DOI 10.7603/s40570-016-0007-4

Volume 18, Number 3 – September 2016
China Accounting and Finance Review
中国会计与财务研究
2016年9月 第18卷 第3期

Public Information, Private Information, and Financial Analysts' Forecasts*

Yutao Wang [1]

Received 18th of December 2014 Accepted 18th of April 2016

Abstract

This paper investigates the joint impact of public and private information on analysts' forecasting behaviour. Information disclosure determines the quality and quantity of public information. The geographical distance between listed firms' headquarters and brokerage houses determines the difficulty and costs of analysts obtaining private information, which affect the quality and quantity of private information. Firstly, this paper finds results consistent with prior literature that the higher the disclosure level of the firm and the shorter the geographical distance between firms and brokerage headquarters, the more frequently analysts make earnings forecasts and the lower the forecast errors. Furthermore, this paper examines the joint effects of information disclosure and geographical distance on analysts' forecasts and finds that information disclosure quality imposes a stronger impact on analysts' forecast frequency and forecast error than distance. Moreover, the effect of geographical distance on the frequency and errors of analysts' forecasts is manifested only in the case of higher information disclosure quality. These results imply that both public and private information affect analysts' forecasts but that their effects are asymmetric. Public information has a significantly stronger impact on analysts' forecasts than private information, and the effect of private information depends on a better public information environment. This paper has significant value to public policy regulators, market participants, stakeholders, and academics in China and can help to increase their understanding of the impact of public and private information on analysts' forecasts.

Keywords: Public Information, Private Information, Analysts' Forecasts, Information Disclosure, Geographical Distance

---

* This study was supported by the National Natural Science Foundation of China (Project 71102124), grants from the Beijing Municipal Commission of Education's "Pilot Reform of Accounting Discipline Clustering", and the Fundamental Research Funds for the Central Universities.
[1] Yutao Wang, School of Accountancy, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing, P. R. China 100081; email: wyt_cufe@163.com; wangyutao@cufe.edu.cn.

Published online: 29 August 2016

I. Introduction

As its securities market develops, China's Ministry of Finance continues to deepen its fiscal and tax reforms. Since a major change in China's accounting standards in 2006, the pace of reform has accelerated. Accounting standards and taxation system reforms have prompted tremendous changes in China's securities market information environment, and these changes have inevitably significantly impacted the behaviour of accounting information providers and information intermediaries. Utilising public and private information, financial analysts, as information intermediaries, make earnings forecasts, stock ratings and recommendations, and so forth in order to provide decision-relevant information to capital market participants. Thus financial analysts play an important role in the mitigation of information asymmetry between listed firms and outside investors as well as in the improvement of capital market efficiency. On the basis of the market environment of the past 10 years, this paper attempts to examine generally how public and private information jointly impact the behaviour of financial analysts. The examination herein is significantly different from prior studies in that it studies the two types of information in the same model.

Prior studies have examined the impact of public information on analysts' forecasts, including financial statements, management earnings forecasts (Baldwin, 1984; Hodder *et al.,* 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.,* 2006), and comprehensive information disclosure (Hope, 2003a; Bai, 2009; Fang, 2007). In recent years, researchers such as Green *et al.* (2014), Mayew *et al.* (2013), and Malloy (2005) have measured the differences in private information obtained by financial analysts by analysing their behaviour and have investigated the impact of private information on financial analysts' forecasts. This article is different from these studies in that it attempts to comprehensively investigate the joint impact of public and private information on financial analysts' forecasts. Using the information disclosure index to measure the quantity and quality of public information, and geographical distance to measure the quantity and quality of private information, this paper examines the joint impact of such information on the analysts' forecasting behaviour and finds that higher levels of information disclosure and shorter geographical distance lead to lower analysts' forecast errors and more frequent forecast reports. These findings are consistent with the previous literature. This paper also finds, through theoretical discussion and empirical analysis, that under the combined effect of public and private information on analysts' forecasts, information disclosure has a significantly greater influence on analysts' forecasts than geographical distance. In addition, when the full data sample is divided into two groups according to the level of information disclosure, geographical distance has a significant impact on forecast errors and frequency only in the higher group. These results imply that compared to private information, public information has a greater impact on analysts' forecasting behaviour.

This study makes two contributions to the literature. First, it complements previous research results. Unlike previous separate investigations of the impact of public and private information on analysts' forecasts (e.g. Baldwin, 1984; Hodder *et al.,* 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.,* 2006; Bai, 2009; Fang, 2007), this paper focuses on the combined effect of the two, yet it validates previous research results. Second, its research findings have practical significance. Financial analysts play an important intermediary role in capital markets, reducing the information asymmetry between investors and listed companies. With the continuing reform of China's securities market and the gradual improvement of the public information environment, the impact of the combined effect of public and private information on financial analysts' forecasts has become an important issue of concern to all capital market participants. This study provides an answer to this important question. Our research findings will also help investors better understand and predict the behaviour of financial analysts and enable them to make rational decisions.

II. Literature Review

Financial analysts play an important role in information processing and transfer in capital markets. Due to the unavoidable information asymmetry between investors and listed companies, professional analysts evaluate situations and provide research reports on the basis of public and private information, thereby reducing information asymmetry. Financial analysts simultaneously observe public information, including periodic reports disclosed by companies, management earnings forecasts, and audit reports, as well as some specific types of managerial behaviour (e.g. manipulating profits, investments, and financing behaviour). Such information can be observed publicly by capital market participants and financial analysts. The prior literature on public information, such as Pankoff and Virgil (1970), McEwen and Hunton (1999), and Vergoossen (1993), focuses on how the accuracy of financial analysts' forecasts relates to accounting numbers. Subsequent studies have started to emphasise the specific structure and quality of financial statements and to examine whether and how specific contents and different qualities have influence over analysts' earnings forecasts: for example, whether the content of segment reports has an impact on analysts' earnings forecasts (Baldwin, 1984) and whether earnings management, audit quality, and professionalisation (Li and Jia, 2009) affect analysts' earnings forecasts (Behn *et al.,* 2008; Burgstahler and Eames, 2003; Payne, 2008; Robertson, 1988; Yu, 2008). In addition to regularly reported public information, management also disclose future corporate earnings forecasts. Hence, some researchers have examined whether and how these future corporate earnings forecasts affect financial analysts' forecasts (Libby *et al.,* 2006; Wang *et al.,* 2015; Wang and Wang, 2012). Recent studies have focused on the integrated information environment, such as the information disclosure index and accounting standards, investigating how the external environment influences analysts' forecasting behaviour. O'Brien and Bhushan (1990), Lang and Lundholm (1996), and Bai (2009) find that the higher a company's level of information disclosure, the more analysts are going to follow the company. Lang and Lundholm (1996) and Bai (2009) further find a higher accuracy and lower dispersion of analysts' forecasts among companies with higher disclosure levels. Hope (2003b), through an examination of the relationships among the levels of information disclosure, analysts' forecast errors, and forecast dispersions, finds that the level of information disclosure is negatively correlated with forecast error and dispersion. The above studies consistently conclude that higher levels of information disclosure lead to more accurate forecasts and lower forecast dispersion. These studies imply that either public information or the information environment has a significant impact on analysts' forecasting behaviour.

Public information is what can be observed from, and what is disclosed by, companies. However, the interpretations and utilisation of such public information by analysts can be defined as private information. Hu and Lin (2005), through the survey research method, find that the information obtained by financial analysts in China can be classified into four categories: publicly disclosed information, research information, collateral information, and informal information. This means that research information, collateral information, and informal information can be defined as private information. Since private information is difficult to observe and measure, there are fewer studies on the impact of private information on analysts' forecasting behaviour. Chen and Jiang (2006) investigate analysts' forecasting processes and find that analysts rely more on private than public information to make earnings forecasts. Green *et al.* (2014) find that the more private information financial analysts obtain, the higher the accuracy of analysts' forecasts. Mayew *et al.* (2013) investigate the impact of private information obtained via conference calls on analysts' forecast accuracy and find that financial analysts participating in these conference calls issue timelier and more accurate forecast reports. Hence, they determine that participating in conference calls can be useful for identifying whether financial analysts have private information advantages.

Analysts rely on both public and private information to make forecasts. Therefore, the first question is: What is the relationship between private information and public information? Verrecchia (1982) and Bushman (1991) suggest that in a perfect and competitive market, information disclosure reduces information asymmetry and increases investor welfare such that public information and private information are substitutes for each other. However, McNichols and Trueman (1994) and Lundholm (1988) believe that more public information will lead to an increase in private information, resulting in no change in information asymmetry, because some market participants make use of public information as resources to obtain private information; therefore, public and private information are complementary to each other. Wang (2006) examines the price and liquidity effects of information disclosure in China and finds that public and private information are complementary to each other in the Chinese stock market.

Therefore, the second question is: How do public and private information jointly impact financial analysts' forecasts? The results of prior studies are inconsistent. Gintschel and Markov (2004) believe that if public information disclosure rules are effective, private information will be reduced, analysts' forecast accuracy will decrease, and forecast dispersions will increase. De Jong and Apilado (2009) find that greater information disclosure can have the effect of reducing or removing private information, which helps improve the accuracy of analysts' forecasts. Although the above researchers consistently agree that a higher level of public information disclosure leads to a reduction in private information, their conclusions about the impact on forecast accuracy are inconsistent. The inconsistencies in the findings prompted this study to investigate how public and private information jointly influence analysts' forecasts in China.

III. Research Hypotheses and Regression Model

3.1 Theoretical Model and Research Hypotheses

Earlier researchers have analysed the influence of public information disclosure (Hope, 2003b; Langberg and Sivaramakrishnan, 2008; Libby *et al.,* 2006; Bai, 2009) and private information (Chen and Jiang, 2006; Mayew *et al.,* 2013) on analysts' forecasts. Compared with the prior research above, this paper focuses on how securities analysts predict earnings and make decisions when simultaneously facing public and private information and discusses the joint influence of public and private information on securities analysts' forecasts following the theoretical framework of Barronet *et al.* (1998).

N represents the number of financial analysts following a company. The earnings forecasts of each financial analyst are defined as y. Assume that y is normally distributed with mean $\bar{y}$ and variance $1/h$ (h is the precision). Generally, analysts can gain two kinds of information: common (public) information and idiosyncratic (private) information. Further assume that public information is equal to the actual earnings during a previous period (with mean $\bar{y}$ and precision h). Private information refers to the information some analysts can observe but others cannot; it is defined as $z_i = y + \varepsilon_i$, where ε_i is independent and identically distributed (i.i.d.) with mean zero and variance $1/s_i$. Different analysts have access to private information of different quality. Therefore, $s_i s$ varies among analysts.

It is assumed that each analyst's earnings forecasts reflect his or her best estimate of earnings based on the available public and/or private information, and thus the analyst's conditional expectation is a weighted average of the public and private information, with the precision of each piece of information as the weight. The analyst's earnings forecasts, denoted by μ_i, are defined as

$$\mu_i \equiv E[y|z_i] = \frac{h\bar{y} + s_i z_i}{h + s_i} \tag{1}$$

Equation (1) shows that financial analysts' best estimate of earnings is the weighted average of the mean actual earnings during the past period and earnings forecasts based on private information. Equation (1) reveals the decision-making procedure of analysts' earnings forecasting. First, if there is no private information, the analyst will predict current earnings only on the basis of the average actual earnings during the past period. Second, financial analysts obtain private information from two sources: their understanding and analysis of public information and their own investigations to gain such information, such as on-site visits. Generally, both public and private information are flexible, so the decision-making processes of financial analysts are based on all the information they obtain, taking the precision of each piece as the weight.

From Equation (1), the average forecast value of each analyst is calculated as

$$\mu \equiv \frac{1}{N}\sum_{i=1}^{N}\mu_i = \sum_{i=1}^{N}\frac{h\bar{y} + s_i z_i}{h + s_i} \tag{2}$$

Thus, the observed dispersion in forecasts, denoted by *d*, is defined as the sample variance of μ_i: that is,

$$d \equiv \frac{1}{N-1}\sum_{i=1}^{N}(\mu_i - \mu)^2 \tag{3}$$

Generally, the non-random dispersion measure of all analysts, denoted by *D*, is simply the unconditional expectation of *d*: that is,

$$D \equiv E[d] = \frac{1}{N-1}E\left[\sum_{i=1}^{N}(\mu_i - \mu)^2\right] = \frac{1}{N-1}\sum_{i=1}^{N}Var(\mu_i - \mu) \tag{4}$$

Equation (4) shows that the forecast dispersions of all financial analysts is the average variance of the difference between the earnings forecast (μ_i) of individual analysts and the mean value of all analysts' forecasts. Because both μ_i and μ contain common information *y* with mean $\bar{y}$, *D* is further influenced by the quality of private information gained by each individual analyst. Hence, from the perspective of the entire information environment, *D* can be viewed as the average quality of private information obtained by financial analysts.

Furthermore, the analysts' forecast error, denoted by e, is the actual earnings (y) minus the mean forecast (μ): that is,

$$e \equiv y - \mu \tag{5}$$

The forecast error for all analysts, denoted by SE, is the expectation of the squared error in the mean forecast: that is,

$$SE \equiv E(e^2) = E[(y-\mu)^2] \tag{6}$$

What is discussed above is the decision-making procedure of analysts' forecasts based on the public and private information they obtain. Then, the influence of the information environment on analysts' earnings forecasts is analysed. Two general properties of the analysts' information environment are introduced. The first property is the level of residual uncertainty for individual analysts, which is defined as follows:

$$V_i \equiv E[(y-\mu_i)^2] = \frac{1}{h+s_i} \tag{7}$$

Equation (7) shows that the expected variance of the difference between actual earnings and the earnings forecast predicted by financial analysts is conditional upon the information available to them and is equal to the derivative of the sum of the precision of public and private information. The overall uncertainty level V is defined as simply the individual uncertainty averaged over N analysts: that is,

$$V \equiv \frac{1}{N}\sum_{i=1}^{N} V_i = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h+s_i} \tag{8}$$

Equation (8) shows that the overall uncertainty level is the variance of the difference between analysts' predicted earnings and actual earnings. Then, the average value for all analysts is calculated. In other words, the overall uncertainty level is a function of the accuracy of the public and private information available to all financial analysts.

The second measure introduced is the degree of consensus among analysts. This is defined as the average covariance of the difference between earnings forecast and actual earnings among all analysts.

$$C \equiv \frac{1}{N}\sum_{i=1}^{N} C_i = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j\neq i}^{N} Cov(y-\mu_i, y-\mu_j)\right] \tag{9}$$

One can interpret the average covariance, *C*, as common uncertainty – that is, the uncertainty shared by all analysts attributable to their reliance on imprecise public information.

According to Equation (9), the forecast dispersion and (expected squared) error in the mean forecast are *D* and *SE*, respectively; uncertainty is *V*; and average covariance is *C*. The equation is stated as follows:

$$V = \frac{1}{N}\sum_{i=1}^{N} V_i = \frac{1}{N}\sum_{i=1}^{N} E[(y-\mu_i)^2] = C + D \tag{10}$$

$$SE = C + \frac{D}{N} \tag{11}$$

V is the overall uncertainty for all analysts when they make earnings forecasts. It equals the average variance of the difference between actual earnings and earnings forecasts. *D* is the average variance of the difference between the earnings forecast of analyst *i* and the mean earnings forecasts of all analysts. *C* is the average covariance of the difference between the earnings forecast of analyst *i* and actual earnings. Hence, Equation (10) implies that the overall uncertainty of analysts' forecasts is the sum of the variance of the difference in earnings forecast between an individual analyst *i* and all analysts following the firm and the covariance of the difference between an individual analyst's earnings forecast and actual earnings.

In Equation (11), *SE* is forecast error, which equals the sum of the covariance of the difference between an individual analyst's earnings forecast and actual earnings (*C*), and part of variance of the difference between the earnings forecast from individual analyst *i* and the mean earnings forecast from all analysts (i.e. *D*/N). According to the analysis above, *D* is influenced mainly by the quality (or uncertainty) of the private information obtained by each financial analyst. Private information produces the difference between the earnings forecast from a specific analyst and the mean earnings forecast from all analysts (consensus forecast). The more private information an analyst obtains, the larger the difference. Meanwhile, *C* consistently shows the extent of the difference between earnings forecast and the actual value among analysts. *C* is mainly influenced by the public information obtained by financial analysts. Public information can be observed by all analysts simultaneously, which helps them make consistent earnings forecasts. Therefore, a difference between the consistent earnings forecast made by all analysts and the actual value of earnings only occurs when analysts make forecasts on the basis of public information. All in all, analysts' forecast error is dependent upon public information (quantity or quality) and some of their private information (quantity and quality).

Following the theoretical model above, public information is denoted by *PubIn and* private information by *PrvIinf*, and the regression model is built according to Equation (11) as follows:

$$SE = \alpha + \beta * PubInf + \gamma * PrvInf + \varepsilon \tag{12}$$

Compared with Equation (11), it is inferred from Equation (12) that $\beta = 1$ and $\gamma = 1/N$. Because N>1,[2] it is reasonable to predict that $\beta > \gamma$.[3] This means that the influence of public information is greater than idiosyncratic, or private, information. The first hypothesis in this paper is as follows:

H1: The impact of public information on analysts' forecast errors is greater than that of private information.

