Does soft information matter for financial analysts' forecasts? A gravity model approach

Régis Breton, Sébastien Galanti, Christophe Hurlin, Anne-Gaël Vaubourg

Does soft information matter for financial analysts' forecasts? A gravity model approach

Régis BRETON, Sébastien GALANTI, Christophe HURLIN et Anne-Gaël VAUBOURG

Résumé

Cet article évalue dans quelle mesure le souci des analystes financiers de maintenir de bonnes relations avec les entreprises, afin de conserver un accès à de l'information financière privée, les conduit à émettre des prévisions optimistes ou pessimistes. On utilise une approche en terme de modèles de gravité des relations entreprises-analystes par une régression en panel. Les données portent sur les prévisions à un an des bénéfices par actions par 4648 analystes sur 241 entreprises françaises (1997-2007). On trouve que plus l'effet-couple est faible (fort), plus l'erreur de prévision est faible (forte). Ceci accrédité l'idée selon laquelle le pessimisme ou l'optimisme des prévisions découle du souci de conserver un accès à l'information privée révélée par les dirigeants d'entreprise.

Codes JEL Classification : C58, D84, G17, G24

Mots-Clés : analystes financiers, prévisions de bénéfices, anticipation, information privée, régression en panel, modèle de gravité.

Does soft information matter for financial analysts' forecasts? A gravity model approach

Régis Breton[*], Sébastien Galanti[†], Christophe Hurlin[‡] and Anne-Gaël Vaubourg[§]

Abstract

We study whether the financial analysts' concern to maintain good relationships with firms' managers in order to preserve their access to 'soft' qualitative information entice them to issue pessimistic or optimistic forecasts. We use a gravity model approach to firms-analysts relationships and propose a measure of soft information. Our database contains the one-year ahead EPS forecasts issued by 4 648 analysts about 241 French firms (1997-2007). We find that a low (high) pair-effect is associated with a low (high) forecast error. This suggests that pessimism and optimism result from analysts' concern to preserve access to soft information released by managers.

JEL Classification : C58, D84, G17, G24

Keywords: financial analysts, earnings forecasts, expectations, soft information, panel regression, gravity models.

[*]Banque de France, CNRS-Laboratoire d'Economie d'Orlans (LEO)

[†]University of Orléans, Laboratoire d'Economie d'Orléans (LEO). *Corresponding author*: Sebastien.Galanti@univ-orleans.fr. *Contact*: Faculté de Droit d'Economie et de Gestion, Rue de Blois BP 26739, F-45067, Orléans, France; Tel +33 2 38 41 73 62, Fax +33 2 38 41 73 80

[‡]University of Orléans, Laboratoire d'Economie d'Orléans (LEO)

[§]University of Orléans, Laboratoire d'Economie d'Orléans (LEO)

1 Introduction

The role of financial analysts as information producers is crucial for financial markets. By issuing forecasts about the value of firms' share or earnings per share (EPS), they contribute to alleviate information asymmetries between firms and investors or fund managers. Mainly issued on behalf of brokers, their forecasts and selling or buying recommendations are widely used by fund managers when taking portfolio allocation decisions. However, many studies have shown that the earnings forecasts by analysts can be inaccurate (Brown, 1997), thus increasing corporate agency costs, and worsening the informational efficiency of financial markets.

A first set of litterature shows that analysts forecasts are excessively optimistic[1]. There exist at least two reasons for this. On the one hand, some firms may refuse to contract with an investment bank if 'sell-side' analysts do not issue favourable forecasts about its EPS (Barber, Lehavy & Trueman 2007, Dugar & Nathan 1995, Hayward & Boeker 1998, O'Brien, Hsiou-Wei & McNichols 2005). On the other hand, it is in the interest of the analyst to release optimistic recommendations or EPS forecasts in order to generate buying transactions (Jackson, 2005).

A second set of papers suggest that some analysts' EPS forecasts turn out to be pessimistic, i.e. below, instead of above, the actual EPS. To explain such a behavior, the literature mainly refers to the 'earning management strategy'. It consists in the discretionary manipulation of earnings by firm managers such that the actual EPS finally appears higher than the forecast (Payne & Robb, 2000, Matsumoto, 2002, Burgstahler & Eames, 2006). The goal of this strategy is to create a 'positive earning surprise' on financial markets when the firm manager finally releases the amount of actual EPS. If the realized EPS appears higher than its forecast, investors on financial markets may have a favourable reaction, thus contributing to increase the price of firm shares.

But analysts not only produce forecasts. They also provide different kinds of qualitative

[1] Throughout this article, we will denote as optimistic a forecast that exceeds the realized value (forecast-earnings>0), and conversely denote as pessimitic a forecast that is below the realized value (forecast-earnings<0).

information about a given firm to their brokerage clients, mainly fund managers. To do this, they organize phone calls, one-to-one meetings and invite fund managers to conference calls with the firm's management (Fogarty & Rogers, 2005, Breton & Taffler, 2001). This access to so-called 'soft information' is possible only if the analyst has a good relationship with the firm management. In this context, issuing biased forecasts appears as a way for analysts to maintain their access to 'soft information' about firms (Libby, Hunton, Tan & Seybert, 2008, Pratt, 1993, Gibson, 1995, Womack, 1996 and Boni & Womack, 2002). If the firm management is mainly interested by 'beating the forecasts', the analyst tries to maintain a friendly relationship with the firm by issuing pessimistic forecasts. If, conversely, the management prefers a short-term increase in the firm's market value, the analyst is enticed to issue optimistic forecasts. The incentive to produce either pessimistic or optimistic forecasts also depends on the analyst himself. For example, some analysts may pay less attention than others to maintain a good relationship with a given firm because of weaker pressure from their customers (Yu, 2008) or because of any unobservable reason that makes them less sensitive to pressure . This suggests that the impact of soft information on forecast errors mainly passes through a firm-analyst effect. In other words, high (negative or positive) forecast errors should be associated with high firm-analyst pair effect.

