firm in our sample employs a median of 48 analysts. An analyst in our sample has a median working experience of 10 years, a firm-specific coverage experience of 2 years, a coverage portfolio of 16 firms and 3 distinct industries, and issues 4 forecasts for each firm-year.

The median value of forecast horizon is 100 days. These summary statistics are generally comparable to those in prior studies (e.g., Bradley et al. 2017).

Panel B of Table 4 presents the regression results using Equation (2). We include firm, year, and analyst fixed effects in the first three columns. In Column 1, the coefficient on *Translate* is negative and statistically significant at the 1% level, suggesting that the rollout of Google Translate results in lower individual forecast errors, consistent with our main findings using the firm-year sample in Table 3. The magnitude of the economic significance is also similar, representing a decrease of forecast errors by 0.19*,* which is 4.8% of the sample standard deviation (0.19/3.96).

In Column 2, we separate *Translate* into two indicators based on an analyst's potential foreign language proficiency. The coefficient on *Translate (Non-Proficient)* is negative and statistically significant, while the loading on *Translate (Proficient)* is statistically indistinguishable from zero. The difference between these two coefficients is also significant at the 5% level (*p*-value = 0.04). The results suggest that Google Translate mainly reduces the processing costs of foreign information for analysts who are less likely to understand the relevant foreign languages. This finding supports the idea that machine translation technology plays a complementary role with respect to the innate ability of human analysts.

In Column 3, we separate *Translate* into two indicators based on the availability of foreign coverage resources in analysts' brokerage firms. The coefficient on *Translate (Non-Covered)* is negative and statistically significant, while the coefficient is insignificant for *Translate (Covered)*. The difference between these two coefficients is also significant at the 1% level (*p*-value = 0.00). The results indicate that Google Translate is mainly helpful for analysts whose brokerage firms lack corresponding foreign coverage activity, supporting the complementary relation between machine translation technology and brokerage firms' resources.

In Columns 4 to 6, we report the results of the regressions using analyst-firm fixed effects and year fixed effects. Such a strict fixed-effect specification allows us to identify the treatment effect within a given analyst-firm pair, which is less subject to any changes in analysts' coverage decisions. Our findings remain similar using this specification. In Table OA7 of the Online Appendix, we present the analyst-level results using continuous treatment variables instead of indicators as in Table 4 and obtain similar inferences. Overall, our evidence suggests that the effects of translation technology are stronger for analysts who are subject to higher costs of processing foreign information due to the lack of language skills or brokerage resources.

5. Additional Analyses

We conduct additional analyses to explore how machine translation technology may assist analysts' information activity. Ideally, we need to examine the specific pieces of information that analysts process using machine translation tools, but such information is generally not available to researchers (Brown et al. 2015). As such, we focus on observable but related empirical proxies to provide suggestive evidence. Specifically, we examine analysts' questions during conference calls and their incorporation of foreign economic information into forecasts. We further test the incremental information content of analysts' forecasts using market outcomes and explore alternative information sources that may also explain analysts' improved forecast accuracy.

5.1. Analysts' Questions during Conference Calls

First, we examine the questions that analysts raise during earnings conference calls, which represent an important source to infer the information set that analysts possess (Mayew et al. 2013; Cen et al. 2022). As suggested by Cen et al. (2022), analysts are more likely to ask questions on certain issues for which they have gathered relevant information with their skills and expertise. If machine translation tools allow analysts to process more existing information related to firms' foreign operations, such information may allow analysts to ask more related questions during conference calls.[22]

We conduct textual analysis using detailed transcripts of a firm's annual earnings conference calls to identify analysts' questions regarding the firm's foreign exposure during the Q&A session. Specifically, we define *Foreign Question* as the number of questions that mention at least one of the foreign countries where the major language is ever supported by Google Translate (Supported countries), scaled by the total number of all questions. The summary statistics from Table 2 show that the average value of *Foreign Question* is 2.40%, suggesting that around 2 percent of analysts' questions in conference calls mention at least one of Supported countries. Column 1 of Table 5 shows that analysts raise a higher proportion of questions related to a firm's foreign exposure in Supported countries after the rollout of Google Translate. In terms of economic significance, *Translate* increasing from 0 to 1 is associated with an 8.3% increase in foreign questions relative to the sample standard deviation (0.44/5.27).

To better understand the types of foreign information that analysts may possess, we employ the Latent Dirichlet Allocation approach to analyze the thematic content of analysts' foreign questions (Huang, Lehavy, Zang, and Zheng 2018). We use the full sample of foreign questions collected from conference calls and classify these questions into 10 topics.[23] Appendix C provides our labels of these topics and the most frequent words under each topic. We find that these topics

[22] It's also possible that these analysts may raise fewer questions related to firms' foreign operations during conference calls if the additional information that analysts possess has helped resolve their questions already.
[23] Our number of topics is smaller than prior work (e.g., Huang et al. 2018) because our analysis is conditional on analysts' foreign questions, not the full sample of all analysts' questions. We also use other number of topics (e.g., 20, 30, 50), but do not observe a meaningful improvement in performance. We remove stop words as well as the names of specific countries or regions, before running the LDA analyses.

appear to cluster in three broad categories. The first category includes topics that are focused on firm-specific information, such as a firm's operational performance, financial conditions, and business opportunities. The second category includes topics on a firm's related entities including local suppliers, business partners, industry rivals, banks etc. The third category includes macro-level topics, such as currency risk, the impact of exports and tariffs, and regional market conditions.

Columns 2 to 4 of Table 5 present the effects of Google Translate on the quantity of foreign questions by these three categories. We find that the increase in foreign questions is mainly driven by questions from the third category, which suggests that machine translation technology may especially facilitate analysts' processing of foreign economic information. We also observe a modest increase in analysts' questions from the second category, consistent with analysts possessing more information about a firm's related foreign entities with the help of machine translation technology.

5.2. Incorporation of Foreign Economic Information

Next, we test whether analysts integrate more foreign information into their forecasts after the rollout of Google Translate. Prior research suggests that macroeconomic factors have a material impact on a firm's financial performance (Li et al. 2014) and represent a key source of information that analysts integrate into their forecasts (Hutton et al. 2012). Our analyses on conference calls provide consistent evidence that translation technology may particularly facilitate analysts' processing of foreign economic information. If such information helps analysts better understand a firm's specific exposure to foreign economic conditions, we expect analysts' forecasts to incorporate more information on foreign economic conditions.

In our empirical analyses, we rely on foreign GDP information as our proxy for foreign economic conditions.[24] Our empirical design is motivated by Li et al. (2014) to examine the relation between a firm's forecast-implied ROA and the weighted-average GDP growth across countries where the firm operates.[25] In our study, we separately aggregate real GDP growth for Supported and Non-Supported countries, where Supported countries, as defined previously, include countries where the major language is ever supported by Google Translate, and Non-Supported countries include those with languages that are not supported by Google Translate as well as English-speaking countries. Accordingly, we construct *GDP Growth (Supported)* and *GDP Growth (Non-Supported)*. We retain firm-year observations with sales from both Supported and Non-Supported countries in our regressions. We further control for a set of firm fundamentals in the prior year that may explain the actual ROA or analysts' forecasts. We also include firm and year fixed effects. Table 6 presents the results. As a benchmark, we estimate the relation between the actual *ROA* and GDP growth in Column 1. The results show that macroeconomic factors in both Supported- and Non-Supported countries are associated with firms' financial performance, consistent with the evidence from Li et al. (2014).

Next, we examine whether analysts incorporate firms' exposure to macroeconomic factors into their forecasts. In Column 2, we replace the dependent variable with *Forecast-Implied ROA*,

[24] We do not posit that analysts need to rely on Google Translate to process foreign GDP information as such information is mostly quantitative in nature. Instead, we argue that Google Translate can facilitate analysts' assessment of a firm's specific exposure to a set of foreign economic information, such as local market conditions, government policies, and economic environments, which are challenging to capture using empirical proxies. To the extent that GDP growth captures part of this information, we should observe a stronger relation between analysts' forecasts and GDP growth after the support of Google Translate.

[25] Li et al. (2014) show that macroeconomic conditions can help predict firms' ROA. We follow Li et al. (2014) to aggregate the GDP growth using a firm's geographic segment sales. We calculate the weight for each country based on the segment sales information and then take the weighted-average GDP growth across countries to construct the GDP growth measure at the firm-year level. If a firm only discloses total sales at the regional level (e.g., North America) instead of the country level, we use the specific GDP of member countries in the region as the weight to allocate aggregated sales to each country. Our inferences are robust to aggregating the GDP growth using a firm's number of subsidiaries across countries as the weight. The results are tabulated in Table OA8 of the Online Appendix.

calculated as the consensus analyst forecast of earnings per share multiplied by the number of common shares outstanding divided by total assets. The coefficients for the two *GDP Growth* variables remain positive and statistically significant, suggesting that analysts on average incorporate a significant portion of firms' exposure to macroeconomic factors into their forecasts. In Column 3, we include the interaction terms between *Translate* and the two *GDP Growth* variables. The coefficient on *GDP Growth (Supported)* × *Translate* is positive and significant, while the coefficient on *GDP Growth (Non-Supported)* × *Translate* is insignificant. The results suggest that Google Translate facilitates analysts' integration of economic information from Supported countries, but not Non-Supported countries. These results shed light on the potential types of information activity through which translation technology enhances the performance of analysts.

5.3. Incremental Information Content of Analysts' Forecasts

We further examine whether analysts' improved forecasting performance conveys incremental information to the market. One may argue that other market participants (e.g., journalists, managers, and investors) can also make use of machine translation tools to process more foreign information at a lower cost and analysts may simply rely on these participants' information outputs to enhance their forecasts. As such, the improvement in analysts' forecasting performance might convey little information to the market.

In Table OA9 of the Online Appendix, we explore the potential effects of Google Translate on a firm's several other information sources, including corporate filings, managers' earnings guidance, and media coverage. We do not find strong evidence of any significant changes in these alternative information sources.[26, 27] Given that we cannot examine all possible alternative information sources, we rely on the market reaction to analysts' forecasts as a proxy for the incremental information content to further alleviate the previous concern. In Table 7, we adopt the analyst-firm-year level regressions as in Equation (2). The dependent variable is the cumulative market-adjusted return over the three-day window around the forecast announcement date. We proxy for the information of each analyst forecast using *Revision*, calculated as the difference between the analyst' last forecast and her previous forecast, scaled by stock price. We further interact *Revision* with *Translate*, to gauge the incremental effect of Google Translate on the informativeness of analysts' forecasts. As expected, the coefficients on *Revision* are positive and highly significant in both columns of Table 8, using two different fixed-effect specifications. More importantly, the coefficients on the interaction term, *Revision* × *Translate*, are also positive and significant. The results suggest that analysts' forecasts become more informative to the market after the support of Google Translate.

In Table 8, we evaluate whether the rollout of Google Translate improves a firm's information environment. We employ the same regression specification as in Equation (1) and replace the dependent variables with the dispersion of analysts' forecasts and the stock liquidity of the focal firm. *Dispersion* is calculated as the standard deviation of analysts' forecasts, scaled by stock price, and *Liquidity* is defined as the annual average of the daily Amihud illiquidity measure,

[26] We only document a significant increase in the quantity of annual earnings guidance issued by managers, which is consistent with our prior analyses on conference calls and suggests that managers respond to analysts' greater demand for information. More importantly, we include the number of earnings guidance as a control variable that may explain analysts' forecast errors through our regressions. Our main findings are also robust to using the sample of firm-year observations without any annual earnings guidance, reducing the concern that our results are mainly driven by changes in firms' disclosure behaviors.