According to H1, if the impact of public information on analysts' forecast errors is greater than that of private information, the second related question is whether the effect of private information is dependent upon a better public information environment. In the theoretical analysis above, private information is defined as $z_i = y + \varepsilon_i$. This means that private information is related to the public information obtained by analysts during the past period (i.e. y) and specific idiosyncratic information (i.e. ε_i). The latter is dependent upon the information process of financial analysts, such as understanding public information, or his or her efforts and actions. Public information is the most available information an analyst can obtain. Firstly, analysts need to explain, understand, and process the common information, which will help them to gain more effective private information. McNichols and Trueman (1994) and Lundholm (1988) argue that public information can lead to increases in private information because some market participants use public information as a source of private information. The availability of a large quantity of high-quality information improves the reliability of analysts' public information, which saves analysts' time and effort, enabling them to concentrate on analysis and obtaining private information. Obtaining a higher quantity or quality of private information can help analysts improve the accuracy of their forecasts. So, it is reasonable to predict that the influence of private information on analysts' forecasts is greater in a better public information environment. The second hypothesis is as follows:

---

[2] In practice, N=1 is also possible, meaning that some companies are in a particular year covered by only one analyst. If only one analyst issues a forecast report, the forecast dispersion cannot be calculated. In this paper, Equations (3) and (4) cannot be established. Hence, N > 1 is a necessary precondition.

[3] The purpose of this study is to investigate whether the impact of public information or private information is stronger, as well as their joint effect. So, the theoretical model is limited to inferring the relative effects of β and γ. Therefore, the main empirical test of this paper is to evaluate and analyse whether β is greater than γ, not to test if $\beta = 1$ and $\gamma = 1 / N$.

H2: The impact of private information on analysts' forecasts is greater in a better public information environment.

3.2 Research Design and Linear Model

Before discussing the definitions of key variables and the regression model, two caveats have to be noted. First, this study mainly focuses on the differences in terms of impact on analysts' forecast error between public and private information. Another main test is to find out whether and how public and private information impact the number of forecast reports issued by financial analysts (i.e. forecast frequency), which is additional evidence for verifying the hypotheses. Second, before testing hypotheses H1 and H2, the separate influences of public and private information on forecast error and frequency have to be tested. It is very important to test H1 and H2 because if the separate effects of public and private information do not exist, testing the joint effects would not be necessary.

Since it is very difficult to observe the efforts or actions of financial analysts to gain private information, private information is measured as the geographical distance between the headquarters of the brokerage house the analyst works with and the listed firm followed by the same analyst. Recent studies have begun to focus on the relationship between this geographical distance, or location, and the behaviour of economic participants: for example, the impact of geographical distance on bank loans (Agarwal and Hauswald, 2010; García and Norli, 2012; Hauswald and Marquez, 2006; Petersen and Rajan, 2002), the relationship between stock returns and geographical factors (García and Norli, 2012; Kim *et al.,* 2012), the impact of location on corporate actions such as the dividend payout policy (John *et al.,* 2011), the impact of geographical factors on the performance of mutual funds (Teo, 2009), and the impact of location on auditor independence (DeFond *et al.,* 2008). All such research is based on the same theoretical logic that geographical location, or proximity, increases the information advantage of economic participants. These economic participants can gain private information at a lower cost, thereby decreasing transaction costs. From the perspective of geographical location, some researchers have investigated the relationship between location and analysts' forecasts. For example, Bae *et al.* (2008) and Wang *et al.* (2010) investigate the differences in the accuracy of forecasts on local firms between local and foreign analysts. They find that local analysts generate more accurate forecasts than foreign analysts because the former have an information advantage due to their closer proximity to local firms. Hence, this paper also adopts the geographical distance between the headquarters of analysts and the listed firm as the measurement of private information.

A comprehensive indicator is used to measure public information – the information disclosure index of the listed firm, which is a disclosure rating (*DC_rating*). The Shenzhen Stock Exchange evaluates the disclosure quality of each listed firm and publishes the index annually. On the basis of this measure, some papers, such as Fang (2007) and Bai (2009), have investigated the relationship between information transparency and securities analysts' forecasts. They find that the greater the transparency of information, the higher the accuracy of analysts' forecasts. Hence, the geographical distance between the analyst's headquarters and the listed firm is used as a proxy for private information and the disclosure rating for public information. The regression models are set up as follows:

$$AnlBeh_{i,j,t} = \alpha + \beta_1 DC\_rating_{i,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (13)$$

$$AnlBeh_{i,j,t} = \alpha + \beta_1 Distance_{i,j,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (14)$$

$$AnlBeh_{i,j,t} = \alpha + \beta_1 DC\_rating_{i,t} + \beta_2 Distance_{i,j,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (15)$$

Regression models (13) and (14) examine the separate impact of public and private information, and model (15) examines their joint effect on analysts' forecasting behaviour. In the models above, the dependent variable is analysts' forecasting behaviour, denoted by *AnlBeh*, which includes the number of analysts' forecast reports (*Anlrptnum*) and the forecast error (*Ferror*). *Anlrptnum* is equal to the logarithm of the number of forecast reports issued by all analysts in a brokerage house covering firm *i* in year *t*. *Ferror* is the forecast error or accuracy, which equals the absolute value of the difference between the earnings per share (EPS) forecast of analysts and the actual value divided by the total market value at the end of the previous year. Since a brokerage house may issue forecasts about firm *i* several times in year *t*, the average earnings forecast is used. The two dependent variables are taken at the firm-year-analyst level.

Regarding the main explanatory variables, *Distance* is the geographical distance between company *i* and analyst *j* in year *t*; its natural logarithm is taken as the proxy for private information. The shorter the distance between firms and analysts, the more private information analysts have. *DC_rating* is the information disclosure rating derived from the Shenzhen Stock Exchange disclosure index, which is the proxy for public information. All the firms listed on the Shenzhen Stock Exchange are divided into four categories: D (fail), C (pass), B (good), and A (excellent). For *DC_rating*, 1, 2, 3, and 4 are used to represent D (fail), C (pass), B (good), and A (excellent), respectively. The smaller the value of the *DC_rating*, the lower the disclosure rating and the worse the public information environment. Thus, model (13) tests the public information effect on analysts' forecasts, model (14) tests the private information effect on analysts' forecasts, and model (15) tests both jointly. For hypothesis H2, the full sample is divided into two groups according to *DC_rating*, and then, on the basis of model (14), the effect of private information on analysts' forecasts in different public information environments is examined.

Table 1 Variable Definitions

Variable	Definition
Ferror	Absolute value of the difference between the analyst's EPS forecast and the actual EPS, multiplied by 100 and then divided by the total market value at the end of last year.
Anlrptnum	Number of analysts' forecast reports for each listed firm.
Distance	Natural logarithm of the geographic distance between the headquarters of the financial analyst and that of the listed firm, multiplied by -1. This means that the greater the measure, the shorter the geographic distance and the more private information gained.
DC_rating	Natural logarithm of information disclosure index.
Ananum	Natural logarithm of the number of analysts covering each listed firm.
Gexp	General experience of analysts, which equals the natural logarithm of the year difference between the year in which an analyst issue his or her first forecast report and the current year.
Fexp	Specific experience of analysts on following a listed firm, which equals the natural logarithm of the year difference between the year in which analyst *j* issues his or her first forecast report for firm *i* and year *t*.
Brksize	Brokerage size, which equals the natural logarithm of the number of analysts the brokerage firm has.
FH	Analysts' forecast horizon, which equals the natural logarithm of the number of calendar days difference between the analyst's forecast report date and the firm's earnings announcement date.
Intan	Ratio of intangible assets to total assets.
Lev	Debt ratio, which equals the total liabilities of each firm divided by its net assets
BM	The ratio of the book value of each firm's net assets to its market value
Size	The natural logarithm of the total market value of stock equity

With regard to the control variables, *Ananum* is the natural logarithm of the number of analysts who issue at least one forecast report for firm *i* in year *t*. *Gexp* refers to the general experience of analysts, which is equal to the natural logarithm of the time span between the year in which an analyst issues his or her first forecast report and the current year. *Fexp* refers to the specific experience of analyst *j* following firm *i* in year *t*, calculated as the natural logarithm of the time span between the year in which analyst *j* issues the first forecast report for firm *i* and year *t*. *Brksize* refers to the size of the brokerage house, measured as the natural logarithm of the number of analysts each brokerage house has. *FH* refers to the natural logarithm of the number of calendar days between each analyst's earnings forecast date and the earnings announcement date. *Intan*, *Lev*, *BM*, and *Size* are firm characteristics. *Intan* equals intangible assets divided by total assets. *Lev* equals the total liabilities divided by net assets. *BM* equals the book value of stock equity divided by its market value. *Size* equals the natural logarithm of the total market value of firm *i*. *Year*/*Ind* refers to the year and industry dummy variables.

IV. Empirical Analysis

4.1 Sample Selection and Descriptive Statistics

In this paper, the initial sample includes all analysts' forecasts from 2002 to 2011 for the 1,583 firms listed on the Shenzhen Stock Exchange,[4] totalling 11,245 firm-year observations. Because each financial analyst may forecast earnings several times a year, and because there may be multiple analysts covering each firm, there is a total of 62,330 firm-year-analyst observations and 130,170 firm-year-analyst-forecast observations. Table 2 presents the sample selection procedures. First, B-share firms are excluded. Second, following Cuijpers and Buijink (2005), only those analysts' forecast reports issued within 180 calendar days before the earnings announcement date are included. Third, firms in the finance and insurance industries are excluded. Finally, the missing values of the mean variables are dropped from the sample. The final sample includes 21,901 firm-year-analyst observations. All the data, including the financial data and firm location information, are taken from the CSMAR database. All the variables are winsorised at the highest and lowest 1% extreme values.

Table 2 Sample Selection

Sample Selection Procedure	Obs	#firm-year-analysts	#firm-years	#firms
All analysts' forecasts from 2002 to 2011	130,170	62,330	11,245	1,583
After deleting B-share firms	129,590	61,772	10,688	1,524
After keeping analysts' forecasts during past six months prior to earnings announcement dates	62,194	35,094	4,632	1,375
After keeping samples at the firm-year-analyst level	35,094	35,094	4,632	1,375
After deleting firms in finance and insurance industries	34,369	34,369	4,575	1,365
After deleting the observations with missing values	21,901	21,901	3,318	695

4 Since only the Shenzhen Stock Exchange requires disclosure of the information disclosure index of listed companies, sample firms are limited to Shenzhen-listed companies.

Panel A of Table 3 lists the yearly distribution of the information disclosure rating of Shenzhen-listed firms. In 2002, out of the 99 listed firms that had disclosed information, 3.03%, 28.28%, 54.55%, and 14.14% were rated D (fail), C (pass), B (good), and A (excellent), respectively. But in 2011, the disclosure rating percentages changed significantly to 1.07%, 21.49%, 64.3%, and 13.14%, respectively. It is obvious that most of the Chinese listed firms are rated “good”, and during the years 2002 to 2011, the number of companies that were rated “pass” decreased and the number of “good” firms increased. This implies that the information disclosure level of Chinese listed firms is improving. Panel B of Table 3 lists the distribution of the geographical distance between listed firms and brokerage houses. As shown by the table, the distances are mostly between 1,000km and 2,000km, accounting for 43.06% of the listed firms; the distance is within 500km for 18.33% of the listed firms, between 500 and 1,000km for 17.14%, and over 2,000km for 21.47%. This means that there is certain dispersion in the geographical location, which helps to distinguish the difference in private information between analysts.

Table 3 Distribution of Information Disclosure Ratings and Geographical Distance

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Total
Panel A: Distribution of Information Disclosure Ratings											
D (Fail)	3	5	1	2	5	5	1	1	6	6	35
	3.03	4	0.88	0.92	1.69	1.34	0.21	0.19	1.13	1.07	
C (Pass)	28	26	15	36	70	96	90	98	106	121	686
	28.28	20.8	13.27	16.59	23.73	25.81	18.48	19.07	19.89	21.49	
B (Good)	54	71	78	138	171	218	326	341	347	362	2,106
	54.55	56.8	69.03	63.59	57.97	58.6	66.94	66.34	65.1	64.3	
A (Excellent)	14	23	19	41	49	53	70	74	74	74	491
	14.14	18.4	16.81	18.89	16.61	14.25	14.37	14.4	13.88	13.14	
Total	99	125	113	217	295	372	487	514	533	563	3,318
Panel B: Distribution of Geographical Distance (km)											
<=500	29	47	36	168	158	273	521	374	394	276	2,276
	25.44%	20.43%	15.32%	20.44%	18.08%	19.50%	18.48%	18.94%	18.27%	15.38%	18.33%
>500 and	16	36	28	119	158	262	480	343	393	294	2,129
<=1000	14.04%	15.65%	11.91%	14.48%	18.08%	18.71%	17.03%	17.37%	18.23%	16.39%	17.14%
>1000 and	44	99	117	347	379	579	1,175	853	931	824	5,348
<=2000	38.60%	43.04%	49.79%	42.21%	43.36%	41.36%	41.68%	43.19%	43.18%	45.93%	43.06%
>2000 and	24	34	47	149	155	231	495	318	317	293	2,063
<=3000	21.05%	14.78%	20%	18.13%	17.73%	16.50%	17.56%	16.10%	14.70%	16.33%	16.61%
>3000	1	14	7	39	24	55	148	87	121	107	603
	0.88%	6.09%	2.98%	4.74%	2.75%	3.93%	5.25%	4.41%	5.61%	5.96%	4.86%
Total	114	230	235	822	874	1,400	2,819	1,975	2,156	1,794	12,419

Panel A and Panel B of Table 4 list the descriptive statistics and correlation matrix of correlated variables. *Ferror* is the error of analysts' forecasts, the mean value of which is 1.955%, which suggests that the average error of analysts' forecasts is nearly 2% of each firm's market value in China. *Anlrptnum* shows that the average number of forecast reports issued by each broker for every year is 1.866. *Distance* is the geographical distance between the brokerage headquarters and listed companies, the mean value of which is 1,341km. *DC_rating* is the information disclosure rating of listed firms, the mean value of which is 3.097, which suggests that the information disclosure level of most listed companies reaches at least B, similar to the results presented in Table 3. *Ananum* shows that the average number of analysts following each company is 15.02. *Gexp* shows that the experience of analysts is, in general, 3.49 years. *Fexp* shows that the average firm-specific experience of analysts is 1.62 years. During the sample period, the average number of analysts that brokerage houses have is 29.1 (*Brksize*).

Table 4 Descriptive Statistics and Correlation Matrix

Panel A: Descriptive Statistics

Variable	Obs.	Mean	Std.	P25	Median	P75
Ferror	21,901	1.955	3.640	0.258	0.725	2.037
Anlrptnum	21,901	1.866	1.399	1.000	1.000	2.000
Distance (original value)	21,901	1,341	879	741	1304	1890
DC_rating (original value)	21,901	3.097	0.626	3.000	3.000	3.000
Ananum (original value)	21,901	15.02	10.37	7.00	13.00	22.00
Size (million yuan)	21,901	12,958	19,577	3,039	6,053	13,470
Gexp (original value)	21,901	3.49	2.06	2.00	3.00	5.00
Fexp (original value)	21,901	1.62	1.13	1.00	1.00	2.00
Brksize (original value)	21,901	29.10	14.30	18.00	29.00	38.00
Intan	21,901	0.041	0.044	0.013	0.030	0.054
Lev	21,901	0.138	0.155	0.012	0.068	0.240
BM	21,901	0.354	0.257	0.171	0.280	0.465
FH	21,901	264.84	90.48	196.00	263.00	339.50

Panel B of Table 4 reports the correlation matrix of the variables. The Pearson correlation coefficients are in the upper right corner, and the Spearman correlation coefficients are in the lower left corner. As shown by the table, there is a negative correlation between *Distance* and *Ferror*, which means that the forecast error is less when the distance between listed companies and brokers is shorter. There is a positive correlation between *Distance* and *Anlrptnum* (the Spearman coefficient is not significant), which means that analysts update forecast reports more frequently when the distance between listed companies and brokers is shorter. There is a negative correlation between *DC_rating* and

Table 4 Descriptive Statistics and Correlation Matrix (Continued)

Panel B: Correlation Matrix

Variable	*Ferror*	*Anlrptnum*	*Distance*	*DC_rating*	*Ananum*	*Size*	*Gexp*	*Fexp*	*Brksize*	*Intan*	*Lev*	*BM*	*FH*
Ferror		-0.01	-0.028***	-0.051***	-0.047***	-0.111***	0.009	0.028***	0.008	-0.045***	0.193***	0.399***	0.262***
Anlrptnum	0.009		0.035***	0.081***	0.252***	0.189***	0.191***	0.248***	0.181***	0.005	0.005	-0.015**	0.084***
Distance	-0.028***	0.008		0.021***	-0.028***	-0.001	0.014**	0.032***	0.009	-0.003	-0.014**	0.004	0.006
DC_rating	-0.017**	0.089***	0.006		0.293***	0.212***	0.021***	0.124***	0.012*	-0.059***	-0.001	0.062***	0.023***
Ananum	-0.043***	0.276***	-0.029***	0.298***		0.576***	0.191***	0.269***	0.049***	-0.034***	0.005	-0.077***	0.079***
Size	-0.125***	0.19***	-0.051***	0.222***	0.587***		0.148***	0.268***	0.035***	-0.071***	0.167***	-0.182***	-0.025***
Gexp	0.048***	0.198***	0.005	0.021***	0.189***	0.154***		0.456***	0.274***	0.04***	0.034***	-0.04***	0.174***
Fexp	0.055***	0.274***	-0.006	0.126***	0.281***	0.263***	0.436***		0.22***	-0.008	0.073***	0.025***	0.17***
Brksize	0.015**	0.205***	0.016**	0.011	0.06***	0.052***	0.295***	0.239***		0.025***	-0.003	-0.026***	0.083***
Intan	-0.031***	0.018***	0.032***	-0.072***	0.002	-0.074***	0.07***	0.016**	0.046***		-0.071***	-0.107***	0.02***
Lev	0.205***	0.005	-0.055***	-0.004	-0.007	0.155***	0.039***	0.071***	0.003	-0.05***		0.257***	0.03***
BM	0.44***	-0.024***	0.005	0.079***	-0.1***	-0.251***	-0.041***	0.001	-0.045***	-0.102***	0.301***		0.145***
FH	0.386***	0.023***	-0.016**	0.017**	0.061***	-0.034***	0.158***	0.184***	0.084***	0.028***	0.037***	0.153***	

Note: (1) The number of observations is 21,901; (2) Pearson correlation coefficients are in the upper right corner, and Spearman correlation coefficients are in the lower left corner; (3) ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

Ferror, which means that the forecast error is less when the information disclosure level is higher. There is a positive correlation between *DC_rating* and *Anlrptnum*, which means that analysts update their forecast reports more frequently when the information disclosure level is higher. As shown from the correlation coefficients between other variables, there is no serious multicollinearity among them.