The goal of this paper is precisely to check the empirical relevance of this theoretical idea by investigating whether soft information or close relationship between a firm and an analyst implies optimistic or pessimistic forecasts[2]. Soft information, which is qualitative and delivered only in case of friendly relationship with firm managers, is based on unobservable characteristics. It is thus a difficult task to provide a direct and explicit measure of the role played by soft information in the forecast process. Moreover, we need to distinguish unobservable factors steming from the firm, or from the analyst, or from a specific analys-firm pair. As far as we know, this has not yet been dealt with in the empirical literature on financial analysts' forecast.

[2]In this article, we do not consider insider information (i.e. quantitative, precise, timely information which is likely to affect prices). It is worth noting that firms executives are generally very cautious not to disclose fraudulent insider information, which are forbidden by financial regulation in most countries.

The main contribution of this paper is to fill this gap. Using a IBES data set provided by ThomsonReuters, which contains the forecasts issued by 4 648 analysts about the earnings of 241 French firms between 1997 and 2007, we use a gravity model approach. We regress analysts' absolute forecast accuracy on observable determinants surveyed in the theoretical and empirical literature. As the effect of soft information on forecast accuracy is not captured by observable determinants, a measure of soft information should be provided by the pair-specific effect of the regression. We finally use this measure to examine whether the need to access to soft information contributes to optimistic or pessimistic forecasts by financial analysts. Using a gravity model approach to investigate financial analysts' forecasts appears particularly innovative since this approach is usually used in international economics in order to analyse relationships between two trading partners.

The paper is organized as follows. Section 2 sets up the theoretical and empirical background of our research as well as our testable assumption. Section 3 presents our empirical investigation while our results are reported in Section 4. Section 5 considers some robustness checks. Section 6 concludes.

2 Literature and testable assumption

In this section, we review the literature dedicated to forecast accuracy and its determinants. We first focus on observable determinants of forecast accuracy (firms', analysts' and pairs' caractheristics). Then we present the role of soft information and we state our testable assumption.

2.1 Firms' characteristics as observable determinants of forecast accuracy

Concerning firms' characteristics, the main idea developed in the literature is that analysts are concerned about keeping the contact with corporate top executives in order to maintain their source of information about firms. Pratt (1993), Gibson (1995), Womack (1996) and Boni & Womack (2002) provide several examples of this phenomenon. They report situations in which analysts lost their access to a firm's manager consequently to an unfavourable recommendation or earnings forecast about the firm. The need to take care of their relationships with firms' top executives leads analysts to issue optimistic forecasts, i.e. forecasts higher than earnings realisations (Easterwood & Nutt, 1999, Lim, 2001). Hence, as underlined by Lim (2001), forecasts about a firm should be all the more optimistic as information availability and earnings predictability are low.

These theoretical arguments are examined by empirical studies. On a data set of 605 firms from the Value Line Survey between 1989 and 1993, Das, Levine & Sivaramakrishnan (1998) document that the forecast error (the difference between forecasted and realized earnings) increases with firms' profit volatility. This result is confirmed by Lim (2001) on a set of 89 959 forecasts provided by I/B/E/S (*Institutional Brokers Estimate System*) for the period 1984-1996. It is also corroborated by Jackson (2005) who makes use of a data set of 23 brokers on the Australian security market over the period 1992-2002.

Another result of Lim (2001)'s study is that the optimism bias is higher for firms with negative past earnings surprise and poor past stock returns. This supports the view that firms having bad past performance should be more reluctant to release public information.

Morever, as public information availability is stronger for large firms and for those followed by a large number of analysts, optimism is shown to decrease with firm size (Das & Levine, 1998, Lim, 2001, Jackson, 2005) as well as with analyst coverage (Lim, 2001).

Considering interactions between forecasts' accuracy determinants, Das et al. (1998) establish that analyst coverage mitigates the increasing effect of earnings volatility on forecast optimism.

Another driving factor of analysts' forecast behaviour is past optimistic consensus about a firm, which should deter analysts from being right against the crowd and thus increase their forecast error. This is confirmed empirically by Lim (2001).

Finally, Das et al. (1998) find that the need to preserve their relationship with managers entices analysts to issue particularly optimist forecasts about firms who received unfavourable rating from the famous American financial publication Value Line. The same result is reached by Francis & Philbrick (1993) on a set of Value Line recommendations for 1987 (313 firms), 1988 (310 firms) and 1989 (295 firms).

2.2 Analysts' characteristics as observable determinants of forecast accuracy

The literature also investigates how analysts' characteristics affects earnings forecasts. While Clement (1999) focuses on forecast accuracy (defined as the *absolute value* of the difference between forecasted and realized earnings), Lim (2001) concentrates on optimism. For Clement (1999) accuracy increases with the ability of analysts, the resources they devote to analysing firms as well as with the incentive to preserve their reputation towards their employer. For Lim (2001), analysts who have valuable information about firms are less concerned by taking care of their relationships with top executives. This explains why higher analysts' ability and resources should reduce optimism.

Once again, econometric studies provide some support to these theoretical assertions. Using a set of 198 639 I/B/E/S forecasts over the period 1983-1994, Clement (1999) finds that, on the whole sample period, general experience, measured by the number of years for which an analyst supplied at least one forecast, favours accuracy. When general experience increases, two effects are at play. First, the skill of the analyst rises, due to a learning-by-doing process. Second, the analyst is identified as very capable since low-skilled analysts do not last in the profession. General experience does not only affect accuracy but also the optimistic bias. Lim (2001) provides evidence that high-experienced analysts, who are less concerned by preserving their relationship with firms' managers, are generally less optimistic than low-experienced ones.

Accuracy also rises with the complexity of analyst's portfolio (Clement, 1999). When an analyst follows a large number of firms, he devotes less resources to each one. When he follows a large number of industrial sectors, he benefits less from sector specialisation.

Another worthwhile characteristic is the size of the broker employing the analyst. It affects forecast accuracy as well as the optimism bias. As large broker can devote more resources to analysing firms, analysts issue more accurate forecasts (Clement, 1999). For the same reason, maintaining a relationship with firms' top executives becomes less crucial so that forecasts are less optimistic (Lim, 2001).