[27] A possible explanation for these results might be that managers and media outlets are less subject to the cross-border information frictions than analysts. Specifically, managers can have access to a rich set of foreign information through internal accounting systems or local agents (e.g., Liu and Lu 2023). Media outlets are also known to operate local branches across foreign countries with journalists who can understand local languages.

multiplied by -1. The coefficient on *Translate* is negative and significant in Column 1 and positive and significant in Column 2. The results suggest that Google Translate leads to a lower level of performance difference among analysts and better stock liquidity for the focal firm. Overall, the findings are consistent with our main analyses and corroborate the inference that Google Translate has a positive impact on analysts' forecasting performance, which leads to an improvement in firms' information environment.

5.4. Analysts' Coverage Decisions

As our last test, we explore whether Google Translate can affect the coverage decisions of analysts in Table OA10 of the Online Appendix. Prior research suggests that technological innovations may substitute the work of human labor (e.g., Grennan and Michaely 2021). As such, machine translation technology may reduce the market demand for analysts' information processing activity by lowering the processing costs for all market participants. For each analyst-firm pair in our sample, we code an indicator variable (*Coverage*) as equal to one if a firm year is covered by the analyst and zero otherwise. We estimate a linear-probability model with *Coverage* as the dependent variable and *Translate* as the key test variable, and we include a list of firm-level controls. We do not find significant evidence in support of the substitution effect between translation technology and human analysts. One possible explanation is that analysts' knowledge of local cultural and regulatory factors still matters to the market even when the language barrier to foreign information is reduced (e.g., Du et al. 2017; Merkley et al. 2020) and such knowledge may not be easily replaced by machines.

6. Conclusion

We study the impact of machine translation technology on sell-side analysts' forecasts for multinational firms. Exploiting the staggered rollout of Google Translate's support to different foreign languages over time, we find that the innovation of machine translation technology leads to a meaningful improvement in analysts' forecast accuracy for multinational firms. The improvement is stronger for analysts who lack foreign language skills or foreign coverage resources in their brokerage houses. The market reaction analysis suggests that the technology improves the informativeness of analyst research to the market. Overall, our findings suggest that machine translation technology aids the research of sell-side analysts and such a complementary effect varies with analysts' innate ability and brokerage firms' resources.

We also provide evidence to better understand how machine translation technology improves analysts' forecasting performance. Our findings suggest that Google Translate leads to 1) analysts asking more questions about firms' foreign operations and 2) incorporating more foreign economic information into forecasts. The results are consistent with the idea that Google Translate lowers the costs of processing foreign information, which in turn facilitates analysts' incorporation of such information.

The past two decades have witnessed a technological revolution in the investment research industry. The distinct features of technological innovations have the potential to transform and disrupt traditional market participants, which has drawn significant interest from both researchers and industry practitioners (Goldstein, Jiang, and Karolyi 2019). While prior work largely focuses on financial technologies and their potential competitive or substitutive effect on human labor, our research helps paint a more complete picture by showing how the public availability of certain non-financial technology may also have a complementary effect on the tasks of human market participants.

We conclude with a limitation to our study and our suggestion for future research. Our inferences on the role of machine translation technology in the sell-side research industry primarily rely on a collection of empirical evidence, anecdotes, and our conversations with a professional translator and sell-side analysts. As demonstrated by Brown et al. (2015), survey evidence based on a broad sample of analysts has become increasingly relevant for researchers to obtain direct evidence on the work of analysts, such as the inputs that they employ and their incentives. We believe that future research may benefit from such large-sample survey evidence that sheds light on how sell-side analysts adapt to technological innovations, including translation technology, to their work.

References

Acemoglu, D., and Restrepo, P. 2019. Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6), 2188–2244.

Asthana, S., Balsam, S., and Sankaraguruswamy, S. 2004. Differential Response of Small versus Large Investors to 10-K Filings on EDGAR. *The Accounting Review*, *79*(3), 571–589.

Baker, A., Larcker, D., and Wang, C. 2022. How Much Should We Trust Staggered Difference-In-Differences Estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Berger, P. G., and Hann, R. 2003. The Impact of SFAS No. 131 on Information and Monitoring. *Journal of Accounting Research*, *41*(2), 163–223.

Blackwood, M. 2020. Translation Rate in America. https://thewordpoint.com/blog/translation-rates-in-america

Borusyak, K., Jaravel, X., Spiess, J., 2021. Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv preprint. arXiv:2108.12419

Bradley, D., Gokkaya, S., and Liu, X. 2017. Before an Analyst Becomes an Analyst: Does Industry Experience Matter? *The Journal of Finance*, *72*(2), 751–792.

Bradshaw, M. T., Lock, B., Wang, X., and Zhou, D. 2021. Soft Information in the Financial Press and Analyst Revisions. *The Accounting Review*, *96*(5), 107–132.

Brochet, F., Naranjo, P., and Yu, G. 2016. The Capital Market Consequences of Language Barriers in the Conference Calls of Non-U.S. Firms. *The Accounting Review*, *91*(4), 1023–1049.

Brown, L., Call, A., Clement, M., and Sharp, N. 2015. Inside the "Black Box" of Sell-Side Financial Analysts. *Journal of Accounting Research*, *53*(1), 1–47.

Brynjolfsson, E., Hui, X., and Liu, M. 2019. Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. *Management Science*, *65*(12), 5449–5460.

Brynjolfsson, E., Li, D., and Raymond, L. R. 2023. Generative AI at Work. *National Bureau of Economic Research Working Paper Series*, *No. 31161*.

Bursztyn, L., Chaney, T., Hassan, T., and Rao, A. 2022. The Immigrant Next Door: Long-Term Contact, Generosity, and Prejudice. Working Paper, University of Chicago

Callaway, B., and Sant'Anna, P. H. C. 2021. Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Callen, J., Hope, O-K., and Segal, D., 2005. Domestic and Foreign Earnings, Stock Return Variability, and the Impact of Investor Sophistication. *Journal of Accounting Research* 43, 377–412.

Cao, S., Jiang, W., Wang, J. L., and Yang, B. 2022. From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses. Working Paper.

Cen, L., Han, Y., and Harford, J. 2022. What Do Questions Reveal? Skill Acquisition, Detection, and Recognition in the Capital Markets. Working Paper.

Chan, K., Covrig, V., and Ng, L. 2005. What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide. *The Journal of Finance*, *60*(3), 1495–1534.

Chen, D., Ma, Y., Martin, X., and Michaely, R. 2022. On the Fast Track: Information Acquisition Costs and Information Production. *Journal of Financial Economics*, *143*(2), 794–823.

Chen, X., Acosta, S., and Barry, A. 2016. Evaluating the Accuracy of Google Translate for Diabetes Education Material. *JMIR Diabetes*, 1(1), e3.

Chen, H., De, P., Hu, Y., and Hwang, B.-H. 2014. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, *27*(5), 1367–1403.

Chi, F., Hwang, B.H., and Zheng, Y. 2022. The Use and Usefulness of Big Data in Finance. Working Paper.

Clement, M. 1999. Analyst Forecast Accuracy: Do Ability, Resources, and Portfolio Complexity Matter? *Journal of Accounting and Economics*, *27*(3), 285–303.

Cohen, L., Frazzini, A., and Malloy, C. 2010. Sell-Side School Ties. *The Journal of Finance*, *65*(4), 1409–1437.

Coleman, B., Merkley, K., and Pacelli, J. 2022. Human Versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations. *The Accounting Review*, *97*(5), 221–244.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110 (9): 2964-96.

De Franco, G., and Zhou, Y. 2009. The Performance of Analysts with a CFA® Designation: The Role of Human-Capital and Signaling Theories. *The Accounting Review*, *84*(2), 383–404.

Desai, M., Foley, C., Hines Jr., J., 2008. Capital structure with risky foreign investment. *Journal of Financial Economics* 88, 534–553.

De Vries, E., Schoonvelde, M., and Schumacher, G. 2018. No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications. *Political Analysis*, 26(4), 417-430.

Do, T., and Zhang, H. 2020. Peer Effects among Financial Analysts. *Contemporary Accounting Research*, *37*(1), 358–391.

Du, Q., Yu, F., and Yu, X. 2017. Cultural Proximity and the Processing of Financial Information. *Journal of Financial and Quantitative Analysis*, *52*(6), 2703–2726.

Duru, A., and Reeb, D. 2002. International Diversification and Analysts' Forecast Accuracy and Bias. *The Accounting Review*, *77*(2), 415–433.

Dyreng, S., Hoopes, J., Langetieg, P., and Wilde, J. 2020. Strategic Subsidiary Disclosure. *Journal of Accounting Research*, *58*(3), 643–692.

Dyreng, S., and Lindsey, B. 2009. Using Financial Accounting Data to Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U.S. Multinational Firms' Tax Rates. *Journal of Accounting Research*, *47*(5), 1283–1316.

Gibbons, B., Iliev, P., and Kalodimos, J. 2021. Analyst Information Acquisition via EDGAR. *Management Science*, *67*(2), 769–793.

Giles, J. 2006. Google Tops Translation Ranking. *Nature News.* https://www.nature.com/news/2006/061106/full/news061106-6.html

Goldstein, I., Jiang, W., and Karolyi, G. A. 2019. To FinTech and Beyond. *The Review of Financial Studies*, 32(5), 1647-1661.

Grennan, J., and Michaely, R. 2020. Artificial Intelligence and High-Skilled Work: Evidence from Analysts. Working Paper.

Grennan, J., and Michaely, R. 2021. Fintechs and the Market for Financial Analysis. *Journal of Financial and Quantitative Analysis*, 56(6), 1877-1907.

Grinblatt, M., and Keloharju, M., 2001. How Distance, Language, and Culture Influence Stockholdings and Trades. *Journal of Finance*. 56, 1053–1073.

Groysberg, B., and Healy, P. 2013. Wall Street Research: Past, Present, and Future. Redwood City: Stanford University Press.

Hope, O.-K. 2003. Disclosure Practices, Enforcement of Accounting Standards, and Analysts' Forecast Accuracy: An International Study. *Journal of Accounting Research*, 41: 235-272.