4.2 Separate Influence of Public and Private Information on Analysts' Forecasts

We first use model (13) to test the influence of public information, and the results are presented in Table 5. Panel A of Table 5 lists the univariate test results, and Panel B lists the multiple regression results. In Panel A, the lower information disclosure rating group includes companies rated D (fail) or C (pass) and the higher rating group includes those rated B (good) or A (excellent). As shown by the table, the average value of *Ferror* is 2.35% for the lower rating group and 1.89% for the higher rating group. In addition, the average value of *Ferror* for the lower rating group is significantly greater than that for the higher rating group. This means that analysts' forecast errors are fewer for firms with a higher disclosure rating. Similarly, the value of *Anlrptnum* is 1.7559 for the lower group and 1.8845 for the higher group, and the difference between them is significant. This means that analysts update their forecast reports more frequently for firms with a higher disclosure rating. As shown by the multiple regression results in Panel B of Table 5, when other variables, such as firm size (*Size*) and brokerage size (*Brksize*), are controlled for, *DC_rating* has a significantly negative influence on *Ferror*. This means that the forecast errors are fewer when the information disclosure rating is higher. The research result is consistent with previous findings (e.g. Bai, 2009; Fang, 2007). In addition, *DC_rating* has a positive but insignificant influence on *Anlrptnum*. This means that analysts update their forecast reports more frequently when the information disclosure level is higher, but the difference is not significant.

Model (14) is used to test the influence of private information on *Anlrptnum* and *Ferror*. The results are presented in Table 6. Panel A and Panel B of Table 6 list the univariate and multiple test results, respectively. As shown by the results of the univariate tests, the average value of *Ferror* is 1.90% for the shorter distance group and 2.02% for the greater distance group and the difference between them is significant at the 5% level. Analysts' forecast error is therefore less when the distance between analysts and firms is shorter. Moreover, the average value of *Anlrprnum* is 1.8833 for the shorter distance group and 1.8465 for the greater distance group and the difference between them is significant at the 10% level. This means that analysts update their forecast reports more frequently for firms at a shorter rather than greater distance from the analysts. Panel B of Table 6 shows that *Distance* has a significantly negative influence on *Ferror* and a significantly positive influence on

Anlrptnum. These results are the same as the univariate results, which are also consistent with previous studies (e.g. Malloy, 2005).

Table 5 Impact of Information Disclosure Rating on Analysts' Forecasts

Panel A: Univariate Test

Variable	*DC_rating*	Obs.	Mean	Median	T-value	Z-value
Ferror	Lower	3,216	2.3504	0.7833	5.52***	4.12***
	Higher	18,685	1.8868	0.7143		
Anlrptnum	Lower	3,216	1.7559	1.0000	-5.09***	-5.49***
	Higher	18,685	1.8845	1.0000		

Panel B: Multivariate Test

Variable	*Ferror*	*Anlrptnum*
DC_rating	-1.0169***	0.0093
	(-3.07)	(0.61)
Ananum	-0.4212***	0.0674***
	(-3.86)	(12.74)
Size	0.1121	0.0339***
	(1.31)	(7.29)
Gexp	0.0329	0.0369***
	(0.60)	(5.84)
Fexp	0.1628*	0.1535***
	(1.66)	(13.61)
Brksize	-0.0826**	0.0968***
	(-1.98)	(17.79)
Intan	-0.6574	0.0925
	(-0.55)	(1.09)
Lev	3.2283***	-0.0331
	(4.64)	(-1.38)
BM	3.9314***	-0.0318**
	(6.36)	(-1.99)
FH	1.8827***	0.0173***
	(17.34)	(2.85)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.278	0.148

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

Table 6 Impact of Geographical Distance on Analysts' Forecasts

Panel A: Univariate Test

Variable	Distance	Obs.	Mean	Median	T-value	Z-value
Ferror	Shorter	11,409	1.8984	0.7078	-2.39**	-2.65***
	Greater	10,492	2.0163	0.7412		
Anlrptnum	Shorter	11,409	1.8833	1.0000	1.95*	0.75
	Greater	10,492	1.8465	1.0000		

Panel B: Multivariate Test

Variable	*Ferror*	*Anlrptnum*
Distance	-0.0333**	0.0054***
	(-2.21)	(3.05)
Ananum	-0.5131***	0.0687***
	(-4.89)	(13.30)
Size	0.1091	0.0338***
	(1.32)	(7.30)
Gexp	0.0442	0.0365***
	(0.81)	(5.77)
Fexp	0.1307	0.1527***
	(1.36)	(13.42)
Brksize	-0.0843**	0.0968***
	(-2.01)	(17.86)
Intan	-0.4705	0.0877
	(-0.40)	(1.02)
Lev	3.3017***	-0.0308
	(4.71)	(-1.29)
BM	3.8425***	-0.0315**
	(6.25)	(-2.00)
FH	1.8877***	0.0173***
	(17.23)	(2.84)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.275	0.149

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

4.3 Joint Influence of Public and Private Information on Analysts' Forecasts

This study mainly focuses on how public and private information jointly influence analysts' forecasts when analysts gain two types of information at the same time. First, public and private information are incorporated into one model to test the difference in the coefficient between them (i.e. to test H1). Second, the full sample is classified into two subsamples according to firms' information disclosure ratings to examine whether the effect of private information varies with a different public information environment (i.e. to test H2). The regression results of model (15) are listed in Table 7. Columns (1) and (2) of Table 7 report analysts' forecast errors and forecast frequency, respectively. The results show that the coefficients of both *DC_rating* and *Distance* are significantly negative in the forecast error model (Column (1)) and the magnitude of the former is larger than that of the latter. The F value is 8.88 in the bottom row, which means that the difference in effect between *DC_rating* and *Distance* is significant.[5] In other words, the effect of public information on forecast error is significantly greater than the effect of private information, which supports H1.

Table 7 Joint Impact of Public and Private Information on Analysts' Forecasts

Variable	(1) *Ferror*	(2) *Anlrptnum*
DC_rating	-1.0136***	0.0088
	(-3.07)	(0.58)
Distance	-0.0321**	0.0054***
	(-2.16)	(3.05)
Ananum	-0.4242***	0.0679***
	(-3.90)	(12.82)
Size	0.1128	0.0338***
	(1.32)	(7.31)
Gexp	0.0345	0.0366***
	(0.63)	(5.79)
Fexp	0.1700*	0.1523***
	(1.73)	(13.51)
Brksize	-0.0826**	0.0968***
	(-1.98)	(17.85)
Intan	-0.6371	0.0891
	(-0.54)	(1.04)
Lev	3.2095***	-0.0300
	(4.62)	(-1.25)
BM	3.9340***	-0.0323**
	(6.37)	(-2.04)
FH	1.8825***	0.0173***
	(17.35)	(2.85)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.278	0.149
DC_rating vs. *Distance* (F value)	8.88***	0.05

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

[5] According to models (11) and (12), we can infer that the influence coefficient of *DC_rating* is 1. Although this study focuses on the difference in influence between *DC_rating* and *Distance*, it tests whether the coefficient of *DC_rating* (-1.0136) equals 1 to enhance the reliability of the inference of the theoretical model. The corresponding F value is 0.00 and the corresponding P value is 0.97, which cannot refute the original hypotheses. This result supports the inference of the theoretical model herein.

Another method to test H2 is to divide the full sample into two groups according to information disclosure ratings and then test the difference in the effect of private information between the two groups. The higher rating group includes firms rated A or B and the lower rating group includes firms rated C or D.[6] The regression results of model (14) are listed in Table 8. Panel A and Panel B of Table 8 report univariate test results and multivariate regression results, respectively. Panel A shows that in the worst public information situation, the average value of *Ferror* is 2.354% for the shorter distance group and 2.345% for the greater distance group and the difference between them is not significant. The effect of geographical distance on analysts' forecasts is not significant when the information disclosure rating is lower. Meanwhile, analysts' forecast frequency (*Anlrptnum*) is 1.6887 for the shorter distance group and 1.8214 for the greater distance group, and the difference is significant. This means that in a lower information disclosure environment, analysts in locations closer to the firms they follow issue forecast reports less frequently than those in locations farther away from the firms they follow, which is inconsistent with our prediction.

In an external environment with a higher disclosure rating, the mean of the forecast error (*Ferror*) is 1.82% and the mean of analysts' forecast frequency (*Anlrptnum*) is 1.9147 for the shorter distance group, but for the greater distance group, the corresponding figures are 1.96% and 1.8511, respectively. The differences between the two groups in respect of both variables are significant. This means that analysts issue forecast reports more frequently and make fewer forecast errors when the information disclosure environment is better.

Panel B of Table 8 shows the multivariate regression results based on different information disclosure ratings after controlling for other variables. Columns (2) and (3) are based on the forecast error model, and Columns (4) and (5) are based on the forecast frequency model. For the lower disclosure rating group, the coefficient of *Distance* on forecast error is -0.0236 (Column 2) but not significant. However, for the higher disclosure rating, the coefficient of *Distance* is -0.0315 with t value -2.01 (Column 3), which is significant at the 1% level. Meanwhile, the F test shows that the difference in the coefficient of *Distance* between the two groups is significant at the 10% level. The coefficient of *Distance* on analysts' forecast frequency is 0.0011 (t value is 0.35) when the information environment is bad and 0.0060 (t value is 3.14) when the information environment is good. The F test shows that the difference between the two groups is not significant. Both the univariate and multivariate test results support H2. This means that the effect of private information on forecast error occurs only in a better public information environment.

---

[6] The full sample is not divided into two groups, one rated D and another rated higher than D (i.e. rated A, B or C), because there are very few observations rated D (only 58, accounting for 0.26% of the total).

Table 8 Impact of Geographical Distance on Analysts' Forecasts under Different Information Disclosure Ratings

Panel A: Univariate Test

Classification	Distance	Obs.	*Ferror*		*Anlrptnum*	
			Mean	Median	Mean	Median
Lower disclosure rating	Shorter	1,587	2.3536	0.7676	1.6887	1.0000
	Greater	1,629	2.3473	0.8078	1.8214	1.0000
	Shorter-Greater		0.0063	-0.0402*	-0.1326***	-0.0000**
Higher disclosure rating	Shorter	9,822	1.8248	0.6978	1.9147	1.0000
	Greater	8,863	1.9554	0.7322	1.8511	1.0000
	Shorter-Greater		-0.1306**	-0.0344**	0.0636***	0.000*

Panel B: Multivariate Test

Variable	*Ferror*		*Anlrptnum*	
	Lower disclosure rating	Higher disclosure rating	Lower disclosure rating	Higher disclosure rating
Distance	-0.0236	-0.0315**	0.0011	0.0060***
	(-0.60)	(-2.01)	(0.35)	(3.14)
Ananum	-0.1704	-0.4764***	0.0546***	0.0716***
	(-0.61)	(-4.16)	(4.86)	(12.29)
Size	-0.0424	0.1480*	0.0434***	0.0321***
	(-0.18)	(1.68)	(4.59)	(6.21)
Gexp	0.3042**	-0.0014	0.0229	0.0389***
	(2.03)	(-0.02)	(1.40)	(5.60)
Fexp	0.4526	0.1164	0.1618***	0.1513***
	(1.52)	(1.18)	(5.59)	(12.53)
Brksize	-0.1011	-0.0830**	0.0715***	0.1011***
	(-0.59)	(-2.19)	(5.78)	(16.90)
Intan	1.6130	-0.7619	0.2518	0.0410
	(0.51)	(-0.60)	(1.55)	(0.42)
Lev	5.2347**	2.8931***	-0.0933***	-0.0219
	(2.47)	(4.18)	(-2.70)	(-0.80)
BM	5.1766***	3.8365***	-0.0101	-0.0308*
	(3.05)	(5.86)	(-0.30)	(-1.82)
FH	2.0846***	1.8325***	0.0214*	0.0160**
	(7.15)	(16.24)	(1.76)	(2.32)
Year Control	Yes	Yes	Yes	Yes
Industry Control	Yes	Yes	Yes	Yes
Observations	3,216	18,685	3,216	18,685
Adj. R^2	0.294	0.281	0.153	0.148

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

4.4. Robustness Tests

The following robustness tests are performed:

1. Public information is measured as the information disclosure rating provided by the Shenzhen Stock Exchange for listed firms. Because the measurement is a comprehensive indicator influenced by multiple factors (e.g. analyst following and analysts' forecast), it is vulnerable to endogeneity problems. To solve this issue, the two-stage least squares method is used to further explore the results presented in Table 5. As regards the influencing factors of the information disclosure rating, referring to Hope (2003a), this paper chooses firm-level characteristics, such as firm size (*Size*), firm profitability (*ROE*), financial leverage (*Lev*), the number of analysts following (*Ananum*), audit quality (*Auditors*), government control (*SOE*), and institutional investor holdings *(Inst_holding*), to construct the first-stage model.

With the exceptions of *Size*, *Lev*, and *Ananum*, the firm-level characteristics in the first-stage model are newly added exogenous variables. *ROE* is the return on net assets. *Auditor* is a dummy variable which equals 1 if the auditor of the listed company is one of the Big Four and 0 otherwise. *SOE* is a dummy variable which equals 1 if the ultimate controller is the central or a local government and 0 otherwise. *Inst_holding* is the ownership of institutional holders. The regression results of the first-stage model are listed in Panel B of Table 9; they show that *Ananum*, *Size*, *Auditor*, and *Inst_holding* have a positive and significant influence on the information disclosure rating, consistent with the prediction. In the second stage, the predicted value of *DC_rating* is incorporated and the results are similar to those shown in Table 5. Panel A of Table 9 shows that *DC_rating* has a negative and significant influence on *Ferror* and a positive and significant influence on *Anlrptnum*. These outcomes imply that this study's results remain unchanged when endogenous problems are considered.

2. The relationship between private information and analysts' forecasting behaviour (presented in Table 6) also has endogeneity problems. Analysts' decisions to follow a certain company are endogenous, thus leading to a natural relationship between forecasting results and information acquisition (Malloy, 2005). Prior researchers (e.g. Malloy, 2005) have summarised three possibilities that can cause endogenous problems: (1) Local analysts, especially those whose registered addresses are located in remote areas, follow few companies and focus more on specific firms, which can help their forecast capabilities; this indicates that the influence of short geographical distance on analysts' forecast error emanates from the perspective of analysts' efforts rather than information acquisition. (2) Local analysts are more professional in certain industries. (3) Local analysts spend more time and effort on analysing customers' stocks; as a result, the improvement in their forecasting accuracy is based on their customers' demands.

Table 9 Endogeneity and Two-Stage Regression Results of Public Information

Variable	Panel A		Panel B
	Ferror	*Anlrptnum*	*DC_rating*
DC_rating	-11.5335***	0.4362***	
	(-3.01)	(2.90)	
Ananum	0.5079	0.0296**	0.0700***
	(1.47)	(2.13)	(5.70)
Size	0.1496	0.0325***	0.0254*
	(0.70)	(3.97)	(1.78)
Gexp	-0.0904	0.0405***	
	(-1.16)	(5.96)	
Fexp	0.5622***	0.1375***	
	(2.72)	(10.67)	
Brksize	-0.0604	0.0968***	
	(-1.18)	(17.25)	
Intan	-2.3244	0.1649	
	(-0.85)	(1.34)	
Lev	2.3415**	0.0044	-0.0099
	(2.30)	(0.11)	(-0.14)
BM	4.8226***	-0.0699***	
	(6.65)	(-2.75)	
FH	1.8147***	0.0203***	
	(15.88)	(3.12)	
Year Control	Yes	Yes	
Industry Control	Yes	Yes	
Constant	-4.3601	-0.9779***	0.4450
	(-0.98)	(-5.38)	(1.47)
ROE			0.0335
			(0.30)
Auditor			0.0734**
			(2.23)
SOE			0.0143
			(0.71)
Inst_holding			0.0989***
			(2.84)
Observations	21,685	21,685	21,685
Adj. R^2	0.282	0.146	0.093

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

The two-stage least squares method is still used to mitigate endogenous problems. This study follows Malloy's (2005) research, which chooses such variables in the first-stage model as firm size (*Size*), brokerage size (*Brksize*), the number of industries covered by financial analysts (*Indnum*), auditor quality (*Auditor*), institutional investor holdings

(*Inst_holding*), and information disclosure rating (*DC_rating*). The larger the brokerage size, the more likely it is to be located in a big city; otherwise, the brokerage is more likely to be located in a small city. This partly reflects point (1) in Malloy (2005). Meanwhile, in combination with firm size (*Size*), the geographical distance between the financial analysts' headquarters and the listed firm is known. In accordance with point (2) of Malloy (2005), this study chooses the number of industries covered by analysts (*Indnum*) as the proxy for the efforts analysts put in to gain private information because the greater the number of industries analysts follow, the more likely they are to diversify their efforts to gain private information, and this can also complement analysts' knowledge about certain industries, thereby improving the acquisition of private information. Institutional investor holdings (*Inst_holding*) are chosen according to point (3) of Malloy (2005). The higher the institutional investor holding, the better the public information environment, leading to analysts being able to obtain a higher quantity or quality of private information. In addition, two variables of public information that may influence private information are chosen, namely the information disclosure rating (*DC_rating*) and audit quality (*Auditor*). The definitions of these variables are the same as noted above. The results of the first-stage model based on the factors of private information (*Distance*) are presented in Panel B of Table 10. Table 10 shows that *Indnum*, *Auditor,* and *DC_rating* have a positive influence on private information, except that the influence of *DC_rating* is insignificant. Panel A lists the second-stage results where endogenous problems are considered. The results show that the closer the listed firm is located to analysts, the fewer the forecast errors and the more frequently analysts issue forecasts reports, which is consistent with Table 6.