Finally, using a dataset from I/B/E/S containing 1 293 487 forecasts over the period 1990-2004, Krishnan, Lim & Zhou (2006) explore the role played by past accuracy. They find that low past accuracy is associated with high current absolute value of the difference between forecasted and actual earnings. This suggests that being accurate allows analysts to mitigate the harmful effect of past forecast errors on their reputation and career.

2.3 Pairs' characteristics as observable determinants of forecast accuracy

For the same reasons as above (learning-by-doing and analysts' survival effects), specific experience, measured by the number of years for which an analyst supplied at least one forecast on a given firm, should rise forecast accuracy. Clement (1999)'s result is in accordance with this assumption. It reveals a negative relationship between specific experience and the absolute value of the forecast error. On a set of 182 188 I/B/E/S forecasts between 1988 and 2000, Krishnan et al. (2006) confirm that analysts' forecast error increases with the number of day between the forecast and the end of the fiscal year and as well as with past forecast error.

2.4 The role of soft information

Besides all the observable variables described above, the access to soft information appears as an important, unobservable, determinant of forecast errors by financial analysts (Libby, Hunton, Tan and Seybert, 2008, Pratt, 1993, Gibson, 1995, Womack, 1996, and Boni & Womack, 2002).

In order to inform their customers (portfolio and fund managers) about the projects or the strategy of a particular firm, analysts organize phone calls, conference calls and one-to-one meetings with corporate managers (Fogarty & Rogers, 2005, Breton & Taffler, 2001). This access to qualitative soft information is possible only if the analyst has a good relationship with the firm management. That is why analysts are prone to adopt a forecast behavior that satisfies firms' managers. If the manager pays particular attention to 'beating the forecasts', the analyst is driven to manage the forecast downward in order to provide a positive earning surprise to financial markets. This implies a negative forecast error. Conversely, if the management is mainly interested in a short-term increase in the firm's market value, the analyst is enticed to issue optimistic forecasts, giving birth to a positive forecast error. However, it is unlikely that all analysts react the same way. Some analysts may be less interested than others in maintaining a friendly relationship with a given firm manager because of weak pressure from their customers (Yu, 2008) or because of any observable reason that makes them less sensitive to external or internal pressure. Finally, this suggests that the impact of soft information on forecast errors must be captured by an analyst-firm pair effect. This leads us to state the following testable assumption:

H1: High (negative or positive) forecast errors are associated with strong firm-analyst pair effect.

The following section presents the econometric methodology that will allow us tu check this testable assumption.

3 Empirical investigation

Turning to our empirical investigation, we present successively our data and our econometric model.

3.1 Data

We use data provided by ThomsonReuters. It includes I/B/E/S earnings forecasts and additional data from Worldscope. Our sample concerns 241 French firms from the largest Paris' stock index *SBF 250*, diversified according to firms' size and sector. We study 1-year ahead EPS forecasts by 4 648 analysts from 1997 to 2007 on a monthly basis. This raw database consists of 265 238 firm-analyst-time observations. Several database cleanings were needed. First, once issued, a forecast is frequently repeated several months in the database. We kept track of the number of monthly occurrences of each forecast by storing it in a variable called $DURATION$. Then, for each forecast, we dropped repeated occurrences of the same forecast, to avoid artificially counting it several times. Second, the date of realized EPS, i.e. the fiscal year's ending day, was carefully checked. Whereas some firms do not close their fiscal year by the 31st of december, the database systematically reports the realized EPS each month from january to december. Thus, some EPS artificially appear in january in the database although they are issued in march, for example. When a difference was detected, forecast errors were computed using fiscal years and not calendar years. Third, we dropped aberrant observations (for example when there exist several different forecasts from the same analyst, on the same day, about the same firm, etc). As the reported forecasts are supposed to be one-year ahead earnings forecasts, we created a variable denoted $TERM$, measuring the number of days between the earnings announcement date and the forecast release date. We then dropped forecasts with a negative 'horizon' value, or with a 'horizon' value exceeding 365 days (366 for bissextile years). Finally we obtain 102 876 firm-analyst-time of forecast' observations.

3.2 Econometric model

3.2.1 General methodology

The goal of our econometric study is to assess the importance of soft information for analysts' forecasts. By definition, the effect of soft information on forecast accuracy is not captured by observable determinants. Hence, when regressing the analysts' forecast accuracy on variables described in Sections 2.1 to 2.3, the impact of soft information, highlighted in Section 2.4, should be captured by the disturbance term of the regression.

The main contribution of our empirical approach is to resort to a gravity model approach, usually used in international economics in order to analyse trade relationships between two countries. Soft information is produced by *a given analyst* about *a given firm*. Hence the idea of our investigation is to measure a so-called pair-specific effect (or, more precisely, a firm-analyst-specific effect) by using the methodology of gravity models. Having extracted a measure of soft information, we study the relationship between this measure and analysts' forecast in order to check whether soft information contributes to biased (pessimistic or optimistic) EPS forecasts. This requires two steps, which are presented in the two following subsections respectively.

3.2.2 First step: estimation using a gravity model approach

The first step of our investigation consists in estimating the following model:

$$AFE_{i,j,t} = \alpha + \beta X_{i,t} + \gamma Y_{j,t} + \delta Z_{i,j,t} + \lambda_i D_i + \mu_j D_j + \eta_{i,j} + \upsilon_t + \varepsilon_{i,j,t} \tag{1}$$

The dependent variable, denoted $AFE_{i,j,t}$, is the abolute forecast error on the firm i's EPS, forecasted by analyst j at date t.

Our empirical model contains three sets of explanatory variables. $X_{i,t}$ denote firms' characteristics which are invariant across analysts in t. Symmetrically, $Y_{j,t}$ denote analysts' characteristics which are invariant across firms in t. Finally, $Z_{i,j,t}$ contains a set of variables which are specific both to firms i and analysts j in t. D's are dummy variables indicating a specific analyst j or a firm i.