Hope, O.-K., and Su, X. 2021. Peer-level Analyst Transitions. *Journal of Corporate Finance*, *70*, 102072 pages.
Huang, X. 2015. Thinking Outside the Borders: Investors' Underreaction to Foreign Operations Information. *The Review of Financial Studies*, *28*(11), 3109–3152.
Huang, A. H., Lehavy, R., Zang, A. Y., and Zheng, R. 2018. Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science*, *64*(6), 2833–2855.
Huang, A., Lin, A.-P., and Zang, A. 2022. Cross-Industry Information Sharing Among Colleagues and Analyst Research. *Journal of Accounting and Economics*, *74*(1), 101496.
Hugon, A., Kumar, A., and Lin, A.-P. 2016. Analysts, Macroeconomic News, and the Benefit of Active In-House Economists. *The Accounting Review*, *91*(2), 513–534.
Hutton, A., Lee, L., and Shu, S. 2012. Do Managers Always Know Better? The Relative Accuracy of Management and Analyst Forecasts. *Journal of Accounting Research*, *50*(5), 1217–1244.
Intento. 2022. State of Machine Translation 2022.
Jame, R., Markov, S., and Wolfe, M. 2022. Can FinTech Competition Improve Sell-Side Research Quality? *The Accounting Review*, *97*(4), 287–316.
Jennings, J., Lee, J., and Matsumoto, D. 2017. The Effect of Industry Co-Location on Analysts' Information Acquisition Costs. *The Accounting Review*, *92*(6), 103–127.
Johnson, S. 2016. Six Ways to Find New Investments: Stock Picking. *The Australian Financial Review.*
Kakaes, K. 2011. Google, Yahoo! Babelfish Use Math Principles to Translate Documents Online. *The Washington Post.*
Kawoosa, V. 2018. How Languages Intersect in India, *Hindustan Times* https://www.hindustantimes.com/india-news/how-languagesintersect-in-india/story-g3nzNwFppYV7XvCumRzlYL.html
Kong, D., Lin, C., Wei, L., and Zhang, J. 2022. Information Accessibility and Corporate Innovation. *Management Science*, *68*(11), 7837–7860.
Krishnan, G., Singer, Z., and Zhang, J. 2023. Audit Partner Ethnicity and Salient Audit Phenomena. *Accounting, Organizations and Society*, 101440.
Lang, T., Stice-Lawrence, L., Wong, F., Wong, T.J. 2023. Differential Treatment and Local Information Advantage: Revelations from Translation Differences. Working Paper.
Li, N., Richardson, S., and Tuna, İ. 2014. Macro to Micro: Country Exposures, Firm Fundamentals and Stock Returns. *Journal of Accounting and Economics*, *58*(1), 1–20.
Lin, L., Mihov, A., Sanz, L., and Stoyanova, D. 2019. Property Rights Institutions, Foreign Investment, and the Valuation of Multinational Firms. *Journal of Financial Economics*, *134*(1), 214–235
Liu, L., and Lu, S. 2023. The Effect of Firms' Information Exposure on Safeguarding Employee Health: Evidence from COVID-19. *Journal of Accounting Research*, *forthcoming*
Lundholm, R., Rahman, N., and Rogo, R. 2018. The Foreign Investor Bias and its Linguistic Origins. *Management Science*, 64(9), 4433-4450.
Mayew, W., Sharp, N., and Venkatachalam, M. 2013. Using Earnings Conference Calls to Identify Analysts with Superior Private Information. *Review of Accounting Studies*, *18*(2), 386–413.
Merkley, K., Michaely, R., and Pacelli, J. 2020. Cultural diversity on Wall Street: Evidence from consensus earnings forecasts. *Journal of Accounting and Economics*, *70*(1), 101330.
Munz, K., Jung, M., and Alter, A. 2020. Name Similarity Encourages Generosity: A Field Experiment in Email Personalization. *Marketing Science*, *39*(6), 1071–1091.

Noy, S., and Zhang, W. 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science*, 381(6654), 187–192.

O'Brien, P., and Tan, H. 2015. Geographic Proximity and Analyst Coverage Decisions: Evidence From IPOs. *Journal of Accounting and Economics*, *59*(1), 41–59.

Patil, S., and Davies, P. 2014. Use of Google Translate in Medical Communication: Evaluation of Accuracy. *BMJ: British Medical Journal*, 349.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. 2023. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. Working paper.

Shroff, N., Verdi, R., and Yu, G. 2014. Information Environment and the Investment Decisions of Multinational Corporations. *The Accounting Review*, 89(2), 759-790.

Stasimitoti, M. 2022. How Europe's Second Largest Bank Built Multilingual Machine Translation In-House. https://slator.com/how-europes-second-largest-bank-built-multilingual-machine-translation-in-house/

Strumpf, D. 2022. Google Pulls Translation App from China. *Wall Street Journal*. https://www.wsj.com/articles/google-pulls-translation-app-from-china-11664795087.

Sultana, N. 2021. The Language Barrier Plaguing India's Stock Market. *Mint*. https://www.livemint.com/market/stock-market-news/why-english-only-markets-breed-inequity-11629994292128.html

Thomas, W. 1999. A Test of the Market's Mispricing of Domestic and Foreign Earnings. *Journal of Accounting and Economics*, 28(3): 243–267.

Turovsky, B. 2016. Ten Years of Google Translate. https://www.blog.google/products/translate/ten-years-of-google-translate/

Wang, K., Yu, X., and Zhang, B. 2022. Panda Games: Corporate Disclosure in the Eclipse of Search. *Management Science*, *69*(6), 3263–3284.

Wu, H. D. 2000. Systemic Determinants of International News Coverage: A Comparison of 38 Countries', *Journal of Communication* 50(2): 110–30.

Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J., 2016. Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation. CoRR abs/1609.08144.

Xu, Y., Xuan, Y., and Zheng, G. 2021. Internet Searching and Stock Price Crash Risk: Evidence From a Quasi-Natural Experiment. *Journal of Financial Economics*, *141*(1), 255–275.

Zhang, R. 2022. Language Commonality and Sell-Side Information Production. *Management Science*, *68*(6), 4435–4453.

Appendix A: Variable Definitions

Variable	Definition	Source
Variables of Interest		
Translate Percentage (%)	The number of pre-existing (lagged) foreign subsidiaries domiciled in countries where the most commonly-used language is supported by Google Translate, scaled by the total number of subsidiaries and multiplied by 100.	EDGAR, Google Translate
Translate	An indicator that equals one if *Translate Percentage* is 10% or greater, zero otherwise.	EDGAR, Google Translate
Translate (Non-Proficient)	An indicator that equals one if the company operates at least 10 percent of subsidiaries in countries where the most commonly-used language is supported by Google Translate and is not a language that the analyst is likely to proficient in, zero otherwise. We use a machine-learning-based software (NamSor) to infer the analyst's most probable ethnicity from her first and last names, and then assign the most commonly-used language by the ethnic group as the language that the analyst is likely to be proficient in.	EDGAR, Google Translate, I/B/E/S, NamSor
Translate (Proficient)	An indicator that equals one if the company operates at least 10 percent of subsidiaries in countries where the most commonly-used language is supported by Google Translate but is also a language that the analyst is likely to be proficient in, zero otherwise.	EDGAR, Google Translate, I/B/E/S, NamSor
Translate (Non-Covered)	An indicator that equals one if the company operates at least 10 percent of subsidiaries in countries where the most commonly-used language is supported by Google Translate and where the analyst's brokerage firm doesn't cover any local companies from the country, zero otherwise.	EDGAR, Google Translate, I/B/E/S
Translate (Covered)	An indicator that equals one if the company operates at least 10 percent of subsidiaries in countries where the most commonly-used language is supported by Google Translate and where the analyst's brokerage firm covers at least one local company from the country, zero otherwise.	EDGAR, Google Translate, I/B/E/S
Dependent Variables		
Error (%)	The absolute difference between consensus forecast (Tables 3-4) or individual analyst forecast (Table 5) and actual earnings per share scaled by the stock price at the end of the fiscal period and multiplied by 100. Consensus forecast is calculated as the mean value of the last forecast issued by each analyst within the 1 to 12 month horizon.	I/B/E/S, Compustat
Foreign Question (%)	The total number of analyst questions that mention at least one of the foreign countries where the most commonly-used language is supported by Google Translate in our sample period, as a percentage of the total number of questions in the Q&A session of conference calls. We use the fourth quarter earnings announcement to construct the measure at the firm-year level.	Capital IQ
Firm Specifics Only (%)	The percentage of analysts' foreign questions related to one of the three topics on firm specifics, including (1) operational costs and profitability, (2) financial details and outlooks, and (3) business opportunities and development. The most frequent keywords under each topic are provided in Appendix C.	Capital IQ
Related Entities (%)	The percentage of analysts' foreign questions related to one of the four topics on a firm's related entities, including (1) production and supply chains, (2) acquisitions and partnerships, (3) industry dynamics and	Capital IQ

	competition, and (4) capital financing and investment. The most frequent keywords under each topic are provided in Appendix C.	
Macro Environment & Policies (%)	The percentage of analysts' foreign questions related to one of the three topics on macroenvironment and policies, including (1) currency risk and trade agreement, (2) impact of exports and tariffs on sales, and (3) regional market conditions and strategy. The most frequent keywords under each topic are provided in Appendix C.	Capital IQ
ROA (%)	Income before extraordinary Items scaled by total assets, multiplied by 100.	Compustat
Forecast-Implied ROA (%)	The consensus forecast of earnings per share multiplied by the number of common shares outstanding, scaled by total assets and multiplied by 100.	I/B/E/S, Compustat
CAR (%)	Market-adjusted return over the three-day window around the forecast announcement date, multiplied by 100	CRSP
Liquidity	The mean value of daily Amihud illiquidity measure over the year, multiplied by −1. Amihud illiquidity is calculated as the absolute value of return divided by stock price times with volume, multiplied by 1 million.	CRSP
Dispersion (%)	The standard deviation of analysts' forecasts included in the consensus, scaled by the stock price at the end of the fiscal period, multiplied by 100.	I/B/E/S
Control Variables		
Size	The logarithm of market capitalization.	Compustat
BTM	Book value of equity divided by market value at the end of the fiscal period.	Compustat
Loss	An indicator that equals one if *ROA* is negative, zero otherwise.	Compustat
EarnVol	Standard deviation of quarterly *ROA* in the past five years.	Compustat
RetVol	Standard deviation of monthly stock returns in the past 12 months.	CRSP
Leverage	Total debt divided by total assets.	Compustat
R&D	Research and development expenses divided by total assets.	Compustat
Guidance Count	The logarithm of one plus the total number of management earnings guidance issued in the year.	I/B/E/S
ForeignSub	The logarithm of one plus the number of foreign subsidiaries.	EDGAR
InstOwn	The percentage of shares owned by institutional investors.	Thomson Reuters
Brokersize	The logarithm of one plus the number of unique analysts employed by the analyst's brokerage firm in the year.	I/B/E/S
General Experience	The logarithm of one plus the number of years since the analysts' first appearance in I/B/E/S.	I/B/E/S
Firm Experience	The logarithm of one plus the number of years since the analyst started covering the firm.	I/B/E/S
Portfolio Size	The number of unique firms covered by the analyst in the year.	I/B/E/S
Industry Count	The logarithm of one plus the number of unique SIC2 industries covered by the analyst in the year.	I/B/E/S, Compustat
Effort	The logarithm of one plus the number of forecasts issued by the analyst for the firm in the year.	I/B/E/S
Horizon	The logarithm of one plus the number of days between the forecast announcement date and earnings announcement date.	I/B/E/S
GDP Growth (Supported)	The weighted average GDP growth across countries where the firm reports a foreign geographic segment and where the most commonly-used language is ever supported by Google Translate in our sample period.	Compustat, World Bank

GDP Growth (Non-Supported)	The weighted average GDP growth across countries where the firm has a foreign geographic segment and where the most commonly-used language is not supported by Google Translate in our sample period.	Compustat, World Bank
Revision (%)	The difference between the analyst's last forecast and their previous forecast, scaled by the stock price at the end of the fiscal period and multiplied by 100.	I/B/E/S

Appendix B. Languages Supported by Google Translate by Year

This table presents the new languages supported by Google Translate by year.