3. This paper uses information disclosure ratings to measure public information and the geographical distance between financial analysts' headquarters and the listed firm to measure private information. However, most financial analysts are located in big cities, such as Beijing, Shanghai, and Shenzhen, where the public information environments are better for listed companies. Therefore, firms located in these big cities are at a shorter geographical distance from financial analysts and have better public information environments. This means that geographical distance is a proxy not only for private information but also for public information to a certain extent. This effect is especially important in Table 8. To mitigate this issue, two methods are adopted. Firstly, public information from geographical distance is excluded (i.e. regressing geographical distance on the information disclosure rating and then obtaining the residual of geographical distance). Incorporating the residual of geographical distance in the model generates the new results displayed in Table 11. Table 11 shows results similar to those in Table 8. Secondly, those firms with headquarters located in Beijing, Shanghai, and Shenzhen are excluded to mitigate the public information effect on geographical distance. As a result of these exclusions, 17,459 observations remain from the

full sample. The new results (untabulated) are similar to those shown in Table 8, especially the results of the analyst forecasting error model.

Table 10 Endogeneity and Two-Stage Regression Results of Private Information

Variable	Panel A		Panel B
	Ferror	*Anlrptnum*	*Distance*
Distance	-1.2287**	0.0618**	
	(-2.13)	(2.08)	
Ananum	-0.5983***	0.0728***	
	(-4.55)	(11.16)	
Size	0.1330	0.0328***	-0.0400
	(1.27)	(5.95)	(-0.92)
Gexp	0.0816	0.0333***	
	(1.03)	(4.71)	
Fexp	0.3931**	0.1404***	
	(2.25)	(10.64)	
Brksize	-0.0775	0.0974***	0.0426
	(-1.31)	(17.39)	(1.30)
Intan	0.2720	0.0579	
	(0.18)	(0.54)	
Lev	2.6193***	0.0031	
	(3.27)	(0.10)	
BM	3.9243***	-0.0374**	
	(6.45)	(-2.16)	
FH	1.8618***	0.0187***	
	(16.02)	(2.85)	
Year Control	Yes	Yes	
Industry Control	Yes	Yes	
Constant	-22.0361***	-0.2045	-6.0685***
	(-4.48)	(-0.80)	(-6.61)
Indnum			0.0566*
			(1.68)
Auditor			0.5786***
			(3.46)
DC_rating			0.1357
			(0.74)
Inst_holding			-0.0887
			(-0.59)
Observations	21,685	21,685	21,685
Adj. R^2	0.278	0.146	0.009

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

Table 11 Robustness Tests

Variable	*Ferror*		*Anlrptnum*	
	Lower disclosure rating	Higher disclosure rating	Lower disclosure rating	Higher disclosure rating
Distance	0.0204	-0.0549**	0.0026	0.0125***
	(0.40)	(-2.05)	(0.50)	(3.57)
Ananum	0.0279	0.1264	0.0855***	0.0904***
	(0.17)	(0.75)	(6.15)	(9.85)
Size	-0.2156	-0.2919***	0.0202***	0.0180***
	(-1.10)	(-3.18)	(3.55)	(3.73)
Gexp	0.1565	-0.1588	0.0482***	0.0447***
	(0.52)	(-0.81)	(3.17)	(3.78)
Fexp	0.2608	-0.0005	0.1639***	0.1454***
	(0.58)	(-0.01)	(5.06)	(5.36)
Brksize	0.0399	0.0733	0.0667***	0.0975***
	(0.28)	(0.94)	(5.89)	(11.34)
Intan	3.2557*	0.4868	0.1966	-0.0139
	(1.79)	(0.56)	(1.51)	(-0.12)
Lev	6.4868**	2.6164*	-0.0402	0.0497
	(2.09)	(1.95)	(-1.43)	(1.60)
BM	7.1339***	4.5863***	0.0409***	0.0320
	(5.20)	(4.18)	(6.36)	(1.15)
FH	2.2207***	1.8580***	0.0292**	0.0305
	(3.37)	(3.15)	(2.20)	(1.43)
Year Control	Yes	Yes	Yes	Yes
Industry Control	Yes	Yes	Yes	Yes
Observations	3,216	18,685	3,216	18,685
Adj. R^2	0.328	0.263	0.148	0.146

Note: The t-values in parentheses are based on standard errors adjusted for firm-level clustering. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed tests.

V. Conclusions and Limitations

5.1 Conclusions

Public and private information are the two main factors that influence analysts' forecasting behaviour. Previous literature has explored the separate effects of public information (Baldwin, 1984; Hodder *et al.,* 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.,* 2006) and private information (Chen and Jiang, 2006; Green *et al.,* 2014; Mayew *et al.,* 2013) on analysts' forecasting behaviour. Unlike these previous studies, this paper uses geographical distance and information disclosure ratings as proxies for public and private information, respectively, and focuses on examining their joint effects on analysts' forecasting behaviour. Empirical results show that the separate effect of public and private information is consistent with the prediction herein and prior findings. Specifically, the higher the information disclosure rating and the shorter the geographical distance, the fewer the number of forecast errors and the more frequently financial analysts issue forecast reports. In its investigation on the joint effects of public and private information, this paper finds that public information has a slightly stronger influence on analysts' forecasts than private information. When categorising the full data sample into two groups according to information disclosure ratings, geographical distance imposes an impact on analysts' forecasts only in the higher disclosure rating group. These results indicate that compared with private information, public information has a greater influence on analysts' forecasts. Meanwhile, the effect of private information depends on a better public information environment.

Different from previous research which concentrates on the separate effects of public and private information on analysts' forecasting behaviour, this paper mainly focuses on their joint effects. Hence, this paper complements previous research. Moreover, this paper has a practical contribution. Securities analysts play an important intermediary role in capital markets and can mitigate information asymmetry between listed companies and outside investors. With the gradual improvement in China's securities market and public information environment, whether and how financial analysts are influenced by public and private information is becoming an increasingly important issue for all market participants. The findings presented in this paper provide evidence on these issues and will also help investors to better understand securities analysts' behaviour and so assist more reasonable and informed decision making.

5.2 Limitations

The major limitation of this research is the measurement of private information. It is difficult to find in the current literature a single reasonable and efficient indicator as a proxy for the efforts and costs that analysts spend in order to gain private information. This paper takes the geographical distance of headquarters location between the brokerage house and the listed firm as a proxy for the private information analysts obtain. Although there are some previous theoretical research studies to support this approach (Agarwal and Hauswald, 2010; García and Norli, 2012; Hauswald and Marquez, 2006; Kim *et al.,* 2012; Malloy, 2005; Petersen and Rajan, 2002), some errors are inevitable. Firstly, geographical distance may reflect certain public information. Secondly, the convenience and diversity of transportation decrease the effects of geographical distance. As a result, geographical distance cannot fully reflect private information. Although the first limitation is discussed and tested in the robustness tests, the negative influence of the second limitation cannot be fully removed. This is a topic warranting future research.

References

Agarwal, S. and Hauswald, R. (2010), 'Distance and Private Information in Lending', *Review of Financial Studies* 23 (7): 2757-2788.

Bae, K., Stulz, R. M., and Tan, H. (2008), 'Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts', *Journal of Financial Economics* 88 (3): 581-606.

Bai, X. (2009), 'Shangshi Gongsi Xinxi Peilu Zhengce dui Fenxishi Yuce de Duochong Yingxiang Yanjiu' (A Study on the Impact of Information Disclosure Policy of Listed Corporations on Analysts' Forecast)，*Journal of Financial Research*, Issue 4: 92-112.

Baldwin, B. A. (1984), 'Segment Earnings Disclosure and the Ability of Security Analysts to Forecast Earnings Per Share', *The Accounting Review* 59 (3): 376-389.

Barron, O. E., Kim, O., Lim, S. C., and Stevens, D. E. (1998), 'Using analysts' forecasts to measure properties of analysts' information environment', *The Accounting Review* 73 (4): 421-433.

Behn, B. K., Choi, J. and Kang, T. (2008), 'Audit Quality and Properties of Analyst Earnings Forecasts', *The Accounting Review* 83 (2): 327-349.

Burgstahler, D. C. and Eames, M. J. (2003), 'Earnings Management to Avoid Losses and Earnings Decrease: Are Analysts Fooled?', *Contemporary Accounting Research* 20 (2): 253-294.

Bushman, R. M. (1991), 'Public Disclosure and the Structure of Private Information Markets', *Journal of Accounting Research* 29 (2): 261-276.

Chen, Q. and Jiang, W. (2006), 'Analysts' Weighting of Private and Public Information', *Review of Financial Studies* 19 (1): 319-355.

Cuijpers, R. and Buijink, W. (2005), 'Voluntary Adoption of Non-Local GAAP in the European Union: A Study of Determinants and Consequences', *European Accounting Review* 14 (3): 487-524.

De Jong, P. J. and Apilado, V. P. (2009), 'The changing relationship between earnings expectations and earnings for value and growth stocks during Reg FD', *Journal of Banking and Finance* 33 (2): 435-442.

DeFond, M., Francis, J. R., and Hu, X. (2008), 'The Geography of Auditor Independence and SEC Enforcement', Working Paper, University of Southern California.

Fang, J. (2007), 'Woguo Shangshi Gongsi Xinxi Peilu Toumingdu yu Zhengquan Fenxishi Yuce' (Transparency of Information Disclosure and Forecast of Securities Analysts in China's Listed Corporations), *Journal of Financial Research*, Issue 6: 136-148.

García, D. and Norli, Ø. (2012), 'Geographic dispersion and stock returns', *Journal of*

Financial Economics 106 (3): 547-565.

Gintschel, A. and Markov, S. (2004), 'The effectiveness of regulation FD', *Journal of Accounting and Economics* 37 (3): 293-314.

Green, T. C., Jame, R., Markov, S., and Subasi, M. (2014), 'Access to Management and the Informativeness of Analyst Research', *Journal of Financial Economics* 114 (2): 239-255.

Hauswald, R. and Marquez, R. (2006), 'Competition and Strategic Information Acquisition in Credit Markets', *Review of Financial Studies* 19 (3): 967-1000.

Hodder, L., Hopkins, P. E., and Wood, D. A. (2008), 'The Effects of Financial Statement and Informational Complexity on Analysts' Cash Flow Forecasts', *The Accounting Review* 83 (4): 915-956.

Hope, O. K. (2003a), 'Disclosure Practices, Enforcement of Accounting Standards, and Analysts' Forecast Accuracy: An International Study', *Journal of Accounting Research* 41 (2): 235-272.

Hope, O. K. (2003b), 'Accounting Policy disclosures and Analysts' Forecasts', *Contemporary Accounting Research* 20 (2): 295-321.

Hu, Y. and Lin, W. (2005), 'Xinxi Guanzhu Shendu, Fenxi Nengli yu Fenxi Zhiliang – dui Woguo Zhengquan Fenxishi de Diaocha Fenxi' (Information Depth, Analytical Ability and Quality of Analysis – An Investigation and Analysis of Securities Analysts in China), *Journal of Financial Research*, Issue 2: 46-58.

John, K., Knyazeva, A., and Knyazeva, D. (2011), 'Does geography matter? Firm location and corporate payout policy', *Journal of Financial Economics* 101 (3): 533-551.

Kim, C. F., Pantzalis, C., and Park, J. C. (2012), 'Political geography and stock returns: The value and risk implications of proximity to political power', *Journal of Financial Economics* 106 (1): 196-228.

Lang, M. H. and Lundholm, R. J. (1996), 'Corporate Disclosure Policy and Analyst Behavior', *The Accounting Review* 71 (4): 467-492.

Langberg, N. and Sivaramakrishnan, K. (2008), 'Voluntary disclosures and information production by analysts', *Journal of Accounting and Economics* 46 (1): 78-100.

Li, D. and Jia, N. (2009), 'Yingyu Zhiliang, Zhidu Huanjing yu Fenxishi Yuce' (Earnings Quality, Insitutional Environment, and Analyst Forecasts), *China Accounting Review*, Issue 4: 351-370.

Libby, R., Tan, H., and Hunton, J. E. (2006), 'Does the Form of Management's Earnings Guidance Affect Analysts' Earnings Forecasts?', *The Accounting Review* 81 (1): 207-225.

Lundholm, R. J. (1988), 'Price-Signal Relations in the Presence of Correlated Public and Private Information', *Journal of Accounting Research* 26 (1): 107-118.

Malloy, C. J. (2005), 'The Geography of Equity Analysis', *Journal of Finance* 60 (2): 719-755.

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.

McEwen, R. A. and Hunton, J. E. (1999), 'Is Analyst Forecast Accuracy Associated With Accounting Information Use?', *Accounting Horizons* 13 (1): 1-16.

McNichols, M. and Trueman, B. (1994), 'Public disclosure, private information collection, and short-term trading', *Journal of Accounting and Economics* 17 (1-2): 69-94.

O'Brien, P. and Bhushan, R. (1990), 'Analyst Following and Institutional Ownership', *Journal of Accounting Research* 28 (3): 55-82.

Pankoff, L. D. and Virgil, R. L. (1970), 'Some Preliminary Findings from a Laboratory Experiment on the Usefulness of Financial Accounting Information to Security Analysts', *Journal of Accounting Research* 8 (3): 1-48.

Payne, J. L. (2008), 'The Influence of Audit Firm Specialization on Analysts' Forecast Errors', *Auditing* 27 (2): 109-136.

Petersen, M. A. and Rajan, R. G. (2002), 'Does Distance Still Matter? The Information Revolution in Small Business Lending', *Journal of Finance* 57 (6): 2533-2570.

Robertson, J. C. (1988), 'Analysts' Reaction to Auditors' Messages in Qualified Reports', *Accounting Horizons* 2 (2): 82-89.

Teo, M. (2009), 'The Geography of Hedge Funds', *Review of Financial Studies* 22 (9): 3531-3561.

Vergoossen, R. G. A. (1993), 'The use and perceived importance of annual reports by investment analysts in the Netherlands', *European Accounting Review* 2 (2): 219-243.

Verrecchia, R. E. (1982), 'Information Acquisition in a Noisy Rational Expectations Economy', *Econometrica* 50 (6): 1415-1430.

Wang, Y. (2006), 'Zhongguo Gupiao Shichang Gongkai Xinxi yu Siyou Xinxi de Hubu Xiaoying' (The complementary effect of public information and private information in the Chinese stock market), *Journal of Financial Research*, Issue 6: 41-52.

Wang, Y., Chen, X., and Hou, Y. (2010), 'Guonei Zhengquan Fenxinshi de Xinxi Youshi: Dili Linjinxing Haishi Kuaiji Zhunze Chayi' (Information advantages of domestic securities analysts: Geographical proximity or accounting standards differences), *Accounting Research*, Issue 12: 34-40.

Wang, Y., Chen, Y., and Wang, J. (2015), 'Management earnings forecasts and analyst forecasts: Evidence from mandatory disclosure system', *China Journal of Accounting Research* 8 (2): 133-146.

Wang, Y. and Wang, Y. (2012), 'Yeji Yugao Xinxi dui Fenxishi Yuce Xingwei You Yingxiang ma?' (Does the information of performance forecast influence the behavior of analysts?), *Journal of Financial Research*, Issue 6: 202-218.

Yu, F. F. (2008), 'Analyst coverage and earnings management', *Journal of Financial Economics* 88 (2): 245-271.