As in all gravity models[3], the disturbance effect is decomposed into three effects: the firm-specific effect λ_i, the analyst-specific effect μ_j, and the pair-specific effect $\eta_{i,j}$. Finally, υ_t denotes the time-specific effect. We assume that these effects are fixed (non stochastic). However, it is well known that the use of a within approach (OLS on demeaned variables) in presence of invariant and/or rarely changing variables may lead to inefficiency and wrong inference. In model (1), this issue could arise not only in the firm's dimension i, but also in the analyst's dimension j. But many of our explanatory variables $Y_{j,t}$, $X_{i,t}$ and/or $Z_{i,j,t}$ show very small variations in one of these dimensions. That is why, we propose here to use the Fixed Effect Vector Decomposition (FEVD) methodology proposed by Plümper and Troeger (2007). It consists in a three-stage procedure (quite similar to that proposed by Mundlack, 1978, for the random effects model): the first stage of the estimator runs a fixed-effects model to obtain the unit effects, the second stage breaks down the unit effects into a part explained by the time-invariant and/or rarely changing variables and an error term, and the third stage reestimates the first stage by pooled OLS including the time-invariant variables plus the error term of stage 2, which then accounts for the unexplained part of the unit effects.

When estimating the model (1), we consider several specifications, inspired by the empirical literature presented in Section 2. The three first specifications only consider firms' characteristics, denoted $X_{i,t}$. In variant [1], we refer to the estimate by Jackson (2005), which introduces three

[3]Our problem differs from standard gravity models in the sense that firms and analysts are not symmetric as two countries are. Hence there are no dyadic variables in our database. But this does not hamper the results, because we focus on the decomposition of the disturbance term into three. Moreover, the idea of a 'distance' between an analyst and a firm is what we want to capture, although the distance in question is rather social than geographical.

determinants of the analysts' forecast error. The absolute forecast error is expected to increase with earinings precitability ($EPSPREV_{i,t}$), decrease with firm size ($SIZE_{i,t}$), decrease with analysts coverage ($COVER_{i,t}$).

Specification [2] relates to the study by Das et al. (1998), which considers the three variables mentioned above, as well as two interactive terms. $EPSPREV_{i,t}.SIZE_{i,t}$ accounts for interactions between EPS predictability and firm size. As firm size should mitigate impact of EPS variability, its coefficient should be negative. $EPSPREV_{i,t}.COVER_{i,t}$ stands for interactions between EPS predictability and coverage. For the same reason as above, its coefficient should be negative.

In variant [3], we include all the variables contained in Specification (2) and we add an additional variable, considered by Lim (2001): past optimistic consensus about a firm ($PASTMDFE_{i,t}$) entice analysts to follow the crowd by being also optimistic -Absolute forecast error should increase.

In the two following specifications, explanatory variables are analysts' characteristics, $Y_{i,t}$, as well as determinants which are specific both to firms i and analysts j in t, $Z_{i,j,t}$.

Specification [4] refers to the study by Clement (1999), which mainly considers five explanatory variables. Absolute forecast errors are expected to: decrease with ($GENEXP_{j,t}$), the general experience of the analyst j; increase with the number of firms ($NBFIRM_{j,t}$) and sectors ($NBSEC_{j,t}$)the analyst follows, decrease with brokerage house size ($BROKERSIZE_{j,t}$), and decrease with $SPECEXP_{i,j,t}$, the specific experience of the analyst i about the firm j.

In specification [5], we follow Krishnan et al. (2006), adding four additional variables to those included in variant [4]. Absolute forecast errors should: be greater if the forecast is far from the end of the fiscal year ($TERM_{i,j,t}$), decrease if the analyst frequently revises his forecasts ($FREQ_{i,j,t}$), and hence decrease with the forecast' lifetime in the database ($DURATION_{i,j,t}$) and increase with past errors ($PASTAFE_{i,j,t}$).

Finally, we also estimate our own encompassing model (specification [6]), in which we include all explanatory variables mentioned above.

Tables 1, in Appendix, reports the list of regression variables mentioned above while Tables 2 provides some summary statistics for each of them.

3.2.3 Second step: analysing the pair-specific effect

In a second step, we focus on the pair-specific effect of our estimate in order to produce a measure of soft information, i.e. information obtained in the framework of a privileged relationship between the firm and the analyst and that is not captured in the observable determinants defined in the previous subsection.

Our methodology decomposes the fixed effect of the panel regression into three: the firm effect (λ_i), the analyst effect (μ_j), and the firm-analyst' pair effect ($\eta_{i,j}$). Each of them accounts for the unobservable determinants influencing the absolute forecast error.

Having estimated our model, for each of our six specifications, we are able to compute $C_{i,j}$, the 'contribution of the pair-specific effect', defined as follows:

$$C_{i,j} = \frac{\mid \eta_{i,j} + \epsilon \mid}{\mid \lambda_i \mid + \mid \mu_j \mid + \mid \eta_{i,j} + \epsilon \mid}$$

, where ϵ is the unconditionnal constant of the regression. The pair-specific effect, computed by the STATA procedure of Plümper & Troeger (2007), is centered on a mean calculated over all observations. Adding the unconditionnal constant ϵ in both the denominator and the numerator of $C_{i,j}$ allows to correct for this bias and to build a consistent indicator.

$C_{i,j}$, defined as the pair-specific effect in absolute value relatively to the absolutes values of firm-specific and pair-specific effects, measures the importance of the pair-specific effect as a determinant of $AFE_{i,j,t}$, relative to the firm-specific and to the analyst-specific effects. For a given forecast, the higher the contribution of the pair-specific effect, the stronger soft information and the closer the relationship between a particular firm and a particular analyst.

Based on results obtained from each of our six specifications, we order observations by twentiles of the (relative) median forecast error $MedFE_{i,j}$, defined as the median, over the whole data set period, of the difference between the EPS forecasts and the EPS realization of each firm i by each analyst j for a forecast issued in t. Then we compute the mean value of $C_{i,j}$ for each twentiles (from the 5% most pessimistic to the 5% most optimistic). We plot the median forecast error twentiles against the pair-effect contribution. Such a graph should allow us to check our testable assumption H1:

- If our graph exhibits a flat horizontal line, it means that there is no link between our pair-effect contribution $C_{i,j}$ and the dependent variable. Hence the forecast error can be positive, zero or negative, whatever the quality of the relationship between analysts and firms. If this is true, H1 is not validated.