Year	Languages	Count
2006	Arabic, Chinese (Simplified), French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish	10
2007	Chinese (Traditional), Dutch, Greek	3
2008	Bulgarian, Catalan, Croatian, Czech, Danish, Finnish, Hebrew, Hindi, Indonesian, Latvian, Lithuanian, Norwegian, Polish, Romanian, Serbian, Slovak, Slovene, Swedish, Tagalog (Filipino), Ukrainian, Vietnamese	21
2009	Afrikaans, Albanian, Belarusian, Estonian, Galician, Hungarian, Icelandic, Irish, Macedonian, Malay, Maltese, Persian, Swahili, Thai, Turkish, Welsh, Yiddish	17
2010	Armenian, Azerbaijani, Basque, Georgian, Haitian Creole, Latin, Urdu	7
2011	Bengali, Gujarati, Kannada, Tamil, Telugu	5
2012	Esperanto, Lao	2
2013	Bosnian, Cebuano, Hausa, Hmong, Igbo, Javanese, Khmer, Maori, Marathi, Mongolian, Nepali, Punjabi, Somali, Yoruba, Zulu	15
2014	Burmese, Chichewa, Kazakh, Malagasy, Malayalam, Sesotho, Sinhala, Sundanese, Tajik, Uzbek	10
2016	Amharic, Corsican, Frisian, Hawaiian, Kurdish, Kyrgyz, Luxembourgish, Pashto, Samoan, Scots Gaelic, Shona, Sindhi, Xhosa	13
2020	Kinyarwanda, Odia, Tatar, Turkmen, Uyghur	5
	Total	108

Appendix C. Topics of Analysts' Questions Related to Firms' Foreign Exposure

This appendix presents the inferred labels of the 10 topics that we model using analysts' questions that mention at least one foreign country or region. The sample of questions is collected from annual earnings conference calls.

Topic	Label	Top 10 Words
	Firm Specifics Only	
1.	Operational Costs and Profitability	margin, cost, operation, basis, profit, improvement, benefit, expense, material, decline
2.	Financial Details and Outlooks	business, number, guidance, rate, consumer, factor, assumption, range, earnings, weakness
3.	Business Opportunities and Development	opportunity, customer, company, plan, project, asset, order, development, activity, base
	Related Entities	
4.	Production and Supply Chains	product, capacity, issue, supplier, data, process, facility, plant, government
5.	Acquisitions and Partnerships	acquisition, contract, deal, discussion, partner, distribution, channel, release, congratulation, press
6.	Industry Dynamics and Competition	price, volume, demand, increase, environment, situation, industry, competitor, inventory
7.	Capital Financing and Investment	cash, investment, capital, bank, capex, flow, program, balance, game, spending
	Macro Environment & Policy	
8.	Currency Risk and Trade Agreement	impact, case, risk, currency, loss, exposure, trade, dollar, ebitda, agreement
9.	Impact of Exports and Tariffs on Sales	revenue, service, percentage, store, export, contribution, unit, client, rest, tariff
10.	Regional Market Conditions and Strategy	market, term, growth, sale, couple, change, line, region, strategy, level

Figure 1. Launch Year of Google Translate's Support to Local Languages

This map labels countries with the launch years since which their most commonly-used languages have been supported by Google Translate.

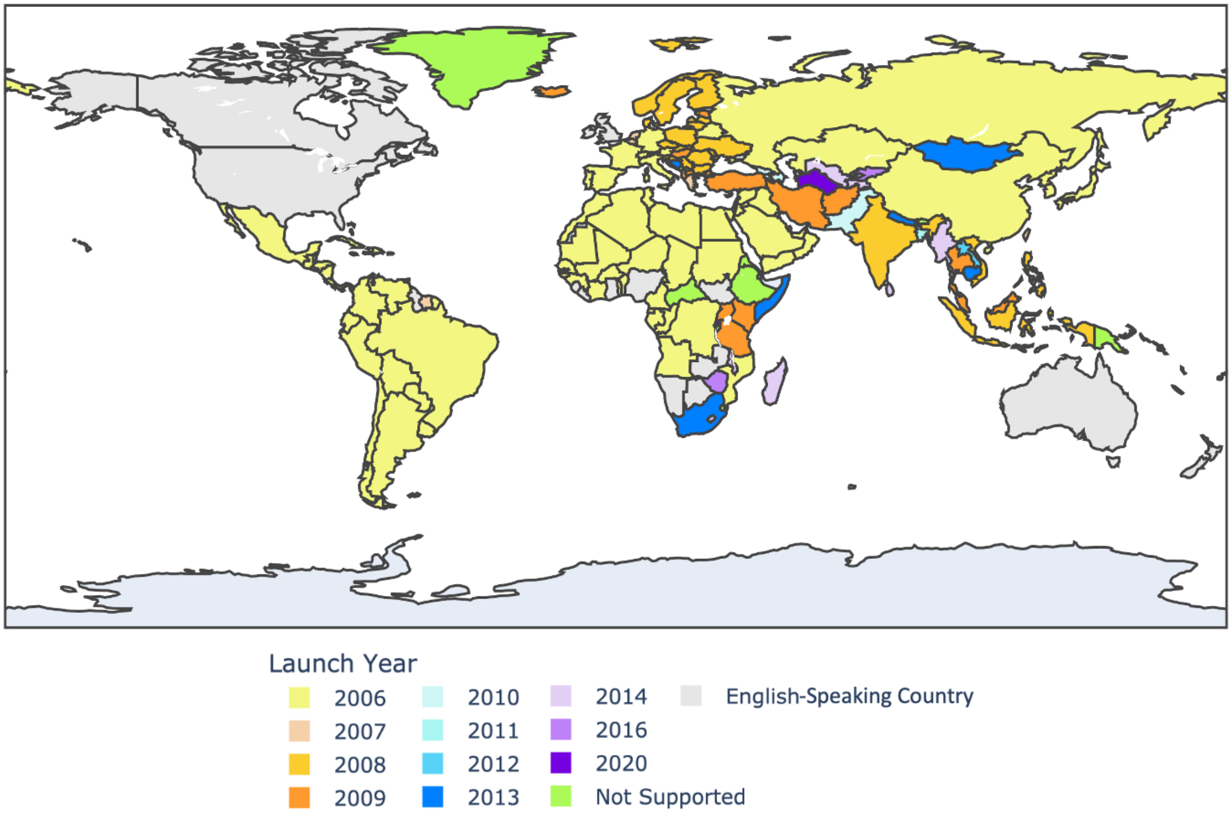


Figure 2. Parallel Trends

This figure examines the parallel trends assumption using the firm-year level sample. We plot the estimated coefficients by year in event time. The x-axis denotes the relative treatment year. The y-axis plots the coefficients for each event year estimated using the regression specification in Equation (1). We control for the percentage of treated subsidiaries below the 10% threshold to remove their effect on the pre-trend estimates. The dots (connected horizontally) represent the estimated coefficient, and the vertical lines represent 95% confidence intervals.

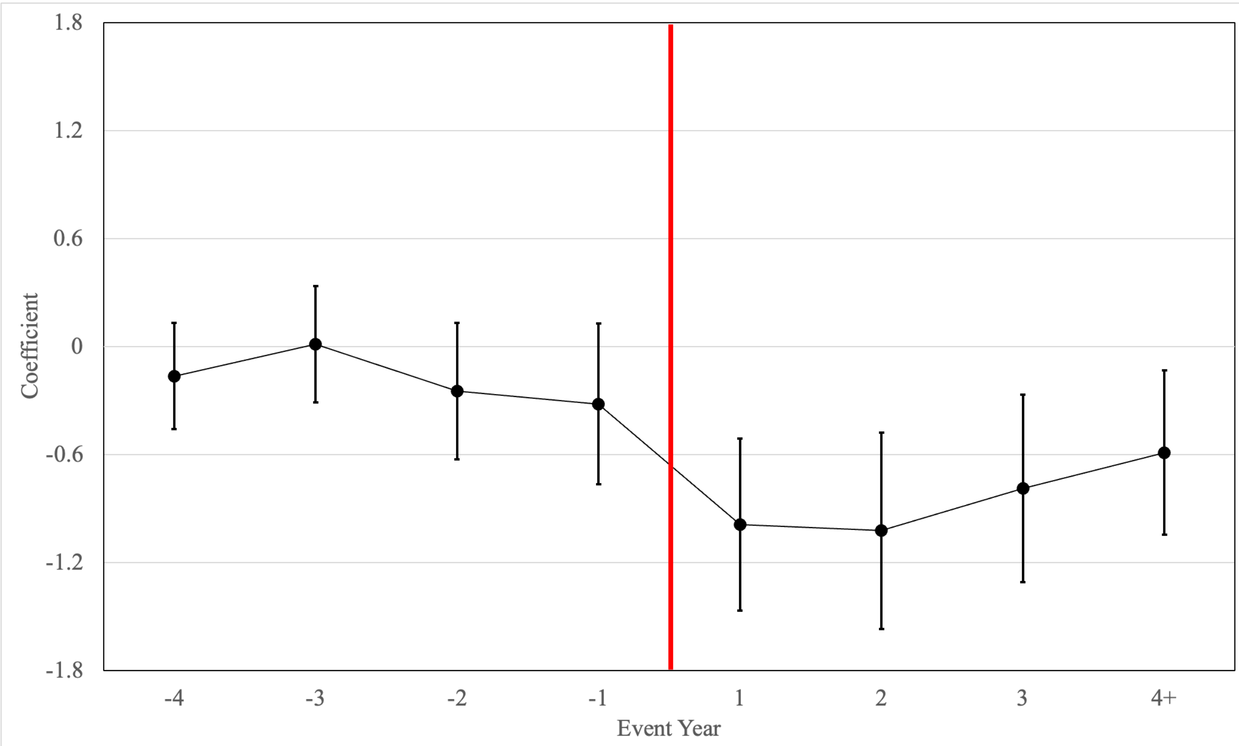


Table 1: Sample Construction and Distribution

This table presents the construction of our main sample.

Panel A. Sample Construction		
	Obs.	N Firms
Compustat firm-year observations with data from I/B/E/S to calculate analysts' forecast error	94,876	12,219
Less: Firms that are not headquartered or incorporated in the U.S.	(14,053)	(1,924)
Less: Observations with missing lagged subsidiary data	(26,495)	(3,271)
Less: Observations with missing control variables	(3,949)	(575)
Less: Firms with only one observation in our sample (singleton observations)	(929)	(929)
Final sample	49,450	5,520

Panel B: Number of Firms by First Treated Year

Fiscal Year	Count	Percentage (%)
2006	35	0.63
2007	898	16.27
2008	231	4.18
2009	171	3.1
2010	90	1.63
2011	81	1.47
2012	100	1.81
2013	83	1.5
2014	92	1.67
2015	85	1.54
2016	87	1.58
2017	87	1.58
2018	67	1.21
2019	68	1.23
2020	31	0.56
Never-Treated	3,314	60.04
Total	5,520	100

Table 2: Summary Statistics

This table presents the summary statistics for the variables used in the firm-year level analyses. Variable definitions are available in Appendix A.