Volume 18, Number 3 – September 2016
China Accounting and Finance Review
中 国 会 计 与 财 务 研 究
2016 年 9 月 第 18 卷 第 3 期

公共信息、私有信息与证券分析师预测*

王玉涛[1]

投稿日：2014 年 12 月 18 日 录用日：2016 年 4 月 18 日

摘要

本文重点考察了公共信息、私有信息以及两者对证券分析师预测的联合影响。信息披露水平决定了分析师获取的公共信息的质量和数量，上市公司与分析师所在机构注册地的地理距离决定了分析师获取私有信息的难度和成本，进而决定私有信息数量和质量。本文首先检验并发现信息披露水平越高、地理距离越近，证券分析师发布预测报告的次数越多，预测误差越低。其次，分析和检验信息披露水平和地理距离如何共同影响证券分析师预测，结果发现，信息披露水平对分析师预测误差和预测次数的影响显著高于地理距离的影响；同时，地理距离对证券分析师预测的影响只在信息披露环境较好时出现。这些结果意味着，虽然公共信息和私有信息对分析师预测均有影响，但影响存在不对称性。公共信息对分析师预测的影响明显高于私有信息，私有信息对分析师预测的影响依赖于更好的公共信息环境。本文的分析对我国资本市场监管者、参与者和学术界理解公共信息和私有信息对证券分析师预测的影响有重要参考价值。

关键词：公共信息、私有信息、证券分析师预测、信息披露、地理距离

* 本文得到国家自然科学基金项目（批准号：71102124）、北京市会计类专业群（改革试点）建设项目和北京市教育委员会共建项目的资助。

[1] 中央财经大学会计学院副教授，博士；研究兴趣：公司估值、审计、分析师行为。北京市海淀区学院南路 39 号中央财经大学会计学院，邮编：100081；Email: wangyutao@cufe.edu.cn。

一、 引言

随着我国证券市场的发展，财政部不断深化财税体制改革，在 2006 年进行了重大会计准则变革之后，又在 2014 年加快了会计改革步伐。[2] 会计准则和财税体制的改革促使我国证券市场信息环境发生着翻天覆地的变化，这些变化必然会对会计信息的提供者、信息中介、信息使用者的行为产生重要影响。作为信息中介，证券分析师利用公共信息（public information）和私有信息（private information）进行盈余预测、股票评级和推荐等，向资本市场参与者提供更多与决策相关的信息，因此能够发挥缓解信息不对称、提高资本市场效率的作用。基于过去近 10 年的市场环境，本文试图考察一个一般性的问题，即公共信息和私有信息如何联合影响我国证券分析师预测行为，这是与以往研究关注两类信息的单独影响是不同的。

以往许多研究考察了公共信息对分析师预测的单独影响，包括财务报表、管理层盈余预测（Baldwin, 1984; Hodder *et al.*, 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.*, 2006）、以及综合信息披露等（Hope, 2003a；白晓宇，2009；方军雄，2007）。近几年，研究者通过分析师的一些行为来衡量其获得私有信息的差异，从而考察私有信息对证券分析师预测的影响，如 Green *et al.*（2014）、Mayew *et al.*（2013）、Malloy（2005）等。与这些研究不同，本文试图综合考察公共信息、私有信息以及两者对证券分析师预测的联合影响。用信息披露指数衡量公共信息的数量和质量，用地理距离衡量私有信息的数量和质量，本文考察了这些信息对证券分析师预测行为的影响，结果发现：信息披露水平越高、地理距离越近，证券分析师预测误差越低、发布预测报告的次数越多，这是与以前研究文献一致的。其次，通过理论探讨和实证分析发现，在信息披露水平和地理距离对分析师预测的联合影响中，信息披露水平的作用显著大于地理距离；同时，当根据信息披露水平高低进行分组后，地理距离对预测误差和预测次数的影响只在高水平组中显著。这些结果意味着，相对于私有信息，公共信息对分析师预测行为的影响作用更大；私有信息作用发挥需要依赖于更好的公共信息环境。

本文主要有以下两方面的贡献：第一，补充了以前的研究成果。与以往单独考察公共信息和私有信息对分析师预测的独立影响不同（如 Baldwin, 1984; Hodder *et al.*, 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.*, 2006；白晓宇，2009；方军雄，2007），本文重点分析了两者的联合影响，补充了以前的研究成果。第二，具有现实指导意义。证券分析师在资本市场扮演重要的中介角色，能够减少上市公司与投资者之间的信息不对称。随着我国证券市场的逐步完善和公共信息环境的改善，证券分析师预测如何受到公共信息和私有信息的共同影响，是资本市场参与者都关心的重要问题。本研究能够回答这一重要问题，能够帮助投资者更好地理解证券分析师预测行为和帮

[2] 2014 年财政部会计司为了加快会计改革步伐，相继制定发布了 10 个重大的会计改革文件，包括《彩票机构新旧会计制度有关衔接问题的处理规定》（财会〔2014〕2 号）、财政部关于印发《企业会计准则第 39 号—公允价值计量》的通知（财会〔2014〕6 号）、财政部关于印发修订《企业会计准则第 30 号—财务报表列报》的通知（财会〔2014〕7 号）、财政部关于印发修订《企业会计准则第 9 号—职工薪酬》的通知（财会〔2014〕8 号）、财政部关于印发修订《企业会计准则第 33 号—合并财务报表》的通知（财会〔2014〕10 号）等。

助他们做出合理决策。

二、 文献回顾

证券分析师在资本市场中起着重要的信息加工和传递作用。由于投资者和上市公司之间存在不可避免的信息不对称，具有专业背景的分析师通过获取公共信息、私有信息对公司经营状况进行分析并提供研究报告，从而降低信息不对称。公共信息是每个证券分析师均可以同时观察到的，包括公司披露的定期报告、管理层盈余预测、审计报告等，以及管理层的一些具体行为，如利润操纵、投融资行为等，这些信息会向资本市场参与者传递一些信息，能够用于证券分析师的预测。关于公共信息与证券分析师的研究，早期的文献关注证券分析师预测准确度是否与会计数字信息的使用有关，如 Pankoff and Virgil（1970）、McEwen and Hunton（1999）、Vergoossen（1993）。后续的研究开始对财务报告的具体构成、质量等方面加以关注，考察哪些具体内容和不同质量是否以及如何影响分析师盈余预测，如考察分部报告的内容是否影响分析师盈余预测（Baldwin, 1984）、以及盈余管理或盈余质量、审计质量和专业化（李丹和贾宁，2009）是否影响分析师盈余预测（Behn *et al.*, 2008; Burgstahler and Eames, 2003; Payne, 2008; Robertson, 1988; Yu, 2008；李丹和贾宁，2009）。除了定期报告信息外，管理层也会披露企业未来盈余预测信息，因此一些学者考察了管理层盈余预测是否以及如何影响证券分析师预测（Libby *et al.*, 2006; Wang *et al.*, 2015；王玉涛和王彦超，2012）。最近的研究关注综合信息环境，如信息披露指数、会计准则等，如何外生影响证券分析师预测行为，如 O'Brien and Bhushan（1990）、Lang and Lundholm（1996）、白晓宇（2009）都发现信息披露水平越高的公司，分析师越喜欢去跟踪，并且 Lang and Lundholm（1996）、白晓宇（2009）还进一步发现，分析师对信息披露水平高的公司预测准确度更高，预测分歧度更低。Hope（2003b）通过检验各国信息披露水平与分析师预测误差、预测分歧度之间的关系，发现信息披露水平与预测误差、预测分歧度呈负相关关系。上述研究一致地发现，信息披露水平越高，分析师预测的准确度越高，分歧度越低。这些研究说明，无论是公共信息本身还是影响公共信息的外生环境或制度，均会对分析师预测行为产生显著影响。

人们都能观察到的、企业公开披露的信息是公共信息，但分析师对企业公开披露的信息进行不同的解读和理解后形成的认识、知识则可归为私有信息的范畴。胡奕明和林文雄（2005）通过调查研究发现后，将我国分析师获取的信息分为四种：公开披露信息、调研信息、间接信息和非正式信息。这意味着，调研信息、间接信息和非正式信息构成了私有信息。由于私有信息难以观察和度量，关于私有信息对分析师预测行为的影响研究较少。Chen and Jiang（2006）在研究分析师预测过程时发现分析师更多地依赖于私有信息做出盈余预测。Green *et al.*（2014）研究发现，证券分析师获得较多的私有信息有利于他们发布更加准确的预测报告。Mayew *et al.*（2013）检验分析师通过电话会议（earnings conference call）获取的私有信息对分析师预测准确度的影响，发现参加电话会议的分析师在会后发布预测报告更加及时和准确。他们认为参加

电话会议可以用于识别证券分析师是否拥有具有相对优势的私有信息。

分析师依赖于公共信息和私有信息进行预测，那么第一个问题是，私有信息和公共信息间的关系究竟如何？Verrecchia（1982）、Bushman（1991）的研究认为在完全竞争的市场上，信息披露减少了信息不对称，增加了投资者福利，所以公共信息和私有信息是替代关系。而 McNichols and Trueman（1994）、Lundholm（1988）则认为公共信息增加会导致私有信息增加，并没有增加信息不对称，因为一部分市场参与者利用公共信息作为获取私有信息的来源，这增加了他们搜集私有信息的动力，因此公共信息和私有信息是互补关系。王燕（2006）通过考察中国股票市场上信息披露的价格效应和流动性效应后发现，中国股票市场上公共信息和私有信息是互补关系。第二个问题是公共信息和私有信息对证券分析师预测行为的交叉影响如何？现有研究并未得到一致结论。Gintschel and Markov（2004）认为，如果公共信息披露规则有效，那么私有信息就会减少，分析师预测准确度会下降且预测分歧度会上升。De Jong and Apilado（2009）发现信息披露水平的提高使证券分析师拥有的私有信息减少甚至不再拥有私人信息，这有助于分析师预测准确度的提高。上述文献得出截然相反的结论，虽然他们都认为公共信息披露水平提高会使私有信息减少，但对预测准确度的影响，前者认为会下降，后者认为会提高。上述研究发现的不一致促使本文检验在中国这一特殊的资本市场背景下，公共信息和私有信息是如何对分析师预测产生联合影响的。

三、 研究假说和模型

（一）理论模型与研究假说

关于公共信息披露（Hope, 2003b; Langberg and Sivaramakrishnan, 2008; Libby *et al.*, 2006；白晓宇，2009）、私有信息（Chen and Jiang, 2006; Mayew *et al.*, 2013）对分析师预测的影响已有一些文献进行考察，在这些研究的基础上，本文重点考察分析师在同时面临公共信息和私有信息时，如何进行预测决策。借鉴 Barron *et al.*（1998）的理论分析过程，本研究讨论公共信息和私有信息对分析师预测行为的联合影响。

假设某一个公司由 N 个分析师跟踪，每个分析师要进行预测的盈余实际值定义为 y，y 符合均值为 $\bar{y}$、方差为 1/h（所以 h 表示准确度）的正态分布。通常情况下，分析师面临两种信息，一般信息（common information）和私有信息（private information）。假设一般信息等于以前的盈余实际值，因此其均值为 $\bar{y}$ 和准确度（precision）为 h。私有信息是指某个分析师可以观察到、但其他分析师无法观察到的信息，可以定义为 $z_i = y + \varepsilon_i$，假设 ε_i 符合均值为 0、方差为 $1/s_i$ 的独立同分布（iid），也允许每个分析师拥有的私有信息质量不同，即 $s_i s$ 在各个分析师之间是不同的。分析师基于拥有的一般信息和私有信息做出最优预测，因此其条件预期（conditional expectation）是一般信息和私有信息的加权平均值，权重为一般信息和私有信息的精度，定义某个分析师的最优预测值为 $\varepsilon_i \sim (0,1/s_i)$：

$$\mu_i \equiv E[y|z_i] = \frac{h\bar{y} + s_i z_i}{h + s_i} \tag{1}$$

公式（1）表明，分析师的最优预期值为过去实际盈余的平均值与基于私有信息所做出的盈余预测值的加权平均。公式（1）实质上揭示了分析师盈余预测的决策过程。首先，如果没有私有信息，分析师会根据过去实际盈余的平均值来预测当年的盈余；其次，在此基础上，分析师的私有信息有两种来源，一是对公开信息（public information）的理解和预测，每个分析师对公共信息的理解可能有差异，这也形成他们的私有信息，二是通过调研、分析获取的特有信息。这些私有信息是其他分析师无法观察的，被定义为 $z_i = y + \varepsilon_i$。$z_i = y + \varepsilon_i$ 表明，私有信息形成的预测是在一般信息预测结果（即 y）的基础上进行的一些修正（即ε_i）。无论是一般信息还是私有信息，都存在一定的波动性，因此用各自的精度作为权重来形成分析师的最优预测值 μ_i。

基于公式（1），可以计算每个分析师的平均预测值为

$$\mu \equiv \frac{1}{N}\sum_{i=1}^{N}\mu_i = \sum_{i=1}^{N}\frac{h\bar{y} + s_i z_i}{h + s_i} \tag{2}$$

分析师的预测分歧度定义为分析师预测值的方差 d，公式如下：

$$d \equiv \frac{1}{N-1}\sum_{i=1}^{N}(\mu_i - \mu)^2 \tag{3}$$

对于整个信息环境而言，所有样本的分析师预测分歧性则定义为特定公司分析师预测分歧度的非条件预期，公式如下：

$$D \equiv E[d] = \frac{1}{N-1}E\left[\sum_{i=1}^{N}(\mu_i - \mu)^2\right] = \frac{1}{N-1}\sum_{i=1}^{N}Var(\mu_i - \mu) \tag{4}$$

公式（4）说明，整个市场中所有分析师的预测分歧性是各个分析师预测值（μ_i）与平均预测值（μ）差异的方差之和。也就是说，当某个分析师根据一般信息和私有信息做出的盈利预测值，与所有分析师预测的平均值形成的差异，存在一定的波动性（即方差），将所有波动性累加起来，并除以参与的分析师数量，则形成整个市场的信息环境质量。但要注意的是，由于 μ_i 和 μ 都包含有一般信息 y，且 y 的均值为 $\bar{y}$，所以 D 更多地受到各个分析师获取的私有信息质量（即异质信息）的影响。因此，从整个信息环境看，D 可以看作分析师获取的私有信息质量的平均。

接下来讨论分析师的预测误差。分析师的预测误差可定义为盈余实际值（y）减去分析师预测平均值（μ），如下：

$$e \equiv y - \mu \tag{5}$$

对于整个样本来说，分析师预测误差定义为预测误差平方和的预期，如下：

$$SE \equiv E(e^2) = E[(y - \mu)^2] \tag{6}$$

本文重点关注 SE。

以上讨论的是分析师面临公共信息和私有信息时，所做的最优预测决策过程和结果，下面接着讨论信息环境如何影响分析师的预测结果。首先引入两个一般性的分析师信息环境特征变量：预测的不确定性程度（uncertainty）和一致性程度（consensus），这两个特征变量会影响分析师预测分歧度和预测误差。将不确定性定义为对实际盈余预测的预期方差，对于单个分析师而言，公式如下：

$$V_i \equiv E[(y-\mu_i)^2] = \frac{1}{h+s_i} \tag{7}$$

公式(7)意味着，对于单个分析师而言，基于一般信息和私有信息做出的盈余预测值，与实际盈余之间差异的方差，它等于一般信息精度与私有信息精度之和的导数。对于所有的分析师而言，信息环境不确定被定义为单个分析师不确定性的平均值，如下：

$$V \equiv \frac{1}{N}\sum_{i=1}^{N} V_i = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h+s_i} \tag{8}$$

公式（8）表明，分析师预测面临的所有不确定性是分析师预测值与实际盈余差异的方差，并在所有分析师之间求平均值，它是分析师面临的一般信息准确度和私有信息准确度的函数。

信息环境的一致性程度被定义为各个分析师预测值与实际值差异的协方差，公式如下：

$$C \equiv \frac{1}{N}\sum_{i=1}^{N} C_i = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N-1}\sum_{j\neq i}^{N} Cov\left(y-\mu_i, y-\mu_j\right)\right] \tag{9}$$

每个分析师做出的最优预测值与实际值都会产生差异，公式（9）表明的是，不同分析师预测差异的相关性。

根据上面的公式，可以推出 SE 的结果，如下：

$$V = \frac{1}{N}\sum_{i=1}^{N} V_i = \frac{1}{N}\sum_{i=1}^{N} E[(y-\mu_i)^2] = C + D \tag{10}$$

$$SE = C + \frac{D}{N} \tag{11}$$

上面公式中，V 指所有分析师盈余预测面临的总体不确定性，它等于单个分析师预测值与实际值差异的方差平均值；D 指单个分析师预测值与预测平均值差异的方差平均值；C 指各个分析师预测值与实际值差异的协方差的平均值，因此公式（10）表明的经济含义是，分析师预测的总体方差（或总体不确定性）由两部分构成：预测值与预

测均值差异的方差，预测值与实际值差异的协方差。

在公式（11）中，*SE* 为预测误差，指盈余实际值与所有分析师预测均值差异的方差，它等于各个分析师预测值与实际值差异的协方差（即 *C*）和一部分单个分析师预测值与预测平均值差异的方差（即 *D/N*）。根据前面的分析，*D* 更多受到单个分析师私有信息质量（或不确定性）的影响，反映的是单个分析师预测值与所有预测平均值（即一致预测）的差异程度，当单个分析师获得私有信息时，其预测值会与一致预测值产生差异，获得的私有信息越多、差异越大；而从经济含义上来说，*C* 表明的是分析师预测值与实际值的差异在各个分析师之间的一致程度，即分析师的预测是一致高于实际值、还是一致低于实际值、或者完全相反的程度，这反应了一般信息质量（common uncertainty）的影响，或者换句话说是所有分析师能够感受到的公共信息质量的影响。这说明，分析师平均预测误差受到公共信息不确定性和一部分异质信息（私有信息）的不确定性的影响。

基于上述理论模型，本文提出相应的实证模型。假设公共信息及质量定义为 *PubInf*，私有信息及质量（即异质信息）定义为 *PrvIinf*，那么根据公式（11），可构建如下回归模型：

$$SE = \alpha + \beta * PubInf + \gamma * PrvInf + \varepsilon \tag{12}$$

由于$PubInf$对应 *C*，$PrvInf$对应 *D*，于是可以推断出$\beta = 1$和 $\gamma = 1/N$。由于 N>1，[3] 因此可以合理预测$\beta > \gamma$，[4] 即公共信息质量带来的影响大于异质信息的影响，因此提出本文的研究假说：