- If we observe that $C_{i,j}$ is greater for the most negative (positive) forecast error twentiles, this means that the need to maintain the access to soft information provide analysts an incentive to issue pessimistic (optimistic) forecasts. Conversely, we should observe that $C_{i,j}$ is weaker for central twentiles (i.e. forecast error around zero). If this is true, H1 is validated.

4 Results

We now present the results of panel regressions as well as graphs representing the relationship between the pair-effect and the forecast error.

4.1 Results of panel regressions with vector decomposition

The results of panel regressions with vector decomposition are reported in Table 3, in Appendix.

To start with, we focus on variables $X_{i,t}$, which capture firm-specific characteristics. We observe that in all specifications in which it is included ([1], [2], [3] and [6]), $EPSPREV$ has a positive impact on the dependent variable. In line with the results obtained by Jackson

(2005), Das et al. (1998), Lim (2001), this suggests that the harder to predict the firm's EPS, the larger the analyst' forecast error. Concerning the coefficient for $SIZE$, specifications [2] and [3] indicate that it is significant with the expected sign (i.e. negative). The larger the firm, the easier its EPS to forecast. This result is consistent with Lim (2001), Jackson (2005) and Das et al. (1998). Reflecting the diversity of findings obtained in the literature, results concerning $COVER$ are less robust. In line with Lim (2001), its coefficient has the expected (i.e. negative) sign in specification [3] but it has a positive sign in specification [1]. Moreover, following Jackson (2005) and Das et al. (1998), it is insignificant in specification [2]. Turning to the interactions terms $EPSPREV.SIZE$ and $EPSPREV.COVER$, one observes that their respective coefficient varies across specifications [2] and [3]. In the former, $EPSPREV.COVER$ has a positive impact on FE but, in line with the theory and with the empirical result of Das et al. (1998), in the latter, its coefficient is negative. Moreover, while its coefficient never appears significant in Das et al. (1998), $EPSPREV.COVER$ has a negative impact on AFE in variant [2] but a positive one in variant [3]. Finally, following Lim (2001), the coefficient for $PASTMDFE_{i,t}$ has the expected positive sign (variant [3]). The higher past optimistic consensus, the larger the analysts' forecast error.

We now turn to variables $Y_{i,t}$, which represent analysts' characteristics. As expected, variants [4], [5] and [6] in Table 3 indicate that general experience ($GENEXP$) contributes to reduce the forecast error in all specifications in which it is included. In line with Lim (2001) and Clement (1999), the longer the general experience of the analyst, the lower its forecast error. In specifications [4] and [6], $NBFIRM$ has a significant and positive sign. This supports the theory as well as Clement (1999)'s finding: when an analysts follows a large number of firms, he dedicates less resource to each of them such that its forecast error is higher. Table 3 also reveals that in specifications [4] and [6], the number of sectors followed by the analyst has no impact on the dependent variable since the coefficient for $NBSECT$ turns out to be insignificant in all specifications in which is introduced. This is consistent with many of annual regression results obtained by Clement (1999). However, its coefficient is significative and negative in specification [5]. Interestingly, as expected, the coefficient for $BROKER$ has a significant and negative sign in specifications [4], [5] and [6]. This result, which is in line with Lim (2001) and Clement (1999), suggests that being employed in a large broker allows to dedicate more resource to prediction and more accurate forecasts.

Finally, we comment on our results concerning variables $Z_{i,j,t}$, which are specific both to firms i and analysts j. As expected, the specific experience of analysts ($SPECEXP$) has a significant and positive impact on the dependent variable (columns [4]). This confirms Clement (1999)'s findings. However, its coefficient is insignificant in column [4] and exhibits a negative sign in column [6]. Variants [5] and [6] in Table 3 also indicate that in line with the theory, the coefficient for $TERM$ is significant and positive: the farther the end of fiscal year, the less accurate the analyst's forecast. $PASTAFE$ also has the expected positive sign, attesting for inertia in forecast dynamics. Finally, we find that $DURATION$ and $FREQ$ do not have the expected impact on the dependent variable.

4.2 Graphs of pair-effect contribution $C_{i,j}$ by twentiles of median forecast error $MedFE_{i,j}$

Graphs 1 to 6 provide interesting representations of $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specifications 1 to 6 respectively. The first twentile represents the most pessimistic forecasts (FE around -2), the 10th twentile represents the most accurate forecasts (about 0) and the 20th twentile represents the most optimistic forecast (FE around +9). In all graphs, we observe that there exists a non-linear relationship between $C_{i,j}$ and $MedFE_{i,j}$. In all specifications except number [4], it is represented by a convex curve. See for example Graph 6, which corresponds to the encompassing model, in which all explanatory variables are included. $C_{i,j}$ is at its lowest level when $MedFE_{i,j}$ lies in the tenfth twentile. When forecast error is weak, the pair effect only accounts for about 30% of the total fixed effect. When the forecast error is strong, the pair effect account for more than 40% of the fixed effect. This is interesting for at least two reasons. First, it suggests that whatever the number of characteristics included in the estimation to explain analysts' forecast error, the relationship between $C_{i,j}$ and $MedFE_{i,j}$ can be represented by a non-linear curve. Second, this indicates that the contribution of the pair-specific effect is at its highest level both for the most optimistic and for the most pessimistic forecast twentile (except for specification [4]). It reaches it lowest level for intermediate values of $MedFE_{i,j}$[4].

Finally, our results mainly validate our testable assumption H1 and provide evidence that some analysts try to maintain friendly relationships with some firms' managers by intentionally biasing their EPS forecasts.

5 Robustness checks

In this section, we propose two robustness checks of our findings. We first discuss the FEVD estimator. We then test for the relationship between pair-effects and forecast errors.