	(1) N	(2) Mean	(3) SD	(4) p25	(5) p50	(6) p75
Error (%)	49,450	2.14	7.40	0.10	0.30	0.99
Translate	49,450	0.30	0.46	0.00	0.00	1.00
Translate Percentage (%)	49,450	13.40	22.86	0.00	0.00	21.43
Foreign Question (%)	26,363	2.40	5.27	0.00	0.00	2.44
Forecast-Implied ROA (%)	49,450	1.77	13.62	0.92	3.61	7.08
Dispersion	46,152	1.65	4.80	0.12	0.33	1.01
Liquidity	47,770	-0.14	0.65	-0.02	0.00	0.00
Size	49,450	7.07	1.83	5.80	7.01	8.26
BTM	49,450	0.60	0.52	0.28	0.50	0.80
Loss	49,450	0.25	0.44	0.00	0.00	1.00
ROA (%)	49,450	0.02	15.65	-0.12	2.74	6.66
EarnVol	49,450	0.02	0.04	0.00	0.01	0.02
RetVol	49,450	0.12	0.07	0.07	0.10	0.15
Leverage	49,450	0.26	0.22	0.07	0.22	0.39
R&D	49,450	0.04	0.08	0.00	0.00	0.03
ForeignSub	49,450	1.70	1.56	0.00	1.39	2.89
Guidance Count	49,450	0.49	0.74	0.00	0.00	1.10
InstOwn	49,450	0.59	0.33	0.35	0.68	0.86

Table 3: Main Results

This table presents the results of the effect of Google Translate on the errors of consensus analyst forecasts at the firm-year level. The dependent variable, *Error*, is calculated as the absolute difference between the consensus analyst earnings forecast and the actual value, scaled by the stock price. The test variable *Translate* is an indicator equal to one if at least ten percent of the firm's pre-existing foreign subsidiaries are located in countries where the most commonly-used language has been supported by Google Translate. Column 1 examines the univariate difference. Column 2 includes firm and year fixed effects. Column 3 further includes a set of common firm characteristics. Variable definitions are available in Appendix A. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)
	Error (%)	Error (%)	Error (%)
Translate	-0.56***	-0.46***	-0.48***
	(-6.04)	(-2.99)	(-3.42)
Size			-2.60***
			(-18.58)
BTM			1.57***
			(6.08)
Loss			0.88***
			(5.52)
ROA			-0.12***
			(-11.94)
EarnVol			1.51
			(0.52)
RetVol			1.84**
			(2.00)
Leverage			1.49***
			(3.31)
R&D			-13.27***
			(-6.46)
ForeignSub			0.32***
			(4.85)
Guidance Count			-0.04
			(-0.60)
InstOwn			-1.50***
			(-4.87)
Firm FE	No	Yes	Yes
Year FE	No	Yes	Yes
Obs.	49,450	49,450	49,450
Adjusted R^2	0.00	0.25	0.39

Table 4: Analyst-Level Results

This table presents the summary statistics and results of the effect of Google Translate on individual analyst forecast errors at the analyst-firm-year level. Panel A presents the summary statistics of the variables used in the analyses. Panel B presents the regression results. Column 1 examines the average effect of Google Translate on individual analyst forecast errors. Column 2 separates the effect of Google Translate into two components, *Translate (Non-Proficient)* and *Translate (Proficient)*, by considering whether the analyst is likely to be proficient in the language supported by Google Translate. Column 3 separates the effect of Google Translate into the other two components, *Translate (Non-Covered)* and *Translate (Covered)*, by considering whether the analyst's brokerage firm employs at least one analyst that covers local companies in the foreign country. Columns 1 to 3 include firm, analyst, and year fixed effects. Columns 4 to 6 provide the results with analyst-firm pair and year fixed effects. Variable definitions are available in Appendix A. Standard errors are clustered by firm and analyst. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

Panel A: Analyst-Level Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Mean	SD	p25	p50	p75
Error (%)	524,078	1.24	3.96	0.06	0.20	0.67
Translate	524,078	0.35	0.48	0.00	0.00	1.00
Translate (Non-Proficient)	497,010	0.36	0.48	0.00	0.00	1.00
Translate (Proficient)	497,010	0.02	0.13	0.00	0.00	0.00
Translate (Non-Covered)	524,078	0.29	0.46	0.00	0.00	1.00
Translate (Covered)	524,078	0.10	0.30	0.00	0.00	0.00
Size	524,078	8.13	1.81	6.88	8.09	9.39
BTM	524,078	0.52	0.44	0.24	0.43	0.71
Loss	524,078	0.21	0.41	0.00	0.00	0.00
ROA (%)	524,078	2.08	12.90	0.64	3.70	7.84
EarnVol	524,078	0.02	0.03	0.00	0.01	0.02
RetVol	524,078	0.11	0.07	0.06	0.09	0.14
Leverage	524,078	0.26	0.21	0.08	0.23	0.38
R&D	524,078	0.03	0.07	0.00	0.00	0.04
ForeignSub	524,078	2.06	1.66	0.69	1.95	3.40
Guidance Count	524,078	0.61	0.80	0.00	0.00	1.61
InstOwn	524,078	0.67	0.31	0.54	0.76	0.89
Brokerage Size	524,078	3.75	1.02	3.09	3.89	4.60
General Experience	524,078	2.18	0.90	1.61	2.40	2.89
Firm Experience	524,078	1.19	0.89	0.69	1.10	1.79
Portfolio Size	524,078	2.75	0.52	2.48	2.83	3.09
Industry Count	524,078	1.43	0.51	1.10	1.39	1.79
Effort	524,078	1.50	0.47	1.10	1.61	1.79
Horizon	524,078	4.69	0.57	4.49	4.62	4.93

Table 4 (Continued)						
Panel B. Analyst-Level Results						
	(1)	(2)	(3)	(4)	(5)	(6)
	Error (%)	Error (%)	Error (%)	Error (%)	Error (%)	Error (%)
Translate	-0.19***			-0.12*		
	(-2.72)			(-1.79)		
Translate (Non-Proficient)		-0.18***			-0.13**	
		(-2.64)			(-2.02)	
Translate (Proficient)		-0.03			0.04	
		(-0.56)			(0.43)	
Translate (Non-Covered)			-0.18***			-0.14***
			(-3.48)			(-2.65)
Translate (Covered)			-0.06			-0.04
			(-1.49)			(-0.84)
Size	-1.35***	-1.36***	-1.35***	-1.72***	-1.71***	-1.72***
	(-15.59)	(-15.41)	(-15.59)	(-17.34)	(-17.23)	(-17.35)
BTM	1.33***	1.32***	1.33***	1.07***	1.06***	1.07***
	(8.60)	(8.37)	(8.60)	(6.26)	(6.14)	(6.26)
Loss	0.89***	0.88***	0.89***	0.70***	0.71***	0.70***
	(9.09)	(8.84)	(9.09)	(7.11)	(7.06)	(7.11)
ROA	-0.05***	-0.05***	-0.05***	-0.04***	-0.04***	-0.04***
	(-9.05)	(-8.73)	(-9.05)	(-8.27)	(-8.06)	(-8.27)
EarnVol	1.43	1.18	1.43	3.01	3.04	3.00
	(0.90)	(0.74)	(0.90)	(1.60)	(1.61)	(1.60)
RetVol	0.50	0.56	0.50	0.02	0.05	0.02
	(1.11)	(1.22)	(1.11)	(0.04)	(0.11)	(0.04)
Leverage	1.24***	1.26***	1.24***	0.89***	0.89***	0.89***
	(5.84)	(5.87)	(5.85)	(3.89)	(3.89)	(3.89)
R&D	-4.56***	-4.62***	-4.57***	-6.31***	-6.36***	-6.31***
	(-4.59)	(-4.58)	(-4.60)	(-5.75)	(-5.73)	(-5.75)
ForeignSub	0.12***	0.12***	0.12***	0.10***	0.10***	0.10***
	(4.68)	(4.58)	(4.63)	(3.43)	(3.42)	(3.46)
Guidance Count	-0.10***	-0.11***	-0.10***	-0.14***	-0.14***	-0.14***
	(-3.12)	(-3.27)	(-3.12)	(-4.16)	(-4.20)	(-4.16)
InstOwn	-1.12***	-1.11***	-1.12***	-1.23***	-1.23***	-1.23***
	(-6.68)	(-6.62)	(-6.67)	(-6.33)	(-6.29)	(-6.32)
Brokerage Size	0.06***	0.06***	0.06***	0.08***	0.08***	0.08***
	(3.84)	(3.92)	(3.64)	(3.92)	(4.03)	(3.75)
General Experience	0.14***	0.14***	0.14***	-0.03	-0.03	-0.03
	(5.31)	(5.31)	(5.27)	(-0.85)	(-0.66)	(-0.86)
Firm Experience	0.06***	0.07***	0.06***	0.38***	0.38***	0.38***
	(6.25)	(6.27)	(6.24)	(11.23)	(11.12)	(11.23)
Portfolio Size	-0.10***	-0.11***	-0.10***	-0.11***	-0.11***	-0.11***
	(-3.45)	(-3.77)	(-3.48)	(-2.91)	(-3.01)	(-2.92)

Table 4 Panel B (Continued)						
Industry Count	-0.05	-0.05	-0.05	-0.06	-0.07	-0.07
	(-1.34)	(-1.27)	(-1.34)	(-1.48)	(-1.52)	(-1.48)
Effort	-0.21***	-0.21***	-0.21***	-0.23***	-0.23***	-0.23***
	(-9.64)	(-9.43)	(-9.63)	(-9.22)	(-9.12)	(-9.22)
Horizon	0.67***	0.67***	0.67***	0.63***	0.63***	0.63***
	(33.99)	(33.39)	(33.99)	(30.71)	(30.14)	(30.71)
Firm FE	Yes	Yes	Yes	No	No	No
Analyst FE	Yes	Yes	Yes	No	No	No
Analyst-Firm FE	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	524,078	496,978	524,078	478,226	459,297	478,226
Adjusted R^2	0.43	0.43	0.43	0.44	0.44	0.44
Coefficient Difference						
Non-Proficient **LESS** Proficient		-0.15**			-0.17*	
p-value		0.04			0.10	
Non-Covered **LESS** Covered			-0.12***			-0.10**
p-value			0.00			0.02

Table 5: Analysts' Questions in Conference Calls

This table presents the results of the effect of Google Translate on analysts' questions in conference calls. Column 1 examines the percentage of analysts' questions that mention at least one of the foreign countries where the most commonly-used language is ever supported by Google Translate during earnings conference calls (*Foreign Question*) using the sample with available data. Columns 2 to 4 separately examine the percentage of these foreign questions across three categories, which are classified based on the questions' thematic content as detailed in Appendix C. Specifically, Column 2 studies questions on firm-specific details, such as foreign operational performance, financial conditions, and business opportunities. Column 3 examines questions that involve a firm's related foreign entities including suppliers, business partners, industry rivals, etc. Column 4 focuses on questions related to foreign macroeconomic environment and policies, such as currency risk, impact of exports and tariffs, and regional market conditions. Detailed variable definitions are available in Appendix A. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

		Foreign Questions by Category		
	(1)	(2)	(3)	(4)
	Foreign Question	Firm Specifics Only	Related Entities	Macro Environment & Policies
Translate	0.44***	0.03	0.05**	0.25**
	(3.01)	(0.93)	(2.02)	(2.32)
Size	0.27***	0.04**	0.04***	0.19***
	(3.36)	(2.17)	(2.81)	(3.25)
BTM	0.12	0.02	0.05**	0.07
	(1.19)	(0.82)	(2.55)	(1.01)
Loss	-0.18*	-0.02	-0.04**	-0.17**
	(-1.86)	(-0.70)	(-2.00)	(-2.26)
ROA	0.01*	-0.00*	0.00	0.00
	(1.91)	(-1.67)	(1.29)	(1.42)
EarnVol	-0.02	0.03	-0.43	0.66
	(-0.01)	(0.07)	(-1.51)	(0.48)
RetVol	-1.01	-0.28*	0.02	-0.51
	(-1.51)	(-1.73)	(0.16)	(-1.03)
Leverage	0.24	-0.10	-0.06	0.44
	(0.63)	(-1.07)	(-0.95)	(1.60)
R&D	-0.24	-0.45	-0.15	0.19
	(-0.16)	(-1.38)	(-0.65)	(0.18)
ForeignSub	0.15*	0.00	0.01	0.08
	(1.84)	(0.22)	(0.47)	(1.36)
Guidance Count	0.08	0.01	-0.01	0.08
	(0.90)	(0.45)	(-0.79)	(1.31)
InstOwn	0.10	0.05	-0.05	0.04
	(0.46)	(0.96)	(-1.35)	(0.27)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs.	26,032	26,004	26,004	26,004
Adjusted R^2	0.41	0.11	0.08	0.33