H1：**公共信息对分析师预测误差的影响大于私有信息的影响；**

如果公共信息对分析师预测误差的作用大于私有信息，那么一个自然的问题是私有信息作用的发挥是否依赖于更好的公共信息环境。根据前面理论模型的论述，私有信息被定义为 $z_i = y + \varepsilon_i$，是在一般信息基础上的调整，每个分析师掌握的私有信息体现在 ε_i，这符合分析师预测决策的一般过程。公共信息（或一般信息）是分析师最容易获得的信息，因此，无论是对公司特征的具体分析，还是对公司的调研前，分析师都需要首先对公司一般信息进行解释、理解、加工和处理，这样有助于他们获得更有效的私有信息。McNichols and Trueman（1994）、Lundholm（1988）认为公共信息增加会导致私有信息增加，因为一部分市场参与者利用公共信息作为获取私有信息的来源。因此，给定分析师投入精力、时间成本固定的情况下，如果公共信息环境好，分析师基于的公共信息更多，也更可靠，那么分析师就能节省出时间和精力，并将这些

[3] N=1 在实践中也是可能的，即某些公司在特定年度只有 1 个分析师跟踪。如果只有 1 个分析师跟踪，就无法计算预测分歧度，正文中的模型（3）和模型（4）就无法成立，所以理论模型的假定还是 N>1。

[4] 本文的研究目的是考察公共信息、私有信息对分析师预测的影响哪个更强，以及如何对其产生联合影响，因此理论模型只是帮助推断 β 和 γ 的相对大小，因此在后面的实证检验中，本文主要检验和分析 β 是否大于 γ，并未对 $\beta = 1$ 和 $\gamma = 1/\mathrm{N}$ 进行检验和分析。

时间和精力投入到获取私有信息的过程中，从而提高私有信息的数量和质量，最终提高分析师的预测准确度。于是，可以合理预期，私有信息对分析师预测的影响在更好的公共信息环境中作用更大，从而提出本文的第二个研究假说：

H2：私有信息对分析师预测误差的影响在更好的公共信息环境中作用更明显。

（二）研究设计与回归模型

在具体讨论关键变量指标设定和回归模型构建之前，有两点需要说明：第一，本研究重点考察公共信息和私有信息对分析师预测误差的影响差异，在后续的实证检验中，除了检验对预测误差的影响，也检验了对分析师预测次数的影响，从而提供更多的证据来验证研究假说。公共信息与私有信息对分析师预测次数的影响逻辑与预测误差的逻辑一致。第二，在检验研究假说 H1 和 H2 之前，首先得检验公共信息与私有信息对分析师预测误差和预测次数的单独影响，如果本文选用的公共信息和私有信息的替代指标不对预测误差和预测次数产生影响，无法得到与以前研究一致的结论，就无法检验本文的研究假说。

关于私有信息的衡量，由于分析师获取私有信息的努力程度难以被外界观察到，本文用分析师所在券商机构与公司注册地之间的地理距离加以衡量。最近的一些研究开始关注地理位置或距离与经济参与者行为的关系，如地理距离对银行贷款的影响（Agarwal and Hauswald, 2010; García and Norli, 2012; Hauswald and Marquez, 2006; Petersen and Rajan, 2002），地理因素与股票回报的关系（García and Norli, 2012; Kim *et al.*, 2012），地理位置对公司行为如股利支付政策的影响（John *et al.*, 2011），地理因素对互助基金绩效的影响（Teo, 2009），地理位置对审计独立性的影响（DeFond *et al.*, 2008）。这些研究基于相同的逻辑，地理位置或邻近程度影响了经济参与者的信息优势（information advantage），他们可以较小的成本获取私有信息，从而节省交易成本。从地理位置的角度，一些研究者也考察过地理位置与分析师预测的关系，如 Bae *et al.*（2008）、王玉涛等（2010）考察过国内外分析师对本地公司（local firm）预测准确度的差异，结果发现，距离较近的本地分析师对本地公司的预测准确度要高于国外分析师。基于这些研究和逻辑，地理距离能够在一定程度上反映分析师获取私有信息的差异，因此本文使用这一指标替代私有信息。关于公共信息的衡量，本文利用一个综合指标 — 信息披露评价指数加以反映。深圳上市公司信息披露指数是反映公司信息披露质量的综合指标，能够在一定程度上反映分析师所处的公共信息环境的差异。方军雄（2007）、白晓宇（2009）检验过信息透明度与证券分析师预测的关系，均发现信息透明度越高，分析师预测准确度越高。基于上述分析，本文用分析师所在券商机构与上市公司注册地之间的距离（*Distance*）衡量私有信息，用上市公司信息披露指数 *DC_rating* 衡量公共信息，构建如下模型对上述研究假说进行检验。

$$AnlBeh_{i,j,t} = \alpha + \beta_1 DC\_rating_{i,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (13)$$

$$AnlBeh_{i,j,t} = \alpha + \beta_1 Distance_{i,j,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (14)$$

$$AnlBeh_{i,j,t} = \alpha + \beta_1 DC\_rating_{i,t} + \beta_2 Distance_{i,j,t} + \gamma_1 Ananum_{i,t} + \gamma_2 Gexp_{j,t} + \gamma_3 Fexp_{i,j,t} + \gamma_4 Brksize_{j,t} + \gamma_5 FH_{i,j,t} + \gamma_6 Intan_{i,t} + \gamma_7 Lev_{i,t} + \gamma_8 BM_{i,t} + \gamma_9 Size_{i,t} + Year/Ind + \varepsilon_{i,j,t} \quad (15)$$

模型（13）和（14）是对公共信息与私有信息的单独影响进行检验，模型（15）则对他们对分析师预测行为的联合影响进行检验。上述模型的被解释变量为分析师预测行为（*AnlBeh*），分别是分析师更新预测报告的次数（*Anlrptnum*）和预测误差（*Ferror*）。*Anlrptnum* 是以券商机构为单位，计算对跟踪的公司 *i* 在 *t* 年度内发布预测报告的次数，并取 Log 值；*Ferror* 反映预测误差，它等于分析师 *EPS* 预测值与实际值之差的绝对值，并除以前一年的公司总市值。由于券商机构 *j* 对公司 *i* 在 *t* 年度可能进行多次预测，因此取了平均值。以上两个指标均是每个公司每个年度每个分析师一个观测值。解释变量中，*Distance* 反映分析师 *j* 与公司 *i* 在 *t* 年度时的实际地理距离，并取自然对数，用于衡量私有信息。该值越小，表示地理距离越近，分析师拥有的私有信息越多。*DC_rating* 反映信息披露水平，该数据取自于深圳证券交易所信息披露指数，用于衡量公共信息。在实证分析中，分别用数值 1、2、3、4 代替 *D*（不及格）、*C*（及格）、*B*（良好）、*A*（优秀）四个档次的信息披露水平。该数值越小，表明公司的公共信息环境越差。因此，模型（13）用于检验公共信息对分析师预测行为的影响，模型（14）检验私有信息对分析师预测行为的影响，模型（15）检验本文的主要研究假说 H1，预期为 $\beta_1 > \beta_2$。对于研究假说 H2，基于模型（14），根据信息披露指数大小分为高低两组，检验在不同公共信息环境中私有信息对分析师预测行为的影响是否存在差异。

模型（13）至模型（15）使用了相同的控制变量。*Ananum* 指在 *t* 年度跟踪公司 *i* 的分析师数量，并取自然对数；*Gexp* 指分析师的从业经验，它等于分析师发布第一份报告所在年度与 *t* 年度之间的时间跨度，由于每个券商机构可能拥有多个分析师，所以以每个机构拥有的分析师从业经验的平均值来衡量券商机构的从业经验长短，并取其自然对数；*Fexp* 指 *t* 年度券商机构 *j* 跟踪公司 *i* 的特定经验，以机构 *j* 对公司 *i* 发布第一份报告所在年度与 *t* 年度之间的时间跨度来衡量，与 *Gexp* 的方法一致，也取了多个证券分析师的平均值，并取其自然对数；*Brksize* 指券商机构的规模，根据其拥有的分析师数量计算，并取自然对数；*FH* 指分析师出具盈利预测报告日离当年度定期报告披露日的天数，并取自然对数；*Intan*、*Lev*、*BM*、*Size* 为公司特征变量，*Intan* 为无形资产占总资产的比例，*Lev* 等于总负债除以净资产，*BM* 是净资产账面值与市场价值之比，*Size* 指公司 *i* 在 *t* 年度的权益总市值的自然对数；*Year/Ind* 为年度和行业 *dummy* 变量。

表 1 变量定义表

变量	定义
Ferror	分析师预测的每股收益减去实际 *EPS* 之差的绝对值，乘以 100 并除以期初总市值
Anlrptnum	每个券商机构每年对每个公司发布预测的次数
Distance	企业注册地与券商注册地的距离,取自然对数后取负值，该值越大，地理距离越近，拥有的私有信息越多
DC_rating	信息披露指数，取自然对数
Ananum	每个公司每年跟踪的分析师数量，取自然对数
Gexp	分析师的一般经验，是指分析师发布第一份预测报告的年度至样本期之间的年数，并基于券商机构求平均值后取自然对数
Fexp	分析师跟踪特定公司的经验，是指分析师对某一公司跟踪的年度起至样本期的年数，并基于券商机构求平均值后取自然对数
Brksize	根据每年拥有的分析师人数计算的券商机构规模，取自然对数
FH	分析师盈余预测报告日离年报公布日的天数，并取自然对数
Intan	无形资产占总资产的比率
Lev	总负债除以净资产
BM	净资产账面值与权益市值之比
Size	公司权益总市值的自然对数

四、 实证分析

（一）样本选择和描述性统计

本研究以 2002-2011 年的所有分析师预测数据为最初样本，涉及深市上市公司 1,583 家。[5] 因为分析师会对未来几年的公司盈余进行预测，于是分析师对每个公司每个年度的预测为 11,245 个观察值。而分析师会对所跟踪公司每年的收益进行多次预测，因此每个分析师对每个公司每一年度的预测为 62,330 个观察值。同时跟踪一家上市公司可能有多个分析师，所有分析师对每个公司每个年度的每一次预测为 130,170 个观察值。在此基础上，进行了如下筛选：（1）删除 B 股公司；（2）借鉴 Cuijpers and Buijink（2005）的研究，将分析师对上市公司进行预测的时间限制在年报公布日前半年内（180 天），在这期间涉及 4,632 个公司年观测，某一分析师会对同一上市公司进行多次预测，得到 35,094 个公司-年-分析师的观测值，而同一上市公司可能会吸引多个分析师跟踪，得到 62,194 个预测值；（3）删除金融保险业；（4）删除相关变量的缺失值，最终得到 21,901 个观察值。所有财务数据和指标来源于 CSMAR 数据库。在此基础上，对所有连续变量上下 1%的观察值进行了 winsorize 处理。样本筛选的详细过程可以见表 2。

[5] 由于上市公司信息披露指数只有深圳证券交易所要求披露，所以样本公司仅限制在深圳上市公司范围内。

表 2　样本筛选

样本筛选过程	观察值	公司-年-分析师	公司-年	公司
选择 2002-2011 年分析师深市数据	130,170	62,330	11,245	1,583
删除 B 股	129,590	61,772	10,688	1,524
将分析师预测限制年报公布日前 180 天内	62,194	35,094	4,632	1,375
生成公司-年-分析师样本	35,094	35,094	4,632	1,375
删除金融保险业	34,369	34,369	4,575	1,365
删除相关变量缺失的样本	21,901	21,901	3,318	695

表 3　公司信息披露水平和地理距离的分布

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	合计
Panel A：信息披露指数的分布											
D（不及格）	3	5	1	2	5	5	1	1	6	6	35
	3.03	4	0.88	0.92	1.69	1.34	0.21	0.19	1.13	1.07	
C（及格）	28	26	15	36	70	96	90	98	106	121	686
	28.28	20.8	13.27	16.59	23.73	25.81	18.48	19.07	19.89	21.49	
B（良好）	54	71	78	138	171	218	326	341	347	362	2,106
	54.55	56.8	69.03	63.59	57.97	58.6	66.94	66.34	65.1	64.3	
A（优秀）	14	23	19	41	49	53	70	74	74	74	491
	14.14	18.4	16.81	18.89	16.61	14.25	14.37	14.4	13.88	13.14	
合计	99	125	113	217	295	372	487	514	533	563	3,318
Panel B：上市公司与券商机构注册地之间地理距离的分布											
<=500	29	47	36	168	158	273	521	374	394	276	2,276
	25.44%	20.43%	15.32%	20.44%	18.08%	19.50%	18.48%	18.94%	18.27%	15.38%	18.33%
>500 and <=1000	16	36	28	119	158	262	480	343	393	294	2,129
	14.04%	15.65%	11.91%	14.48%	18.08%	18.71%	17.03%	17.37%	18.23%	16.39%	17.14%
>1000 and <=2000	44	99	117	347	379	579	1,175	853	931	824	5,348
	38.60%	43.04%	49.79%	42.21%	43.36%	41.36%	41.68%	43.19%	43.18%	45.93%	43.06%
>2000 and <=3000	24	34	47	149	155	231	495	318	317	293	2,063
	21.05%	14.78%	20%	18.13%	17.73%	16.50%	17.56%	16.10%	14.70%	16.33%	16.61%
>3000	1	14	7	39	24	55	148	87	121	107	603
	0.88%	6.09%	2.98%	4.74%	2.75%	3.93%	5.25%	4.41%	5.61%	5.96%	4.86%
合计	114	230	235	822	874	1,400	2,819	1,975	2,156	1,794	12,419

表 3 的 Panel A 列示了我国上市公司信息披露水平的分布情况。2002 年，进行信息披露的 99 家上市公司中，披露水平由低到高所占比例分别为 3.03%、28.28%、54.55%、14.14%。2011 年进行信息披露的公司有 563 家，这一比例分别为 1.07%、21.49%、64.3%、13.14%。可以看出，我国大部分上市公司信息披露水平处于“良好”评级，而且

2002-2011 年间处于“及格”评级的比例明显下降，处于披露水平“良好”的比例明显上升，说明我国上市公司信息披露水平提高了。表 3 的 Panel B 列示了我国上市公司注册地与券商机构地理距离分布的情况。从表中可以看出，在 2002-2011 年间，两者的距离主要集中在 1000-2000 公里之间，占 43.06%。上市公司与券商地理距离在 500 公里以内的占 18.33%，500-1000 公里的占 17.14%，2000 公里以上的占 21.47%。这说明我国券商机构与上市公司的距离总体上存在一定分散性，这有助于区别分析师获取的私有信息差异。

表 4 的 Panel A 和 Panel B 分别列示了相关变量的描述性统计和相关系数矩阵。*Ferror* 表示分析师预测误差，均值为 1.955%，表明我国分析师平均预测误差占公司市值的近 2%。*Anlrptnum* 显示每个券商机构每年对公司进行预测的次数平均为 1.866 次。*Distance* 是券商机构与上市公司的地理距离，均值为 1,341 公里。*DC_rating* 是上市公司信息披露水平，均值为 3.097，表明大部分公司的信息披露水平在 B 级以上，这与表 3 结果类似。*Ananum* 的结果表明跟踪每家公司的分析师数量平均为 15.02 个。*Gexp* 表明样本期内分析师的一般经验为 3.49 年；*Fexp* 显示分析师跟踪某一特定公司的平均经验为 1.62 年；在样本期内，券商机构平均拥有 29.1 个分析师（*Brksize*）。

表 4　描述性统计及相关系数矩阵

Panel A：相关变量的描述性统计

变量	Obs	均值	标准差	25 分位值	中位值	75 分位值
Ferror	21,901	1.955	3.640	0.258	0.725	2.037
Anlrptnum	21,901	1.866	1.399	1.000	1.000	2.000
Distance（原始值）	21,901	1,341	879	741	1304	1890
DC_rating（原始值）	21,901	3.097	0.626	3.000	3.000	3.000
Ananum（原始值）	21,901	15.02	10.37	7.00	13.00	22.00
Size（百万元）	21,901	12958	19577	3039	6053	13470
Gexp（原始值）	21,901	3.49	2.06	2.00	3.00	5.00
Fexp（原始值）	21,901	1.62	1.13	1.00	1.00	2.00
Brksize（原始值）	21,901	29.10	14.30	18.00	29.00	38.00
Intan	21,901	0.041	0.044	0.013	0.030	0.054
Lev	21,901	0.138	0.155	0.012	0.068	0.240
BM	21,901	0.354	0.257	0.171	0.280	0.465
FH	21,901	264.84	90.48	196.00	263.00	339.50

表 4 的 Panel B 列示了变量的相关系数矩阵，右上角为 *Pearson* 相关系数，左下角为 *Spearman* 相关系数。从表中可以看出，地理距离 *Distance* 与预测误差 *Ferror* 之间存在负相关关系，说明上市公司与券商间的地理距离越近，预测误差越小；地理距离 *Distance* 与预测次数 *Anlrptnum* 之间是正相关关系（*Spearman* 系数不显著），说明上市