5.1 Discussion on the FEVD estimator

Firstly, it is worth noting that the FEVD estimator is equivalent to a standard instrumental variables approach, for a specific set of instruments as recently shown by Breusch & al. (2010). Greene (2010) argues that the FEVD approach does not provide an estimator for the coefficients on time invariant variables in a fixed effects model: that part of the parameter vector remains unidentified. However in presence of slowly-changing variables as in our context, this estimator

[4]Concerning specification [4], let us remind that it represents the regression with only analyst-specific variables and only one pair-specific observable determinant. A concave function may then illustrate the fact that, because firm-specific and other pair-specific determinants are excluded, firm- and pair-effects capture all of the missing variables effect. That is why $C_{i,j}$ represents almost 50% of the total fixed effect for median values of $MedFE_{i,j}$. Besides, it is not the case for specification [1] when the regression only contains firm-specific variables, because the firm-effect seems to bear more importance than the analyst effect.

remains consistent, even if the efficiency gains (compared to Hausman & Taylor's, 1981, approach) are controversy. One advantage of the FEVD is that, as the Fixed Effects (Within) estimator, and unlike the Hausman and Taylor one, it does not require specifying the exogeneity status of the explanatory variables. For Greene (2010), the FEVD estimator simply reproduces (identically) the linear fixed effects (dummy variable) estimator, then substituting an inappropriate covariance matrix for the correct one. That is why we propose here an additional estimate of the covariance matrix.

The covariance matrix discussed in Subsection 4.1 is defined in the context of the three stages estimation procedure by Plümper & Troeger (2007). We compute the value of the standard error obtained using the covariance matrix proposed by Greene (2010), that corresponds to the matrix estimated in the first stage of the Plümper & Troeger (2007) procedure. Our goal is to compare both standard errors in order to show that Greene (2010)'s argument is weakly relevant as regards our purpose, which is to investigate the relationship between pair effects and forecast errors. Our results, available upon request to the authors, indicate that in most cases, standard errors using the covariance matrix from the first stage of the estimation procedure of Plümper & Troeger (2007), as it is proposed by Greene (2010), are higher than those obtained through the third stage. This suggests that the estimators showed in Section 4 are consistent, and that our use of the decomposition of the fixed effect is relevant as regards our data and purpose.

5.2 Testing the relationship between pair-effect and forecast error

In this subsection, we test for the relationship between $C_{i,j}$ and $MedFE_{i,j}$ in order to check whether the contribution of the pair effect differs according to the size of the forecast error. We first conduct a median-test that allows us to investigate whether the median of $C_{i,j}$ in a given quantile (here, twentile) equals the median of the whole sample. We then conduct the Bartlett-test which tests for the equality of variance of $C_{i,j}$ across quantiles. We finally use the Krusal-Wallis equality-of-population rank test, which consists to test whether the rank sum of each observation ranked by $C_{i,j}$ differs across quantiles. Results of these three robustness tests are reported in Table 4, in Appendix. They indicate that for each test, the testable assumption H1 is still validated: the contribution of the pair-effect differs according to the size of the forecast error. This result seems particularly robust since its holds for each of our six specifications.

6 Conclusion

The goal of this paper was to check whether the analysts' concern to to preserve their access to soft information provides them an incentive to issue optimistic or pessimistic forecasts. We used a Thomson Reuters data set that contains the forecasts issued by 4 648 analysts about the earnings of 243 French firms over the period 1997-2007.

One important innovation of our approach is to resort to the methodology of gravity models to examine relationships between firms and analysts. The second interest of our paper is to propose a measure of soft information. Having regressed analysts forecast error on observable firm-specific, analyst-specific and pair-specific characteristics, we decompose the disturbance effect in order to extract a pair-specific effect. This provides a measure of soft information, allowing us to check whether soft information contributes to analysts' pessimism or optimism. Finally, we provide interesting evidence that the need to preserve their relationship with firms' manager and their access to soft information entice some analysts to issue pessimistic forecasts while enticing some others to issue optimistic forecasts about firms' EPS.

Our results undoubtedly call for further research. Of course, our work could be extended to the case of other countries, in order to check whether the effect of soft information has a national dimension. More ambitiously, it would be interesting to examine the consequences of analysts' forecast for portfolio investment strategies. For example, this could be done by studying what kind of behaviour (accurate, pessimistic or optimistic) leads to the most profitable investment recommendations for asset managers.

References

Barber B., Lehavy R. & Trueman B. (2007), 'Comparing the stock recommendation performance of investment banks and independent research firms,"', *Journal of Financial Economics* **85**, 490-517.

Boni L. & Womack K. (2002), 'Wall Street credibility problem: misaligned incentives and dubious fixes', *Brookings Wharton Papers on Financial Services*, 93-130.

Breton G. & Taffler R. (2001), 'Accounting information and analyst stock recommendation decisions: a content analysis approach', *Accounting and Business Research* **31)**, 91-101.

Breusch T., Ward M., Nguyen H. & Kompas T. (2010), 'On the fixed-effect vector decomposition', *mimeo http://mpra.ub.uni-muenchen.de/26767/*

Brown L., (1997), 'Analyst forecasting error: Additional evidence', *Financial Analysts Journal* **53**, 81-88.

Burgstahler D. & Eames M. (2006), 'Management of earnings and Analysts' forecasts to achieve zero and small positive earnings surprises', *Journal of Business Finance & Accounting* **33**, 633-652.

Clement M. (1999), 'Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?', *Journal of Accounting and Economics* **27**, 285-303.

Das S., Levine C. & Sivaramakrishnan K. (1998), 'Earnings predictibility and bias in analysts' earnings forecasts', *Accounting Review* **73**, 277-294.

Dugar A. & Nathan S. (1995), 'The effect of investment banking relationships on financial analyst's earnings forecasts and investment recommendations', *Contemporary Accounting Research* **12**, 131-160.

Easterwood J. & Nutt S. (1999), 'Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism', *Journal of Finance* **54**, 1777-1797.

Fogarty T., & Rogers K. (2005), 'Financial analysts' reports: an extended institutional theory evaluation', *Accounting, Organization and Society* **30**, 331-356.

Francis J. & Philbrick D. (1993), 'Analysts' decisions as product of multi-task environment', *Journal of Accounting Research* **31**, 216-230.

Gibson R. (1995), 'Analysts take heat for 'sell' rating on Archer-Daniels', *Wall Street Journal* august 31.