Table 6: Geographic Economic Conditions and Analysts' Forecasts

This table presents the results for the association between analysts' forecasts and firms' exposure to foreign macroeconomic information. Column 1 examines the association between a firm's actual *ROA* and the two sets of macroeconomic information. *GDP Growth (Supported)* is the aggregated GDP growth weighted by the firm's corresponding sales across countries where the most commonly-used language is ever supported by Google Translate in our sample period. *GDP Growth (Non-Supported)* is the GDP growth aggregated across countries where the most commonly-used language is not supported by Google Translate in our sample period or is English. Column 2 examines *Forecast-Implied ROA*, calculated as the consensus analyst forecast of earnings per share multiplied by the number of common shares outstanding divided by total assets. Column 3 further includes the interaction terms between *Translate* and the two *GDP Growth* variables. Variable definitions are available in Appendix A. Control variables are calculated from the previous year. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1) ROA (%)	(2) Forecast-Implied ROA (%)	(3) Forecast-Implied ROA (%)
GDP Growth (Supported)	0.63***	0.30***	0.18*
	(4.70)	(3.85)	(1.67)
GDP Growth (Supported) × Translate			0.21*
			(1.92)
GDP Growth (Non-Supported)	0.30***	0.21***	0.27***
	(2.93)	(3.05)	(2.94)
GDP Growth (Non-Supported) × Translate			-0.14
			(-1.53)
Translate			-0.12
			(-0.49)
Size	-0.61**	0.87***	0.87***
	(-2.21)	(5.05)	(5.07)
BTM	-7.13***	-2.60***	-2.60***
	(-16.09)	(-8.94)	(-8.96)
Loss	-1.30***	-0.11	-0.11
	(-3.98)	(-0.56)	(-0.56)
ROA	0.20***	0.22***	0.22***
	(10.49)	(13.88)	(13.87)
EarnVol	1.50	-3.48	-3.43
	(0.19)	(-0.54)	(-0.53)
RetVol	-6.19***	-3.19**	-3.03**
	(-2.93)	(-2.10)	(-1.98)
Leverage	-2.69***	-2.03***	-2.04***
	(-2.60)	(-2.79)	(-2.80)
R&D	10.12*	-5.37	-5.49
	(1.70)	(-1.26)	(-1.29)
ForeignSub	-0.50***	-0.34***	-0.32***
	(-4.04)	(-4.19)	(-3.91)
Guidance Count	-0.33**	0.08	0.08
	(-2.45)	(1.09)	(1.07)

Table 6 (Continued)			
InstOwn	-0.01	-0.09	-0.09
	(-0.02)	(-0.19)	(-0.20)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs.	23,897	23,897	23,897
Adjusted R^2	0.55	0.71	0.71

Table 7: Market Reaction to Forecast Revisions

This table presents the results of the effect of Google Translate on the market reaction to analysts' forecast revisions. The dependent variable, *CAR*, is the three-day cumulative abnormal return around the forecast announcement date. The independent variable, *Revision*, is the difference between the analyst' last forecast and her previous forecast, scaled by the stock price. Column 1 includes firm, analyst, and year fixed effects. Column 2 includes analyst-firm pair and year fixed effects. Variable definitions are available in Appendix A. Standard errors are clustered by firm and analyst. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1) CAR (%)	(2) CAR (%)
Revision	0.71***	0.74***
	(29.07)	(25.33)
Translate	0.01	0.09
	(0.05)	(0.64)
Revision × Translate	0.24***	0.27***
	(4.82)	(4.53)
Size	0.34***	0.59***
	(4.84)	(6.77)
BTM	-1.13***	-1.12***
	(-9.10)	(-7.46)
Loss	0.38***	0.52***
	(3.24)	(3.89)
ROA	0.00	-0.00
	(0.61)	(-0.37)
EarnVol	-1.06	1.01
	(-0.39)	(0.31)
RetVol	2.68***	2.59***
	(3.45)	(2.94)
Leverage	-0.30	0.23
	(-0.90)	(0.57)
R&D	-4.11**	-2.58
	(-2.48)	(-1.25)
ForeignSub	-0.11**	-0.19***
	(-2.34)	(-2.98)
Guidance Count	-0.28***	-0.30***
	(-3.98)	(-3.68)
InstOwn	0.01	-0.32
	(0.05)	(-1.20)
Brokerage Size	-0.01	0.01
	(-0.27)	(0.32)
General Experience	-0.03	0.00
	(-0.62)	(0.04)
Firm Experience	0.04*	-0.04
	(1.76)	(-0.64)

Table 7 (Continued)		
Portfolio Size	-0.13**	-0.10
	(-2.08)	(-1.18)
Industry Count	0.07	0.17*
	(1.04)	(1.86)
Effort	0.14***	0.20***
	(2.60)	(3.03)
Horizon	0.25***	0.32***
	(5.60)	(6.54)
Firm FE	Yes	No
Analyst FE	Yes	No
Analyst-Firm FE	No	Yes
Year FE	Yes	Yes
Obs.	439,143	395,581
Adjusted R^2	0.10	0.07

Table 8: Forecast Dispersion and Stock Liquidity

This table presents the results for the effect of Google Translate on analyst forecast dispersion and stock liquidity. Column 1 examines analysts' forecast dispersion (*Dispersion*), calculated as the standard deviation of analysts' forecasts, scaled by stock price. Column 2 examines stock liquidity (*Liquidity*), calculated as the annual average of the daily Amihud illiquidity measure, multiplied by −1. Variable definitions are available in Appendix A. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1) Dispersion (%)	(2) Liquidity
Translate	-0.42***	0.02**
	(-4.54)	(2.16)
Size	-1.91***	0.10***
	(-19.45)	(13.38)
BTM	1.32***	-0.08***
	(7.32)	(-5.12)
Loss	0.68***	0.01
	(6.36)	(0.64)
ROA	-0.08***	0.00**
	(-12.09)	(2.30)
EarnVol	0.73	-0.21
	(0.38)	(-1.09)
RetVol	2.03***	0.21***
	(3.32)	(3.41)
Leverage	1.62***	-0.10***
	(5.70)	(-3.43)
R&D	-6.18***	-0.09
	(-4.71)	(-0.60)
ForeignSub	0.23***	-0.01***
	(5.30)	(-2.93)
Guidance Count	0.02	0.00
	(0.36)	(0.74)
InstOwn	-0.99***	0.23***
	(-4.66)	(9.54)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	45,958	47,751
Adjusted R^2	0.44	0.56

Online Appendix for "Breaking the Language Barriers? Machine Translation Technology and Analysts' Forecasts for Multinational Firms"

This online appendix contains the following materials:

A. Additional Figures

Figure OA1: Difference in Language Coverage between Google Translate and Microsoft Translator

This figure employs a Venn diagram to illustrate the relation between the two sets of foreign languages (i.e., languages other than English) supported by Google Translate and Microsoft as of June 30th, 2020. We collect the coverage details of Microsoft Translator following the same procedures as described in Section 3.1 of the main text.

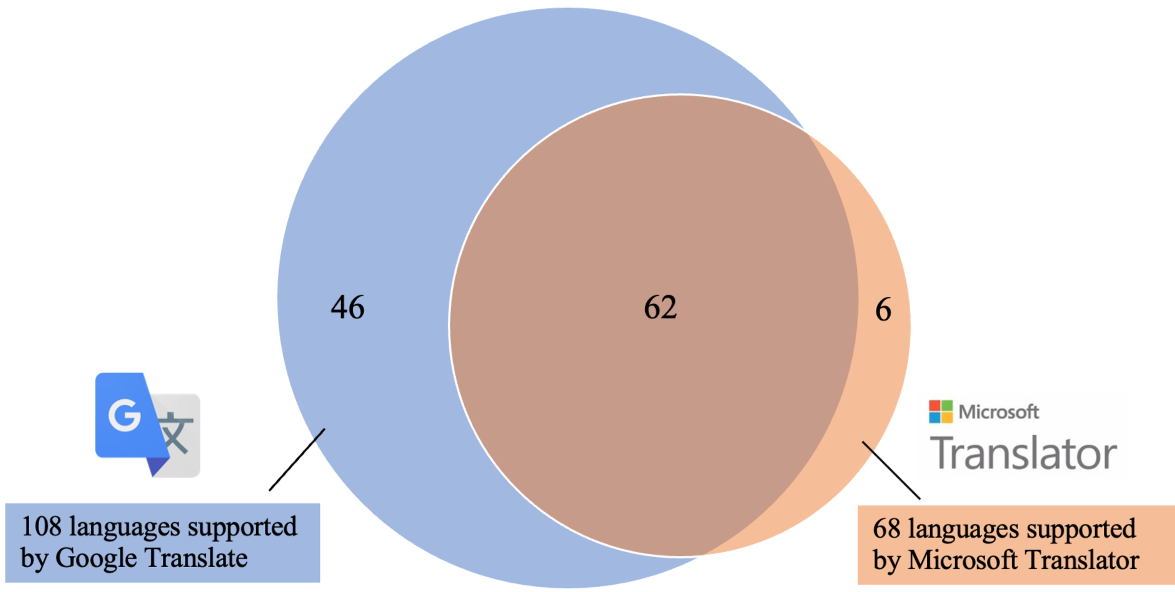


Figure OA2: Timing Difference in the Initial Rollout of the Languages Supported by Google Translate and Microsoft Translator

This figure plots the number of languages by the timing differences (in years) between the initial support date of Microsoft Translator and that of Google Translate for each language that is supported by both platforms (a total of 62 languages as indicated in Figure OA1). A positive (negative) value of the timing difference implies that Microsoft Translator is slower (faster) than Google Translate in introducing translation support for the specific language. We collect the coverage details of Microsoft Translator following the same procedures as described in Section 3.1 of the main text.