表 4 描述性统计及相关系数矩阵

Panel B：相关系数矩阵

变量	*Ferror*	*Anlrptnum*	*Distance*	*DC_rating*	*Ananum*	*Size*	*Gexp*	*Fexp*	*Brksize*	*Intan*	*Lev*	*BM*	*FH*
Ferror		-0.01	-0.028***	-0.051***	-0.047***	-0.111***	0.009	0.028***	0.008	-0.045***	0.193***	0.399***	0.262***
Anlrptnum	0.009		0.035***	0.081***	0.252***	0.189***	0.191***	0.248***	0.181***	0.005	0.005	-0.015**	0.084***
Distance	-0.028***	0.008		0.021***	-0.028***	-0.001	0.014**	0.032***	0.009	-0.003	-0.014**	0.004	0.006
DC_rating	-0.017**	0.089***	0.006		0.293***	0.212***	0.021***	0.124***	0.012*	-0.059***	-0.001	0.062***	0.023***
Ananum	-0.043***	0.276***	-0.029***	0.298***		0.576***	0.191***	0.269***	0.049***	-0.034***	0.005	-0.077***	0.079***
Size	-0.125***	0.19***	-0.051***	0.222***	0.587***		0.148***	0.268***	0.035***	-0.071***	0.167***	-0.182***	-0.025***
Gexp	0.048***	0.198***	0.005	0.021***	0.189***	0.154***		0.456***	0.274***	0.04***	0.034***	-0.04***	0.174***
Fexp	0.055***	0.274***	-0.006	0.126***	0.281***	0.263***	0.436***		0.22***	-0.008	0.073***	0.025***	0.17***
Brksize	0.015**	0.205***	0.016**	0.011	0.06***	0.052***	0.295***	0.239***		0.025***	-0.003	-0.026***	0.083***
Intan	-0.031***	0.018***	0.032***	-0.072***	0.002	-0.074***	0.07***	0.016**	0.046***		-0.071***	-0.107***	0.02***
Lev	0.205***	0.005	-0.055***	-0.004	-0.007	0.155***	0.039***	0.071***	0.003	-0.05***		0.257***	0.03***
BM	0.44***	-0.024***	0.005	0.079***	-0.1***	-0.251***	-0.041***	0.001	-0.045***	-0.102***	0.301***		0.145***
FH	0.386***	0.023***	-0.016**	0.017**	0.061***	-0.034***	0.158***	0.184***	0.084***	0.028***	0.037***	0.153***	

注：（1）观察值为 21, 901；（2）右上角为 Pearson 相关系数，左下角为 Spearman 相关系数；（3）***、**、*分别表示在 1%、5%、10%水平上显著。

公司与券商间的地理距离越近，分析师更新预测报告的次数越多。信息披露水平 *DC_rating* 与预测误差 *Ferror* 间是负相关关系，说明信息披露水平越高，预测误差越小；信息披露水平 *DC_rating* 与预测次数 *Anlrptnum* 间是正相关关系，说明信息披露水平越高，分析师更新预测报告的次数越多。从其他变量之间的相关系数看，变量之间并不存在严重的多重共线性问题。

（二）公共信息、私有信息对分析师预测行为的独立影响

首先利用模型（13）检验公共信息带来的影响，即信息披露水平 *DC_rating* 对分析师预测次数 *Anlrptnum* 和预测误差 *Ferror* 的影响，结果列示于表 5。表 5 的 Panel A 列示了单变量检验结果，Panel B 列示了多元回归结果。Panel A 中，信息披露水平低组包括评级为 D 级（不及格）和 C 级（及格）的公司，水平高组包括评级为 B 级（良好）和 A 级（优秀）的公司，检验的结果显示，信息披露水平低时，*Ferror* 的均值为 2.35%，信息披露水平高时，*Ferror* 的均值为 1.89%，两者差异显著，说明信息披露水平越高，预测误差越低。同样地，信息披露水平低时，*Anlrptnum* 的均值为 1.7559，信息披露水平高时，*Anlrptnum* 的均值为 1.8845，而且两者差异显著，说明信息披露水平越高，分析师更新预测报告的次数越多。

表 5 信息披露水平对分析师预测特征的独立影响

Panel A：单变量检验

Variables	*DC_rating*	Obs.	Mean	Median	T-value	Z-value
Ferror	水平低	3,216	2.3504	0.7833	5.52***	4.12***
	水平高	18,685	1.8868	0.7143		
Anlrptnum	水平低	3,216	1.7559	1.0000	-5.09***	-5.49***
	水平高	18,685	1.8845	1.0000		

Panel B：多元回归结果

Variables	*Ferror*	*Anlrptnum*
DC_rating	-1.0169***	0.0093
	(-3.07)	(0.61)
Ananum	-0.4212***	0.0674***
	(-3.86)	(12.74)
Size	0.1121	0.0339***
	(1.31)	(7.29)
Gexp	0.0329	0.0369***
	(0.60)	(5.84)
Fexp	0.1628*	0.1535***
	(1.66)	(13.61)
Brksize	-0.0826**	0.0968***
	(-1.98)	(17.79)
Intan	-0.6574	0.0925
	(-0.55)	(1.09)
Lev	3.2283***	-0.0331
	(4.64)	(-1.38)
BM	3.9314***	-0.0318**
	(6.36)	(-1.99)
FH	1.8827***	0.0173***
	(17.34)	(2.85)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.278	0.148

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

在 *DC_rating* 对 *Anlrptnum* 和 *Ferror* 的多元回归中，控制了其他因素的影响，包括公司规模（*Size*）、券商规模（*Brksize*）等。由回归结果可以看出，信息披露水平（*DC_rating*）对预测误差（*Ferror*）的影响显著为负，这一结果说明，信息披露水平越高，预测误差越小，与以前的研究发现一致（白晓宇，2009；方军雄，2007）。另外，信息披露水平越高，分析师预测次数（*Anlrptnum*）越多，但差异不显著。

表 6　地理距离对分析师预测特征的独立影响

Panel A：单变量检验

Variables	Distance	Obs.	Mean	Median	T-value	Z-value
Ferror	距离近	11,409	1.8984	0.7078	-2.39**	-2.65***
	距离远	10,492	2.0163	0.7412		
Anlrptnum	距离近	11,409	1.8833	1.0000	1.95*	0.75
	距离远	10,492	1.8465	1.0000		

Panel B：多元回归结果

Variables	*Ferror*	*Anlrptnum*
Distance	-0.0333**	0.0054***
	(-2.21)	(3.05)
Ananum	-0.5131***	0.0687***
	(-4.89)	(13.30)
Size	0.1091	0.0338***
	(1.32)	(7.30)
Gexp	0.0442	0.0365***
	(0.81)	(5.77)
Fexp	0.1307	0.1527***
	(1.36)	(13.42)
Brksize	-0.0843**	0.0968***
	(-2.01)	(17.86)
Intan	-0.4705	0.0877
	(-0.40)	(1.02)
Lev	3.3017***	-0.0308
	(4.71)	(-1.29)
BM	3.8425***	-0.0315**
	(6.25)	(-2.00)
FH	1.8877***	0.0173***
	(17.23)	(2.84)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.275	0.149

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

利用模型（14）检验私有信息的影响，即地理距离 *Distance* 对分析师预测次数 *Anlrptnum* 和预测误差 *Ferror* 的影响，结果列示于表 6。表 6 的 Panel A 列示了单变量检验结果，Panel B 列示了多元回归结果。从单变量检验结果来看，当距离近时，*Ferror* 的均值为 1.90%，距离远时，*Ferror* 的均值为 2.02%，两组之间的差异在 5%水平上显著，说明距离越近，分析师的预测误差越小。同时，当距离近时，*Anlrptnum* 的均值为 1.8833，距离远时，*Anlrptnum* 的均值为 1.8465，两者差异在 10%水平上显著，说明距离越近，分析师更新预测报告的次数越多。

在 Panel B 中，控制了公司规模（*Size*）、券商规模（*Brksize*）等因素的影响后，利用模型（14）进行检验，结果发现：地理距离 *Distance* 对分析师预测误差 *Ferror* 的影响显著为负，对 *Anlrptnum* 的影响显著为正。这意味着距离越近，分析师获取私有信息的成本越低，预测误差越小，更新预测报告的次数越多，与以前的研究发现一致（Malloy, 2005）。

（三）公共信息、私有信息对分析师预测行为的联合影响

当分析师同时面临公共信息和私有信息时，分析师的预测如何受到两种信息的联合影响，是本文考察的重点。本研究采取两种方法检验：第一，将公共信息和私有信息的指标纳入同一个研究模型进行分析，以检验研究假说 1；第二，根据信息披露指数分组，考察私有信息带来的影响在不同公共信息环境中是否存在差异，以检验研究假说 2。利用模型（15）进行回归的结果列示在表 7。表 7 第 2 列为预测误差，第 3 列为预测次数。对于预测误差的影响，*DC_rating* 和 *Distance* 的影响显著为负，分别与表 5 和表 6 相同；通过 F 检验对比两者系数的影响（见本表最后一行），结果发现差异性显著（F 值为 8.88，在 1%水平下显著）。[6] 这说明，信息披露指数越大、地理距离越近，分析师预测误差越小，且信息披露指数衡量的公共信息对预测误差的影响显著大于私有信息的影响，支持了本文的研究假说 1。

根据上面所述的研究设计，将信息披露指数分为两组，一组包含 A 和 B 评级，公司信息披露质量为优秀和良好，定义为信息环境好的组，一组包含 C 和 D 评级，公司信息披露质量为及格和不及格，定义为信息环境较差组。[7] 基于公共信息较好和较差组，利用模型（14）对私有信息进行分析，结果列于表 8。表 8 Panel A 列示了不同信息披露水平下，距离远近对分析师预测行为的影响；Panel B 则列示了多变量回归结果。Panel A 显示，在信息披露水平低的外部环境下，距离近和距离远的预测误差（*Ferror*）均值分别为 2.35%和 2.35%，差异并不显著，这意味着当信息披露水平低时，地理距离对分析师预测误差的影响基本没有差异；而距离近和距离远的预测次数（*Anlrptnum*）分别为 1.6887 和 1.8214，差异是显著的，这表明，在信息披露水平低时，距离近带来

[6] 根据前面的模型（11）和模型（12），可以推断 *DC_rating* 的影响系数为 1。虽然本文检验的重点是对比 *DC_rating* 和 *Distance* 的影响程度差异，但为了增强理论模型推断的可靠性，也对 *DC_rating* 的系数（-1.0136）是否等于-1 进行了检验，相应的 F 值等于 0.00，相应的 P 值为 0.97，因此无法拒绝原假说。这一结果支持了理论模型的推断。

[7] 本文没有将不及格组定义为信息环境差组、将其他评级定义为信息环境好组，原因是这种分类导致信息环境差组的观察值太少，仅有 58 个，占总体样本的 0.26%。

表 7　公共信息与私有信息对分析师预测的联合影响

变量	*Ferror*	*Anlrptnum*
DC_rating	-1.0136***	0.0088
	(-3.07)	(0.58)
Distance	-0.0321**	0.0054***
	(-2.16)	(3.05)
Ananum	-0.4242***	0.0679***
	(-3.90)	(12.82)
Size	0.1128	0.0338***
	(1.32)	(7.31)
Gexp	0.0345	0.0366***
	(0.63)	(5.79)
Fexp	0.1700*	0.1523***
	(1.73)	(13.51)
Brksize	-0.0826**	0.0968***
	(-1.98)	(17.85)
Intan	-0.6371	0.0891
	(-0.54)	(1.04)
Lev	3.2095***	-0.0300
	(4.62)	(-1.25)
BM	3.9340***	-0.0323**
	(6.37)	(-2.04)
FH	1.8825***	0.0173***
	(17.35)	(2.85)
Year Control	Yes	Yes
Industry Control	Yes	Yes
Observations	21,901	21,901
Adj. R^2	0.278	0.149
DC_rating vs. *Distance* (F value)	8.88***	0.05

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

的预测次数反而更低，与预期相反。在信息披露水平高的外部环境下，距离近时预测误差（*Ferror*）均值为 1.82%，预测次数（*Anlrptnum*）均值为 1.9147；距离远时 *Ferror* 均值为 1.96%，*Anlrptnum* 均值为 1.8511，两者均具有显著差异。这一结果表明，当公司处于较好的信息披露环境时，距离越近，分析师更新预测报告的次数越多，预测误差越小。

在控制其他因素的影响后，表 8 Panel B 提供了基于不同信息披露水平下多变量回归结果。Panel B 的第 2 至 3 列为地理距离与预测误差的关系，第 4 至 5 列为地理距离与预测次数的关系。结果显示，在信息环境较差时，*Distance* 对预测误差的影响系数为 -0.0236，不显著；信息环境好时，影响系数为 -0.0315，且在 1%水平上显著。F

检验表明，两组回归系数的差异在 10%水平下是显著的。对于预测次数，信息环境较差时，地理距离对预测次数的影响系数为 0.0011，不显著；信息环境好时影响系数为 0.0060，且在 1%水平上显著。F 检验表明，两组回归系数的差异不显著。无论是单变量结果还是多变量结果，均支持了本文的研究假说 2，私有信息对预测误差的影响在公共信息较好的情况下有作用。

表 8　不同信息披露水平下地理距离对分析师预测特征的影响

Panel A：**单变量分析**

分类	项目	Obs.	预测误差（*Ferror*）		预测次数（*Anlrptnum*）	
			Mean	Median	Mean	Median
信息环境差	距离近	1,587	2.3536	0.7676	1.6887	1.0000
	距离远	1,629	2.3473	0.8078	1.8214	1.0000
	近–远		0.0063	-0.0402*	-0.1326***	-0.0000**
信息环境好	距离近	9,822	1.8248	0.6978	1.9147	1.0000
	距离远	8,863	1.9554	0.7322	1.8511	1.0000
	近–远		-0.1306**	-0.0344**	0.0636***	0.000*

Panel B：**多变量分析**

Variables	*Ferror*		*Anlrptnum*	
	信息披露水平低	信息披露水平高	信息披露水平低	信息披露水平高
Distance	-0.0236	-0.0315**	0.0011	0.0060***
	(-0.60)	(-2.01)	(0.35)	(3.14)
Ananum	-0.1704	-0.4764***	0.0546***	0.0716***
	(-0.61)	(-4.16)	(4.86)	(12.29)
Size	-0.0424	0.1480*	0.0434***	0.0321***
	(-0.18)	(1.68)	(4.59)	(6.21)
Gexp	0.3042**	-0.0014	0.0229	0.0389***
	(2.03)	(-0.02)	(1.40)	(5.60)
Fexp	0.4526	0.1164	0.1618***	0.1513***
	(1.52)	(1.18)	(5.59)	(12.53)
Brksize	-0.1011	-0.0830**	0.0715***	0.1011***
	(-0.59)	(-2.19)	(5.78)	(16.90)
Intan	1.6130	-0.7619	0.2518	0.0410
	(0.51)	(-0.60)	(1.55)	(0.42)
Lev	5.2347**	2.8931***	-0.0933***	-0.0219
	(2.47)	(4.18)	(-2.70)	(-0.80)
BM	5.1766***	3.8365***	-0.0101	-0.0308*
	(3.05)	(5.86)	(-0.30)	(-1.82)
FH	2.0846***	1.8325***	0.0214*	0.0160**
	(7.15)	(16.24)	(1.76)	(2.32)
Year Control	Yes	Yes	Yes	Yes
Industry Control	Yes	Yes	Yes	Yes
Observations	3,216	18,685	3,216	18,685
Adj. R^2	0.294	0.281	0.153	0.148

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

（四）稳健性检验

本研究进行了以下几方面的稳健性检验：

1、公共信息对分析师预测误差的影响（详见表 5），采用了上市公司信息披露指数来衡量分析师面临的公共信息环境。由于信息披露指数是一个综合指标，受到多个因素的影响，包括分析师跟踪和预测，因此会导致内生性问题。为解决这一问题，采用两阶段最小二乘法对表 5 进行进一步分析。对于信息披露指数的影响因素，则借鉴了 Hope（2003a）的研究，选择公司层面的特征变量，包括公司规模（*Size*）、盈利能力（*ROE*）、财务杠杆（*Lev*）、分析师跟踪（*Ananum*）、审计质量（*Auditor*）、政

表 9　公共信息的内生性及两阶段回归结果

Variables	Panel A		Panel B
	Ferror	*Anlrptnum*	*DC_rating*
DC_rating	-11.5335***	0.4362***	
	(-3.01)	(2.90)	
Ananum	0.5079	0.0296**	0.0700***
	(1.47)	(2.13)	(5.70)
Size	0.1496	0.0325***	0.0254*
	(0.70)	(3.97)	(1.78)
Gexp	-0.0904	0.0405***	
	(-1.16)	(5.96)	
Fexp	0.5622***	0.1375***	
	(2.72)	(10.67)	
Brksize	-0.0604	0.0968***	
	(-1.18)	(17.25)	
Intan	-2.3244	0.1649	
	(-0.85)	(1.34)	
Lev	2.3415**	0.0044	-0.0099
	(2.30)	(0.11)	(-0.14)
BM	4.8226***	-0.0699***	
	(6.65)	(-2.75)	
FH	1.8147***	0.0203***	
	(15.88)	(3.12)	
Year Control	Yes	Yes	
Industry Control	Yes	Yes	
Constant	-4.3601	-0.9779***	0.4450
	(-0.98)	(-5.38)	(1.47)
ROE			0.0335
			(0.30)
Auditor			0.0734**
			(2.23)
SOE			0.0143
			(0.71)
Inst_holding			0.0989***
			(2.84)
Observations	21,685	21,685	21,685
Adj. R^2	0.282	0.146	0.093

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

府控制（*SOE*）和机构投资者持股比例（*Inst_holding*），构建第一阶段模型。除了 *Size*、*Lev*、*Ananum* 以外，其他变量为新增的外生影响因素，*ROE* 是净资产回报率；*Auditor* 为虚拟变量，当公司审计师为 4 大会计师事务所时为 1，否则为 0；*SOE* 为虚拟变量，当公司实际控制人为中央政府或地方政府时为 1，否则为 0；*Inst_holding* 为机构投资者的持股比例。第一阶段模型的结果见表 9 的 Panel B，从中可以看出，分析师跟踪数量（*Ananum*）、公司规模（*Size*）、审计质量（*Auditor*）、机构投资者持股比例（*Inst_holding*）对信息披露指数存在正向显著影响，与预期一致。根据第一阶段模型预测的信息披露水平，进一步代入模型（13），形成新的回归结果，列示在表 9 的 Panel A，从中可以看出，*DC_rating* 对 *Ferror* 的影响显著为负，对 *Anlrptnum* 的影响显著为正，这说明，在解决了内生性之后，公共信息环境对分析师预测误差、预测次数的影响是与预期一致的。