Greene W. (2010), 'Fixed effects vector decomposition: A magical solution to the problem of time invariant variables in fixed effects models?', *mimeo.*

Hayward M. & Boecker W.,'Power and conflicts of interest in professional firms: Evidence from investment banking', *Administrative Science Quarterly* **43**, 1-22.

Jackson A. (2005), 'Trade generation, reputation and sell-side analysis', *Journal of Finance* **60**, 673-717.

Hausman J. & Taylor W. (1981), 'Panal data and unobservable individual effects', *Econometrica* **49**, 1377-1398.

Lin H., McNichols M. & O'Brien P. (2005), 'Analyst impartiality and investment banking relationships, *Journal of Accounting Research* **43**, 623-650.

Krishnan M., Lim S. & Zhou P. (2006), 'Analysts' herding propensity: Theory and evidence from earnings forecasts', *mimeo http://ssrn.com/abstract=929467.*

Lim T. (2001), 'Rationality and analysts' forecast bias', *Journal of Finance* **56**, 369-385.

Matsumoto D. (2002), ‘Management’s incentives to avoid negative earnings surprises’, *The Accounting Review* **77**, 483-514.

Mundlak Y. (1978), ’On the pooling of times series and cross section data’, *Econometrica* **46**, 69-85.

Payne J. & Robb S. (2007), ‘Earnings management: The effect of ex-ante earnings expectations’, *Journal of Accounting, Auditing and Finance* **15**, 371-92.

Plümper T. & Troeger V. (2007), ‘Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects’, *Political analysis* **15**, 124-139.

Pratt, T. (1993), ‘Wall Street four letter word’, *Investment Dealers’ Digest* march 29.

Womack K. (1996), ‘Do brokerage analysts recommendations have investment value?’, *Journal of Finance* **51**, 137-167.

Yu P. (2008), ‘Analyst coverage and earnings management’, *Journal of Financial Economics* **88**, 245-271.

Appendix

Table 1: List of regression variables

	DEPENDENT VARIABLE
AFE	Absolute forecast error (absolute difference between the EPS forecast and the EPS realization of each firm i by each analyst j at date t of forecast issue)
	INDEPENDENT VARIABLES AND EXPECTED SIGNS
	Firm characteristics $\mathbf{X}_{i,t}$
$EPSPREV$ (+)	Predictibility of EPS (volatility of the firm i's EPS over the last 3 years)
$SIZE$ (-)	Size of the firm (log of the market capitalization of the firm i in t)
$COVER$ (-)	Coverage of the firm in t (number of analysts who follow the firm i in t)
$PASTMDFE$ (+)	Consensus surprise (for each firm i, the median of the difference between the consensus and the realized EPS in year $t-1$)
	Analyst characteristics $\mathbf{Y}_{i,t}$
$GENEXP$ (-)	General experience of the analyst j (in t, number of days since the analyst's first forecast)
$NBFIRM$ (+)	Number of firms followed by the analyst j in t
$NBSECT$ (+)	Number of sectors followed by the analyst j in t
$BROKER$ (-)	Size of the broker (number of analysts working for the analyst j's broker in t)
	Firm-analyst characteristics $\mathbf{Z}_{i,t}$
$SPECEXP$ (-)	Specific experience of the analyst (in t, number of days since the first forecast by analyst j about the firm i)
$TERM$ (+)	Number of days from t to fiscal year end for a forecast issued by the analyst j on the firm i
$FREQ$ (-)	Frequency of forecasts (number of forecasts per year by the analyst j on a firm j in t)
$PASTAFE$ (+)	forecast error of the analyst j on the firm i in $t-1$
$DURATION$ (+)	forecast lifetime in the database in months (by the analyst j on a firm j in t)

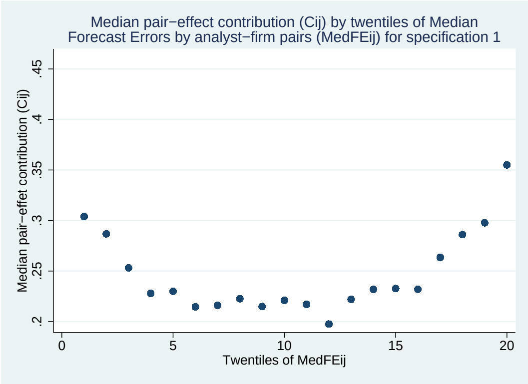
Table 2: Statistical summary for regression variables (1997-2007)

Variables	Mean	Standard deviation	Max	Min	Nonmissing observations
AFE	2.46	6.12	162.57	0	102 876
$EPSPREV$	1.79	4.72	85.17	1.15	94 231
$SIZE$	21.93	1.77	25.95	15.30	102 627
$COVER$	19.49	9.60	47	1	102 876
$PASTMDFE$	2	6.54	75.97	-39.90	94 595
$GENEXP$	1 344.29	1 141.75	6 434	0	102 875
$NBFIRM$	4.61	3.49	24	1	102 876
$NBSECT$	2.22	1.39	9	1	102 876
$BROKER$	17.93	10.15	62	1.	102 876
$SPECEXP$	849.65	912.73	6 253	0	102 875
$TERM$	187.88	102.26	365	0	102 875
$FREQ$	4.01	1.98	13	1	102 876
$PASTAFE$	2.48	6.16	127.27	0	69336
$DURATION$	2.82	2.14	18	1	102 876

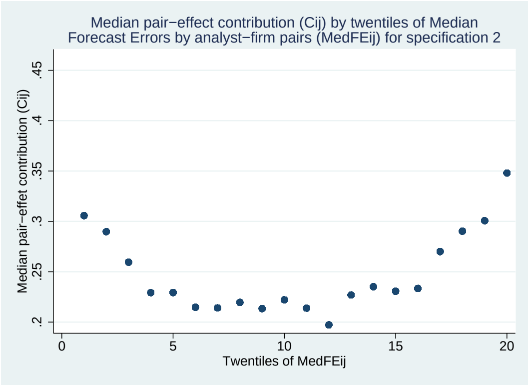
Table 3: Results of panel regression with vector decomposition