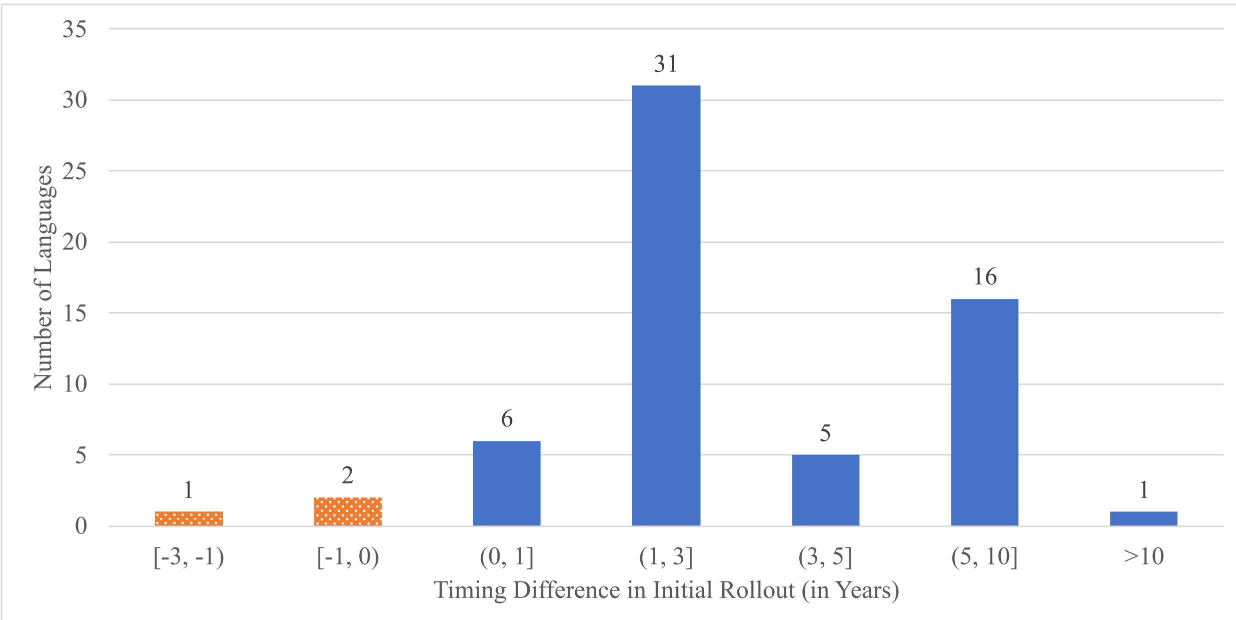
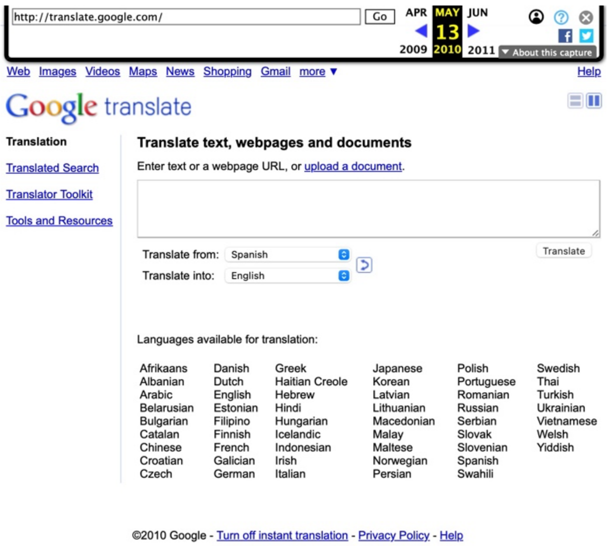
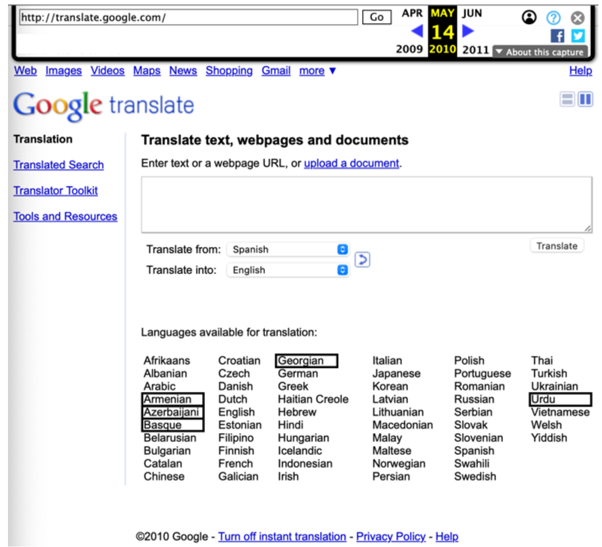


Figure OA3: Historical Web Pages of Google Translate from the Wayback Machine

This figure provides two examples of the historical pages of Google Translate from the Wayback machine to illustrate our process of identifying new languages supported by Google Translate over time. The page on the left is the snapshot on May 13, 2010 and the page on the right corresponds to the snapshot on May 14, 2010. By comparing the list of languages available for translation between the two pages, we identify five new languages (i.e., Armenian Azerbaijani, Basque, Georgian, Urdu) that receives the support in May 14 but not in May 13. We thus identify the initial support date of Google Translate for these five languages as May 14, 2010.



Page on May 13, 2010



Page on May 14, 2010

B. Examples of Foreign Information Discussed in Analyst Reports

This section presents examples of the discussion of foreign information and their corresponding information sources in analyst reports. We randomly select 20 firm-years among firms with significant foreign exposure and manually download 50 reports from Eikon for each firm-year. We browse these analyst reports to identify analysts' discussion of foreign information in the reports.

1. **McNulty, Khaykin, and Rajendran, Credit Suisse, 2010**
   Information source: Field trip

Our Industrials team took a recent field trip to China and met with Shanghai Electric. The company stated that the 2010 wind market has been below expectations, however despite this they are tripling their wind turbine output from 2009 to 2012 up to ~1,200 units. They continue to believe in the longer-term outlook of the industry.

2. **Volkmann, Gilloran, and Patel, Jefferies & Company, 2012**
   Information source: Colleague from the same firm

Our China capital goods colleague remains bearish. Analyst Julian Bu pointed out that Excavator machine hours, a leading indicator for machine sales, fell by 10% in September to 151 hours. As long as existing machines in the field are not busy, underlying demand for new machines should be low, and healthy shipments reported by producers may reflect channel stuffing or credit/promotional inducement.

3. **Jewsikow and Elias, Guggenheim Partners, 2023**
   Information source: Media report with foreign sources

SAIC [Shanghai Automotive Industry Corporation]/Audi [partnership] details. Details are a bit limited. So far, Audi has only confirmed they reached an agreement with SAIC, but reports (here) suggest that Audi will leverage SAIC's IM Motors platform for upcoming electric vehicles in China.

4. **Guo, Chee, and Fan, Jefferies & Company, 2013**
   Information source: Foreign research institution

The relaxation of China's one-child policy should boost near-term momentum for the baby care sector, but the impact on fundamentals are likely to be seen only in 2015. Based on research by China Academy of Social Sciences, this policy should result in 6.3% additional newborns. We expect this to result in a moderate benefit to the baby formula and diaper segments in the medium term.

5. **Suppiger, JMP Group, 2017**
   Information source: Consulting or research report

Figure 2. Enterprise Firewall Gartner Magic Quadrant [which classifies the firm's global competitors into four categories: challengers, leaders, niche players, and visionaries].

6. **Cooley, Stephens, 2018**
   Information source: Corporate announcement

Ruling Follows Recent German Court Victory. Recall at the end of October, BSX announced the District Court of Dusseldorf, Germany ruled EW's Sapien 3 Ultra transcatheter heart valve infringes the German portion of BSX's European Patent (EP) 2 949 292 B1 and; as a result, awarded BSX a preliminary injunction of the Sapien 3 Ultra valve in Germany.

C. Anecdotes of the Use of Google Translate by Financial Professionals

This section presents anecdotal evidence of the use of Google Translate by financial professionals. The excerpts are obtained from articles collected by ProQuest and our manual search.

1. Johnson (2016)

Here is where we consider ideas that others share, either voluntarily or according to law…We follow blogs, not only in English but also in German, French, Swedish, Norwegian and more (thank you, **Google Translate**). And there are plenty of interesting investors posting ideas on Twitter, as well as more fully formed ideas on investing websites such as Value Investors Club (VIC).

2. Root (2020)

After Tesla (ticker: TSLA) reported better-than-expected second-quarter numbers...Plaudits rained down, however, from an unusual place: Germany. Dr. Ferdinand Dudenhoeffer, director of the Center for Automotive Research in Duisburg, sang Tesla's praises in a LinkedIn post written in German. "While almost the entire [car business] sinks into losses in Q2, Tesla makes $327 million in profits in the operating business," he wrote, according to a **Google Translate** translation of the post.

3. Sultana (2021)

[I]nvestors themselves can use technology effectively. It is quite common even for finance professionals to use tools like **Google Translate** to read financial statements and announcements of foreign issuers.

4. Garrahan (2022)

Their [Credit Suisse] due diligence for their client is putting Bulgarian news articles into **Google Translate** and attempting to read them.

D. Additional Tables

Table OA1: Timing of the Initial Language Rollout by Google Translate: The Duration Model

This table reports the estimate of a Cox proportional-hazards model where the "failure event" is the rollout of Google Translate's support to translate a given foreign language into English. The sample consists of all the most commonly-used languages that we identify for countries where subsidiaries of U.S. firms are domiciled, with supported languages dropped from the sample once they were covered by Google Translate. The model is estimated starting from year 2005, when all languages were not supported by Google Translate. *Language Distance Index* is calculated using the method proposed by Joshi and Lahiri (2015) and captures the differences in features between each language and English. All the other independent variables are aggregated at the language-year level across all countries where the language is the most commonly-used. Both *GDP Per Capita* and *GDP Growth* are aggregated using the population of each country as the weight. All coefficients are multiplied by 1,000 for ease of presentation. A more positive coefficient suggests that the language is rolled out faster. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)
Number of U.S. Subsidiaries / Population	-0.45
	(-0.31)
Language Distance Index	-0.96**
	(-2.55)
Log(Population)	0.27***
	(7.23)
Log(GDP Per Capita)	0.33***
	(8.26)
GDP Growth	-0.01
	(-1.28)
Number of Inbound Tourists / Population	0.01***
	(5.01)
Obs.	405
pseudo R^2	0.17

Table OA2: Persistence of Firms' Number of Foreign Subsidiaries

This table examines the relation between the current value of a firm's number of subsidiaries and its future value in the next year. Column 1 adopts a simple linear regression without including any fixed effects. Column 2 adds firm and year fixed effects. Column 3 further includes our main treatment variable, *Translate*, as another independent variable. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)
	ForeignSub_{t+1}	ForeignSub_{t+1}	ForeignSub_{t+1}
ForeignSub	0.98***	0.74***	0.74***
	(793.75)	(87.73)	(84.38)
Translate			0.01
			(0.67)
Firm FE	No	Yes	Yes
Year FE	No	Yes	Yes
Obs.	43,825	43,162	43,162
Adjusted R^2	0.94	0.95	0.95

Table OA3: Robustness: Alternative Treatment Variables

This table evaluates the robustness of our main results (as in Column 3 of Table 3 in the main text) using a set of alternative measures of the treatment variable, *Translate*. Column 1 uses an indicator that equals one if *Translate Percentage* is greater than 0%. Column 2 uses an indicator that equals one if *Translate Percentage* is 20% or greater. Column 3 examines the log transformation of the percentage of foreign subsidiaries domiciled in countries where the most commonly-used language is supported by Google Translate. Column 4 uses an indicator (*Translate Percentage* >= 10%) defined based on a firm's constant set of subsidiaries in the first year of its appearance in our sample period before the initial launch of Google Translate. Column 5 uses an indicator that equals one if the firm derives at least 10% of its revenue from countries where the most commonly-used language is supported by Google Translate. Column 6 uses an indicator that equals one if the firm operates at least 10% of foreign subsidiaries in countries where *any of their official or business languages* (if more than one) are supported by Google Translate. We treat foreign countries with English as one of their official or business languages (e.g., India) as equivalent to English-speaking countries and exclude subsidiaries in those countries from the calculation of treated subsidiaries. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Error	Error	Error	Error	Error	Error
Translate (>0%)	-0.44***					
	(-2.99)					
Translate (≥20%)		-0.52***				
		(-3.66)				
Log(1+Translate Percentage)			-0.16***			
			(-3.80)			
Translate (Using historical subsidiaries)				-0.60***		
				(-3.62)		
Translate (Using segment sales as weight)					-0.37**	
					(-2.41)	
Translate (Any official/business languages)						-0.42***
						(-2.93)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	49,450	49,450	49,450	40,566	40,609	49,450
Adjusted R^2	0.39	0.39	0.39	0.38	0.40	0.39

Table OA4: Robustness: First Batch of Supported Languages vs. Other Batches

This table evaluates the soundness of our staggered Difference-in-Differences design by decomposing the treatment effect into two separate components based on whether the language is one of the first batch of languages supported by Google Translate in April of 2006 (see footnote 6 of the main text). Panel A presents the average value of these two components and tests the difference. Panel B presents the estimates of the treatment effect separately for these two components using the same specification as in Column 3 of Table 3 in the main text. Column 1 uses two indicators defined using the 10% threshold. Column 2 uses the log transformation of the continuous treatment variables. Both columns also present the test results of the statistical difference between the two coefficients. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