2、私有信息对分析师预测行为的影响（表 6）中，也面临内生性问题。分析师做出跟踪某些公司的决策是内生的，因此会自然导致预测结果与信息获取之间的关系（Malloy, 2005）。在 Malloy（2005）的研究中，他们识别了 3 个导致内生性的可能性，包括（1）Local 分析师，尤其是那些注册地在较偏远地区的分析师，跟踪的公司较少，反而能够对跟踪公司给予更多的关注，从而有助于他们的预测。这一逻辑说明，地理距离较近对分析师预测误差的影响并不是从私有信息获取角度，而是从分析师的努力角度进行解释的；（2）Local 分析师比其他分析师在某些行业领域更加专业；（3）Local 分析师投入更多时间和精力分析客户公司的股票，因此其预测准确度的提高是由于 local 客户的需求导致的。本文仍采用两阶段最小二乘法解决内生性问题。对于影响地理距离的变量，参考 Malloy（2005）的分析，本研究选取公司规模（*Size*）、证券分析师所在券商机构规模（*Brksize*）、分析师跟踪的行业数量（*Indnum*）、审计质量（*Auditor*）、机构投资者持股比例（*Inst_holding*）、以及信息披露指数（*DC_rating*）。分析师券商规模越大，越容易出现在大城市，而小规模的券商更可能出现在偏远城市，这可以部分反映 Malloy（2005）关于内生性分析的第（1）点；同时，与公司规模（*Size*）结合在一起分析，可以决定公司与券商机构注册地的地理距离远近。之所以选择分析师跟踪的行业数量，是因为 Malloy（2005）关于内生性分析的第（2）点，这意味着，分析师跟踪的行业越多，可能导致分析师精力分散，同时也可能导致行业知识的互补，最终可能有助于或有损于私有信息的获取；选择机构投资者持股是因为 Malloy（2005）关于内生性分析的第（3）点，同时，机构投资者持股比例越高，也表明公司的信息环境越好，影响分析师获取的私有信息。另外，还选择了两个影响私有信息的公司信息环境指标：信息披露指数（*DC_rating*）[8] 和审计质量（*Auditor*）。*DC_rating* 和 *Auditor* 与前面的定义一致。基于这些变量，构建第一阶段的私有信息（*Distance*）影响模型，结果列于表 10 的 Panel B，从中可以看出，*Indnum* 与 *Auditor* 对私有信息的影响为正，

[8] 信息披露指数在本研究中反映为信息环境，它可能影响地理距离，因为根据本文的统计，很多券商机构注册地为北京、上海和深圳这些大城市，而这些大城市的公司信息披露、公司治理可能也很好，所以地理距离可能受到信息披露指数的影响。审稿人也同样指出这一点，在此表示感谢！

信息披露指数的影响不显著。Panel A 列示了考虑内生性后的私有信息对分析师预测行为的影响，结果显示，地理距离越近，分析师预测误差（*Ferror*）越小，预测次数（*Anlrptnum*）越多，与表 6 的结果一致。

表 10　私有信息内生性及两阶段回归结果

Variables	Panel A		Panel B
	Ferror	*Anlrptnum*	*Distance*
Distance	-1.2287**	0.0618**	
	(-2.13)	(2.08)	
Ananum	-0.5983***	0.0728***	
	(-4.55)	(11.16)	
Size	0.1330	0.0328***	-0.0400
	(1.27)	(5.95)	(-0.92)
Gexp	0.0816	0.0333***	
	(1.03)	(4.71)	
Fexp	0.3931**	0.1404***	
	(2.25)	(10.64)	
Brksize	-0.0775	0.0974***	0.0426
	(-1.31)	(17.39)	(1.30)
Intan	0.2720	0.0579	
	(0.18)	(0.54)	
Lev	2.6193***	0.0031	
	(3.27)	(0.10)	
BM	3.9243***	-0.0374**	
	(6.45)	(-2.16)	
FH	1.8618***	0.0187***	
	(16.02)	(2.85)	
Year Control	Yes	Yes	
Industry Control	Yes	Yes	
Constant	-22.0361***	-0.2045	-6.0685***
	(-4.48)	(-0.80)	(-6.61)
Indnum			0.0566*
			(1.68)
Auditor			0.5786***
			(3.46)
DC_rating			0.1357
			(0.74)
Inst_holding			-0.0887
			(-0.59)
Observations	21,685	21,685	21,685
Adj. R^2	0.278	0.146	0.009

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

3、本研究以信息披露指数来衡量公共信息，用上市公司注册地与分析师所在券商机构注册地的地理距离来衡量私有信息，但由于分析师所在券商机构大部分集中在北京、上海、深圳，[9] 而这些地区的上市公司公共信息环境较好，因此如果上市公司注

[9] 根据统计，北京、上海、深圳三大城市拥有的证券机构的数目合计占全国的 47%。

册地也在这三个城市，那么他们与券商机构的地理距离较近，这意味着，用地理距离替代私有信息的同时，也包含有一定的公共信息。这种影响尤其是在表 8 的分析中更重要。为了解决这一问题，本文采取两种方法加以解决，第一，将地理距离包含的公共信息进行剔除，即用地理距离对信息披露指数进行回归，得到地理距离的剩余（residual），然后在表 8 的基础上进行回归，结果列于表 11，与表 8 的结果一致。第二，将上市公司注册地在北京、上海和深圳的样本剔除，这样地理距离较近（尤其是 *Distance*=0）就不包含公共信息较好的情况，从而减弱变量设定误差的影响。删除样本后剩余 17,459 个观测值，也做了类似表 8 的结果（结果未列出），发现分析师预测误差的结果与表 8 一致，分析师预测次数在公共环境好的情况下为正但不显著。

表 11　稳健性检验

Variables	*Ferror*		*Anlrptnum*	
	信息披露水平低	信息披露水平高	信息披露水平低	信息披露水平高
Distance	0.0204	-0.0549**	0.0026	0.0125***
	(0.40)	(-2.05)	(0.50)	(3.57)
Ananum	0.0279	0.1264	0.0855***	0.0904***
	(0.17)	(0.75)	(6.15)	(9.85)
Size	-0.2156	-0.2919***	0.0202***	0.0180***
	(-1.10)	(-3.18)	(3.55)	(3.73)
Gexp	0.1565	-0.1588	0.0482***	0.0447***
	(0.52)	(-0.81)	(3.17)	(3.78)
Fexp	0.2608	-0.0005	0.1639***	0.1454***
	(0.58)	(-0.01)	(5.06)	(5.36)
Brksize	0.0399	0.0733	0.0667***	0.0975***
	(0.28)	(0.94)	(5.89)	(11.34)
Intan	3.2557*	0.4868	0.1966	-0.0139
	(1.79)	(0.56)	(1.51)	(-0.12)
Lev	6.4868**	2.6164*	-0.0402	0.0497
	(2.09)	(1.95)	(-1.43)	(1.60)
BM	7.1339***	4.5863***	0.0409***	0.0320
	(5.20)	(4.18)	(6.36)	(1.15)
FH	2.2207***	1.8580***	0.0292**	0.0305
	(3.37)	(3.15)	(2.20)	(1.43)
Year Control	Yes	Yes	Yes	Yes
Industry Control	Yes	Yes	Yes	Yes
Observations	3,216	18,685	3,216	18,685
Adj. R^2	0.328	0.263	0.148	0.146

注：括号内为回归系数的 t 值，结果均进行了 Cluster 调整；***、**、*分别表示在 1%、5%、10%水平上显著。

五、 研究结论与研究局限

（一）研究结论

公共信息和私有信息是影响分析师预测行为的两个重要因素，以前学者分开考察公共信息（Baldwin, 1984; Hodder *et al.*, 2008; Langberg and Sivaramakrishnan, 2008; Libby *et al.*, 2006）和私有信息（Chen and Jiang, 2006; Green *et al.*, 2014; Mayew *et al.*, 2013）对分析师预测行为的影响。与上述研究不同，本文分别以地理距离、信息披露水平作为私有信息和公共信息的衡量指标，重点考察了私有信息和公共信息对分析师预测行为的联合影响。实证结果显示，在公共信息与私有信息的单独影响中，本研究发现信息披露水平（公共信息）越高、地理距离（私有信息）越近，证券分析师发布预测报告的次数越多，预测误差越小。在信息披露水平和地理距离对分析师预测行为的联合影响中，本研究发现，信息披露水平对预测误差和预测次数的影响显著高于地理距离的影响；根据信息披露水平分组后，地理距离带来的影响仅在信息披露水平较高的组内出现。这说明，相对于私有信息，公共信息对分析师预测误差和预测次数的影响作用更大，私有信息的作用发挥依赖于更好的公共信息环境。

与以往单独考察公共信息和私有信息对分析师预测的独立影响不同，本文重点分析了两者的联合影响，补充了以前的研究成果。另外，本文的研究也具有现实指导意义。证券分析师在资本市场扮演重要的中介角色，能够减少上市公司与投资者之间的信息不对称。随着我国证券市场的逐步完善，公共信息环境的改善，证券分析师预测如何受到公共信息和私有信息的影响，是资本市场参与者都关心的重要问题。本文的研究能够回答这一重要问题，能够帮助投资者更好地理解证券分析师预测行为，有助于他们做出合理决策。

（二）研究局限

本文一个重要的研究局限是私有信息的衡量。现有文献很难找到一个合理反映分析师获取私有信息努力程度的指标。本文选用分析师所在券商注册地与上市公司注册地之间的地理距离来间接测量分析师获得私有信息的努力和成本，虽然有理论基础和以前研究的支持（Agarwal and Hauswald, 2010; García and Norli, 2012; Hauswald and Marquez, 2006; Kim *et al.*, 2012; Malloy, 2005; Petersen and Rajan, 2002），但不可避免存在许多噪音的影响。这种噪音影响表现在：1、地理距离可能也反映一部分公共信息质量；2、交通便利、交通工具的多样性使地理距离较远导致的信息搜寻成本较高的劣势减弱，可能并不能完全替代私有信息。[10] 虽然在稳健性测试中对第 1 点进行了探讨和检验，但却不能完全排除第 2 点因素对研究结论的影响。

[10] 感谢审稿人指出这一点。

参考文献

白晓宇，2009，“上市公司信息披露政策对分析师预测的多重影响研究”，《金融研究》第 4 期，92-112。

方军雄，2007，“我国上市公司信息披露透明度与证券分析师预测”，《金融研究》第 6 期，136-148。

胡奕明、林文雄，2005，“信息关注深度、分析能力与分析质量 — 对我国证券分析师的调查分析”，《金融研究》第 2 期，46-58。

李丹、贾宁，2009，“盈余质量、制度环境与分析师预测”，《中国会计评论》第 4 期，351-370。

王燕，2006，“中国股票市场公开信息与私有信息的互补效应”，《金融研究》第 6 期，41-52。

王玉涛、陈晓、侯宇，2010，“国内证券分析师的信息优势：地理邻近性还是会计准则差异”，《会计研究》第 12 期，34-40。

王玉涛、王彦超，2012，“业绩预告信息对分析师预测行为有影响吗？”，《金融研究》第 6 期，202-218。

Agarwal, S. and Hauswald, R. (2010), ‘Distance and Private Information in Lending’, *Review of Financial Studies* 23 (7): 2757-2788.

Bae, K., Stulz, R. M., and Tan, H. (2008), ‘Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts’, *Journal of Financial Economics* 88 (3): 581-606.

Baldwin, B. A. (1984), ‘Segment Earnings Disclosure and the Ability of Security Analysts to Forecast Earnings Per Share’, *The Accounting Review* 59 (3): 376-389.

Barron, O. E., Kim, O., Lim, S. C., and Stevens, D. E. (1998), ‘Using analysts’ forecasts to measure properties of analysts’ information environment’, *The Accounting Review* 73 (4): 421-433.

Behn, B. K., Choi, J. and Kang, T. (2008), ‘Audit Quality and Properties of Analyst Earnings Forecasts’, *The Accounting Review* 83 (2): 327-349.

Burgstahler, D. C. and Eames, M. J. (2003), ‘Earnings Management to Avoid Losses and Earnings Decrease: Are Analysts Fooled?’, *Contemporary Accounting Research* 20 (2): 253-294.

Bushman, R. M. (1991), ‘Public Disclosure and the Structure of Private Information Markets’, *Journal of Accounting Research* 29 (2): 261-276.

Chen, Q. and Jiang, W. (2006), ‘Analysts’ Weighting of Private and Public Information’, *Review of Financial Studies* 19 (1): 319-355.

Cuijpers, R. and Buijink, W. (2005), ‘Voluntary Adoption of Non-Local GAAP in the European Union: A Study of Determinants and Consequences’, *European Accounting Review* 14 (3): 487-524.

De Jong, P. J. and Apilado, V. P. (2009), 'The changing relationship between earnings expectations and earnings for value and growth stocks during Reg FD', *Journal of Banking and Finance* 33 (2): 435-442.

DeFond, M., Francis, J. R., and Hu, X. (2008), 'The Geography of Auditor Independence and SEC Enforcement', Working Paper, University of Southern California.

García, D. and Norli, Ø. (2012), 'Geographic dispersion and stock returns', *Journal of Financial Economics* 106 (3): 547-565.

Gintschel, A. and Markov, S. (2004), 'The effectiveness of regulation FD', *Journal of Accounting and Economics* 37 (3): 293-314.

Green, T. C., Jame, R., Markov, S., and Subasi, M. (2014), 'Access to Management and the Informativeness of Analyst Research', *Journal of Financial Economics* 114 (2): 239-255.

Hauswald, R. and Marquez, R. (2006), 'Competition and Strategic Information Acquisition in Credit Markets', *Review of Financial Studies* 19 (3): 967-1000.

Hodder, L., Hopkins, P. E., and Wood, D. A. (2008), 'The Effects of Financial Statement and Informational Complexity on Analysts' Cash Flow Forecasts', *The Accounting Review* 83 (4): 915-956.

Hope, O. K. (2003a), 'Disclosure Practices, Enforcement of Accounting Standards, and Analysts' Forecast Accuracy: An International Study', *Journal of Accounting Research* 41 (2): 235-272.

Hope, O. K. (2003b), 'Accounting Policy disclosures and Analysts' Forecasts', *Contemporary Accounting Research* 20 (2): 295-321.

John, K., Knyazeva, A., and Knyazeva, D. (2011), 'Does geography matter? Firm location and corporate payout policy', *Journal of Financial Economics* 101 (3): 533-551.

Kim, C. F., Pantzalis, C., and Park, J. C. (2012), 'Political geography and stock returns: The value and risk implications of proximity to political power', *Journal of Financial Economics* 106 (1): 196-228.

Lang, M. H. and Lundholm, R. J. (1996), 'Corporate Disclosure Policy and Analyst Behavior', *The Accounting Review* 71 (4): 467-492.

Langberg, N. and Sivaramakrishnan, K. (2008), 'Voluntary disclosures and information production by analysts', *Journal of Accounting and Economics* 46 (1): 78-100.

Libby, R., Tan, H., and Hunton, J. E. (2006), 'Does the Form of Management's Earnings Guidance Affect Analysts' Earnings Forecasts?', *The Accounting Review* 81 (1): 207-225.

Lundholm, R. J. (1988), 'Price-Signal Relations in the Presence of Correlated Public and Private Information', *Journal of Accounting Research* 26 (1): 107-118.

Malloy, C. J. (2005), 'The Geography of Equity Analysis', *Journal of Finance* 60 (2): 719-755.

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.

McEwen, R. A. and Hunton, J. E. (1999), 'Is Analyst Forecast Accuracy Associated With Accounting Information Use?', *Accounting Horizons* 13 (1): 1-16.

McNichols, M. and Trueman, B. (1994), 'Public disclosure, private information collection, and short-term trading', *Journal of Accounting and Economics* 17 (1-2): 69-94.

O'Brien, P. and Bhushan, R. (1990), 'Analyst Following and Institutional Ownership', *Journal of Accounting Research* 28 (3): 55-82.

Pankoff, L. D. and Virgil, R. L. (1970), 'Some Preliminary Findings from a Laboratory Experiment on the Usefulness of Financial Accounting Information to Security Analysts', *Journal of Accounting Research* 8 (3): 1-48.

Payne, J. L. (2008), 'The Influence of Audit Firm Specialization on Analysts' Forecast Errors', *Auditing* 27 (2): 109-136.

Petersen, M. A. and Rajan, R. G. (2002), 'Does Distance Still Matter? The Information Revolution in Small Business Lending', *Journal of Finance* 57 (6): 2533-2570.

Robertson, J. C. (1988), 'Analysts' Reaction to Auditors' Messages in Qualified Reports', *Accounting Horizons* 2 (2): 82-89.

Teo, M. (2009), 'The Geography of Hedge Funds', *Review of Financial Studies* 22 (9): 3531-3561.

Vergoossen, R. G. A. (1993), 'The use and perceived importance of annual reports by investment analysts in the Netherlands', *European Accounting Review* 2 (2): 219-243.

Verrecchia, R. E. (1982), 'Information Acquisition in a Noisy Rational Expectations Economy', *Econometrica* 50 (6): 1415-1430.

Wang, Y., Chen, Y., and Wang, J. (2015), 'Management earnings forecasts and analyst forecasts: Evidence from mandatory disclosure system', *China Journal of Accounting Research* 8 (2): 133-146.

Yu, F. F. (2008), 'Analyst coverage and earnings management', *Journal of Financial Economics* 88 (2): 245-271.