Variables (expected sign)	Specifications					
	(1)	(2)	(3)	(4)	(5)	(6)
Firms' characteristics $\mathbf{X}_{i,t}$						
$EPSPREV$ (+)	0.208***	1.677***	3.521***			2.359***
	(0.003)	(0.070)	(0.088)			(0.081)
$SIZE$ (-)	-1.066	-1.084***	-0.224***			0.016
	(0.033)	(0.032)	(0.037)			(0.041)
$COVER$ (-)	0.023***	-0.002	-0.016***			-0.021***
	(0.002)	(0.003)	(0.003)			(0.004)
$EPSPREV.COVER$ (-)		0.012***	-0.163***			-0.112***
		(0.006)	(0.004)			(0.004)
$EPSPREV.SIZE$ (-)		-0.081***	0.008***			0.012***
		(0.003)	(0.000)			(0.000)
$PASTMDFE$ (+)			0.130***			0.088***
			(0.003)			(0.004)
Analysts' characteristics $\mathbf{Y}_{j,t}$						
$GENEXP$ (-)				-0.003***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
$NBFIRM$ (+)				0.070***	0.053***	0.077***
				(0.008)	(0.010)	(0.009)
$NBSECT$ (+)				0.162	-0.010***	0.000
				(0.002)	(0.002)	(0.336)
$BROKER$ (-)				-0.016***	-0.010***	-0.009***
				(0.002)	(0.002)	(0.002)
Firm-analysts' characteristics $\mathbf{Z}_{i,j,t}$						
$SPECEXP$ (-)				0.002***	0.000	-0.000***
				(0.000)	(0.000)	(0.000)
$TERM$ (+)					0.001***	0.001***
					(0.000)	(0.000)
$DURATION$ (+)					-0.005	-0.003
					(0.008)	(0.008)
$FREQ$ (-)					0.165***	0.145***
					(0.003)	(0.011)
$PASTAFE$ (+)					0.134***	0.091***
					(0.003)	(0.004)
Sector dummies	yes	yes	yes	yes	yes	yes
Analyst dummies	yes	yes	yes	yes	yes	yes
Firm dummies	yes	yes	yes	yes	yes	yes
Nb. obs.	85 398	94 026	79 778	102 875	69 336	65 586

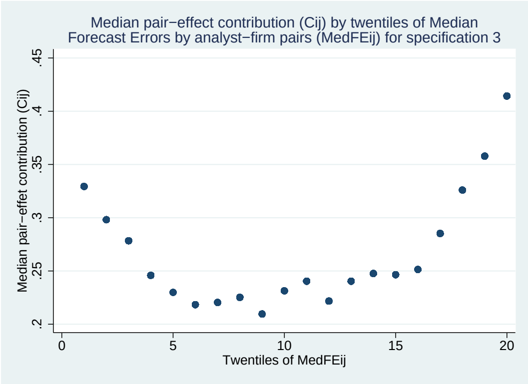
*, ** and *** denote significance respectively at the 10%, 5% and 1% level.



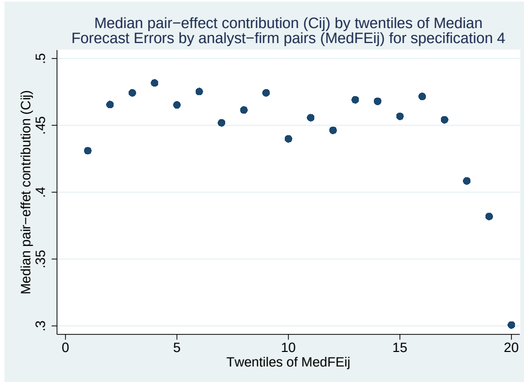
Graph 1: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [1]



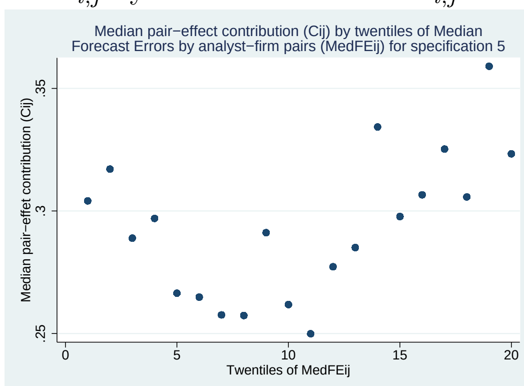
Graph 2: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [2]



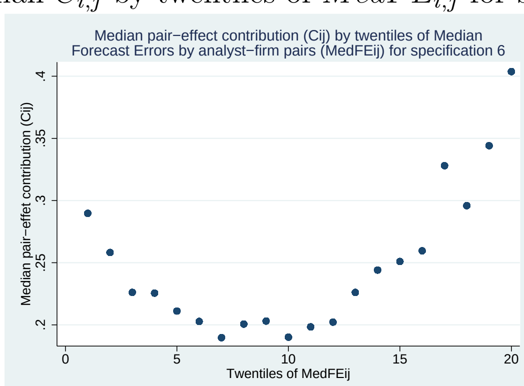
Graph 3: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [3]



Graph 4: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [4]



Graph 5: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [5]



Graph 6: median $C_{i,j}$ by twentiles of $MedFE_{i,j}$ for specification [6]

Table 4: Robustness tests for specifications 1 to 6

Median test (Chi^2 stat)					
(1)	(2)	(3)	(4)	(5)	(6)
1 100***	244***	369***	495***	81***	233***

The null is the equality of the median of $C_{i,j}$ in each twentile to the median of the whole sample.

Bartlett test (Chi^2 stat)					
(1)	(2)	(3)	(4)	(5)	(6)
1 600***	334***	377***	114***	82***	256***

The null is the equality of variances of $C_{i,j}$ across twentiles.

Krusal-Wallis test (Chi^2 stat)					
(1)	(2)	(3)	(4)	(5)	(6)
1693***	314***	524***	524***	99***	341***

The null is the equality of the rank-sum of each observation ranked by $C_{i,j}$ across twentiles.

*, ** and *** denote significance respectively at the 10%, 5% and 1% level.