Panel A: Summary Statistics

	(1) N	(2) Mean
Translate Percentage (First Batch)	49,450	8.19
Translate Percentage (Other Batches)	49,450	5.22
Difference (First Batch - Other Batches)		2.97***

Panel B: Treatment Effect by the First Batch vs. Other Batches

	(1) Error	(2) Error
Translate (First Batch>=10%)	-0.31**	
	(-2.08)	
Translate (Other Batches>=10%)	-0.27*	
	(-1.94)	
Log(1+Translate Percentage) (First Batch)		-0.09*
		(-1.67)
Log(1+Translate Percentage) (Other Batches)		-0.14***
		(-2.62)
Controls	Yes	Yes
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	49,450	49,450
Adjusted R^2	0.39	0.39
Coefficient Difference:		
First Batch **LESS** Other Batches	-0.04	0.05
p-value	0.87	0.53

Table OA5: Robustness: Alternative Difference-in-Differences Estimates

This table evaluates the robustness of our main results (as in Column 3 of Table 3 in the main text) using alternative Difference-in-Differences (DiD) estimators introduced in recent econometric literature. Column 1 computes the DiD estimator introduced in de Chaisemartin and D'Haultfoeuille (2020a), which is designed for a setting where the treatment is not necessarily an absorbing state (i.e., the treatment status can turn on and then off at different times). Column 2 adopts the imputation estimator introduced in Borusyak et al. (2021), which first infers the hypothetical outcome without any treatment for each treated unit using only not-yet-treated units and then estimates the treatment effect by comparing the realized outcome with the hypothetical outcome. Column 3 adopts the Callaway and Sant'Anna estimator introduced in Callaway and Sant'Anna (2021), which similarly uses not-yet-treated units as the control group but differs from Borusyak et al. (2021) in the number of pre-treatment periods that are used to model the pre-treatment outcome. More nuances and differences regarding these estimators can be found in Roth et al. (2023). Column 4 adopts the stacked-regression approach suggested by Baker et al. (2022). We group our treatment firms into different cohorts based on their initial treatment year and then keep their observations in the four-year period around the treatment year. For each cohort of treatment firms, we include control firms that are not yet treated in the respective four-year period. The sample size varies across the four columns depending on the specification of each model. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)	(4)
	Error	Error	Error	Error
Method	de Chaisemartin and D'Haultfoeuille (2020a)	Borusyak et al. (2021)	Callaway and Sant'Anna (2021)	Baker et al. (2022)
Translate	-0.46**	-0.65***	-1.20**	-1.39***
	(-1.97)	(-3.78)	(-2.31)	(-4.73)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	No
Year FE	Yes	Yes	Yes	No
Cohort-Firm FE	No	No	No	Yes
Cohort-Year FE	No	No	No	Yes
Obs.	31,115	46,141	44,401	69,281
Adjusted R^2	-	-	-	0.44

Table OA6: Robustness: Alternative Setting Using Microsoft Translator

This table evaluates the robustness of our main results (as in Column 3 of Table 3 in the main text) using the staggered rollout of Microsoft Translator's support to translate foreign languages into English as an alternative setting. Panel A presents the summary statistics for the two treatment variables defined based on the staggered rollout of Microsoft Translator. *Translate (Microsoft)* is an indicator that equals one if the firm operates at least 10% of subsidiaries in countries where the most commonly-used language is supported by Microsoft Translator in a given year, zero otherwise. *Translate Percentage (Microsoft)* represents the value of the percentage of treated subsidiaries. As presented in Figures OA1 and OA2, Microsoft Translators covers a smaller set of languages at a slower pace compared with Google Translate. As a result, the average value of *Translate Percentage (Microsoft)* and *Translate Percentage (Microsoft)* are both smaller than their counterparts defined based on the staggered rollout of Google Translate (see Table 2 in the main text). Panel B presents the results using the same specification as in Column 3 of Table 3 in the main text with these alternative test variables. Column 1 uses the indicator variable defined based on Microsoft Translator. Column 2 uses the log transformation of the continuous treatment variable. Columns 3 and 4 further include the corresponding treatment variables defined using Google Translate for "horse race" tests. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

Panel A: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Mean	Std.	p25	p50	p75
Translate (Microsoft)	49,450	0.28	0.45	0.00	0.00	1.00
Translate Percentage (Microsoft)	49,450	12.02	21.68	0.00	0.00	15.87

Panel B: Treatment Effect of Microsoft Translator vs. Google Translate

	(1)	(2)	(3)	(4)
	Error	Error	Error	Error
Translate (Microsoft)	-0.35**		0.18	
	(-2.34)		(0.76)	
Log(1+Translate Percentage) (Microsoft)		-0.13***		0.04
		(-2.86)		(0.57)
Translate (Google)			-0.62***	
			(-2.94)	
Log(1+Translate Percentage) (Google)				-0.20***
				(-2.69)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs.	49,450	49,450	49,450	49,450
Adjusted R^2	0.39	0.39	0.39	0.39

Table OA7: Analyst-Level Results Using Continuous Treatment Variables

This table presents the results of the tests as in Panel B of Table 4 in the main text using the log transformation of the continuous treatment variables instead of the indicators. Standard errors are clustered by firm and analyst. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Error	Error	Error	Error	Error	Error
Log(1+Translate Perc.)	-0.09***			-0.08***		
	(-3.89)			(-3.51)		
Log(1+Translate Perc.) (Non-Proficient)		-0.08***			-0.08***	
		(-3.62)			(-3.50)	
Log(1+Translate Perc.) (Proficient)		-0.04***			-0.03	
		(-2.60)			(-1.31)	
Log(1+Translate Perc.) (Non-Covered)			-0.08***			-0.07***
			(-4.31)			(-3.73)
Log(1+Translate Perc.) (Covered)			-0.03**			-0.03*
			(-2.36)			(-1.80)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	No	No	No
Analyst FE	Yes	Yes	Yes	No	No	No
Firm-Analyst FE	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	524,078	496,978	524,078	478,226	459,297	478,226
Adjusted R^2	0.43	0.43	0.43	0.44	0.44	0.44
Coefficient Difference:						
Non-Proficient **LESS** Proficient		-0.04*			-0.05	
p-value		0.06			0.18	
Non-Covered **LESS** Covered			-0.05***			-0.04***
p-value			0.00			0.00

Table OA8: Geographic Economic Conditions Weighted by Firms' Subsidiary Count and Analysts' Forecasts

This table presents the results of the tests as in Table 6 in the main text using an alternative set of GDP growth variables aggregated by the relative weight of the firm's number of subsidiaries in each country instead of its estimated segment sales. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1)	(2)	(3)
	ROA (%)	Forecast-Implied ROA (%)	Forecast-Implied ROA (%)
GDP Growth (Supported)	0.41***	0.20***	-0.03
	(3.88)	(2.94)	(-0.33)
GDP Growth (Supported) × Translate			0.30***
			(2.94)
GDP Growth (Non-Supported)	0.26***	0.13**	0.08
	(2.86)	(2.16)	(1.21)
GDP Growth (Non-Supported) × Translate			0.04
			(0.50)
Translate			-0.15
			(-0.71)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs.	23,897	23,897	23,897
Adjusted R^2	0.55	0.71	0.71

Table OA9: Other Information Sources

This table presents the results of the effect of Google Translate on the output of other information sources. Column 1 examines the total number of 8-Ks filed by the company in a given year. Column 2 tests the total word count of the firm's 10-K filing. Column 3 examines the total number of management earnings guidance issued by the firm in a given year. Column 4 focuses on the accuracy of the last management earnings guidance issued before the annual earnings announcement, which is calculated in a similar way to the analyst forecast error variable (*Error*) in our main text. Column 5 examines the total number of news articles from the Wall Street Journal (WSJ), Barron's, and MarketWatch. Column 6 tests the total number of news articles from Dow Jones Newswires. All count variables are log-transformed in the regressions. Standard errors are clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	Corporate Filings		Earnings Guidance		Media News	
	(1)	(2)	(3)	(4)	(5)	(6)
	8-K Count	10-K Length	Guidance Count	Guidance Error	WSJ etc. Count	DJ Newswires Count
Translate	0.00	-0.01	0.09***	-0.48	0.02	-0.03*
	(0.11)	(-1.27)	(5.55)	(-1.43)	(1.191)	(-1.85)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	49,450	49,395	49,450	15,462	43,874	43,874
Adjusted R^2	0.75	0.57	0.65	0.72	0.72	0.76

Table OA10: Analysts' Coverage Decisions

This table presents the results of the effect of Google Translate on analysts' coverage decisions. We adopt a linear probability model in our regressions. The dependent variable, *Coverage*, is an indicator equal to one if the analyst issues at least one earnings forecast for the firm in the year, and zero otherwise. Column 1 includes firm, analyst, and year fixed effects. Column 2 includes analyst-firm pair and year fixed effects. Standard errors are clustered by firm and analyst. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level using two-tailed tests, respectively.

	(1) Coverage	(2) Coverage
Translate	-0.01	-0.01
	(-1.05)	(-1.19)
Size	0.08***	0.09***
	(30.32)	(30.55)
BTM	0.06***	0.07***
	(15.52)	(16.92)
Loss	0.00	0.00
	(0.98)	(1.26)
ROA	-0.00***	-0.00***
	(-5.35)	(-5.67)
EarnVol	-0.20***	-0.25***
	(-2.70)	(-3.05)
RetVol	-0.08***	-0.07***
	(-3.60)	(-3.46)
Leverage	0.09***	0.10***
	(8.02)	(8.06)
R&D	0.01	0.04
	(0.25)	(0.90)
ForeignSub	-0.00	-0.00
	(-1.55)	(-0.65)
Guidance Count	0.01***	0.01***
	(4.99)	(5.49)
InstOwn	0.06***	0.06***
	(6.92)	(6.56)
Firm FE	Yes	No
Analyst FE	Yes	No
Analyst-Firm FE	No	Yes
Year FE	Yes	Yes
Obs.	1,038,792	1,021,426
Adjusted R^2	0.14	0.30

References for Online Appendix

Baker, A., Larcker, D., and Wang, C. 2022. How Much Should We Trust Staggered Difference-In-Differences Estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Borusyak, K., Jaravel, X., Spiess, J., 2021. Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv preprint. arXiv:2108.12419

Callaway, B., and Sant'Anna, P. H. C. 2021. Difference-In-Differences with Multiple Time Periods. *Journal of Econometrics*, *225*(2), 200–230.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110 (9): 2964-96.

Garrahan. D. 2022. Credit Suisse: What Next for the Crisis-Hit Bank? https://www.ft.com/video/44f4b910-80f3-4728-a846-89e768b8a099

Johnson, S. 2016. Six Ways to Find New Investments: Stock Picking. *The Australian Financial Review.*

Joshi, A., and Lahiri, N. 2015. Language Friction and Partner Selection in Cross-Border R&D Alliance Formation. *Journal of International Business Studies*, *46*(2), 123–152.

Root, A. 2020. VW Might Have a Case of Tesla Envy. Here's Why. *Barron's*.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. 2023. What's Trending in Difference-In-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, *235*(2), 2218–2244.

Sultana, N. 2021. The Language Barrier Plaguing India's Stock Market. *Mint*. https://www.livemint.com/market/stock-market-news/why-english-only-markets-breed-inequity-11629994292128.html