Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

**Elisa Cavezzali, Jacopo Crepaldi
and Ugo Rigoni**

Proximity to Hubs of Expertise in Financial Analyst Forecast Accuracy

**Working Paper n. 8/2013
August 2013**

ISSN: 2239-2734

Proximity to Hubs of Expertise in Financial Analyst Forecast Accuracy

ELISA CAVEZZALI
elisa.cavezzali@unive.it
Department of Management
Università Ca' Foscari di Venezia

JACOPO CREPALDI
jacopocrepaldi@yahoo.it

UGO RIGONI
rigons@unive.it
Department of Management
Università Ca' Foscari di Venezia

July 2013

Abstract

This paper investigates whether the geographical proximity of financial analysts to hubs of information and expertise can influence their forecasting accuracy. Recent studies show that the financial analyst forecasting process show a systematic difference in earnings forecast accuracy dependent on the geographical distance of analysts from the companies which they follow. The literature argues that local analysts issue more accurate forecasts because they have an informational advantage over analysts who are further away.

Industrial centres can constitute important knowledge spillovers by creating formal and informal networks amongst firms and higher education and research institutions. In such a hub, information can easily flow and propagate. Our hypothesis is that physical proximity to these hubs, and not to the companies they follow, is an advantage for financial analysts, leading to the issue of more accurate forecasts.

Using a sample of 205 observations related to 33 firms, across seven countries and ten sectors, our results are consistent with the hypothesis.

Even though preliminary, and probably in part biased by sample selection issues, overall, the empirical evidence confirms the benefit of being part of a network, formal or informal, in which information, knowledge and expertise sharing can flow easily. We try to give some new evidence on what can cause variations in financial analyst accuracy by exploring these concepts, well known and analysed in other fields, but new in the context of financial analysts

Keywords: forecast accuracy, sell-side analysts, geography, hubs of knowledge

Correspondence to:

Elisa Cavezzali	Dept. of Management, Universita Ca' Foscari Venezia San Giobbe, Cannaregio 873 30121 Venezia, Italy
Phone:	[+39] 041-234-6931
Fax:	[+39] 041-234-8701
E-mail:	elisa.cavezzali@unive.it

1. Introduction

Does geographical proximity enhance financial analysts' accuracy? Results from recent literature about the financial analyst forecasting process show a systematic difference in earnings forecast accuracy dependent on the geographical distance of analysts from the companies which they follow (see, e.g. Malloy (2005) and Bae et al. (2008)).

This paper investigates whether the geographical proximity of financial analysts to hubs of information and expertise can influence their forecasting accuracy.

The literature argues that local analysts issue more accurate forecasts because they have an informational advantage over analysts who are further away.[1]

In general, financial analysts are 'intermediaries' between the managers of the firms which they follow and financial markets. They use a heterogeneous set of information (hard and soft, explicit and tacit) about the company which they follow, the industry and the economic system in order to arrive at earnings forecasts, company value and an investment recommendation. Thus, the evaluation process performed by financial analysts has a sequential structure '*input-processing-output*'.

In this framework, local informational advantage could be related either to different information sets available to local and remote analysts or to the superior skills of local analysts in processing the same information set. Specifically, a different set of information could be derived from an analyst's direct contact with company management and premises or from lower information gathering costs. In an international setting, the superior skills could be related to better knowledge of the local language, culture or customs.

The purpose of this study is related to this latter idea, introducing a new concept of proximity. Drawing on both international- and industrial economics-based research and on network analysis and cluster theories, this work aims to explore the role of proximity of analysts to centres of production of soft and not structured knowledge in order to explain the performance of financial analysts.

Industrial centres can constitute important knowledge spillovers by creating formal and informal networks amongst firms and higher education and research institutions. In such a hub, information

[1] Alternative explanations are related to incentive arguments, including compensation and career incentives, and not to information asymmetries.

can easily flow and propagate. Our hypothesis is that physical proximity to these hubs is an advantage for financial analysts, leading to the issue of more accurate forecasts.

Prior literature provides mixed results with respect to geographical advantage. We hypothesise that unstudied aspects of analysts' locations may add important evidence to prior literature.

We test our hypothesis by the collection of both macroeconomic data - to identify the hubs of expertise - and financial analyst data, specifically earnings forecasts, dates of research reports and details about the financial analysts' location. The final filtered sample of 205 matched-observations related to 33 firms, across seven countries and ten sectors over four years (from 2004 to 2008).

Specifically, we first establish the location of the hubs of expertise of the country and industry of our sample, drawing on concepts from industrial and international economics.

Secondly, we test whether the accuracy of financial analyst forecasts depends on the location of financial analysts in regard to the hubs of expertise identified.

Our results are consistent with the hypothesis. In order to establish the robustness of this approach, we employed different measures of both earnings forecast accuracy and proximity.

Even though preliminary, and probably in part biased by sample selection issues, overall, the empirical evidence confirms the benefit of being part of a network, formal or informal, in which information, knowledge and expertise sharing can flow easily. We try to give some new evidence on what can cause variations in financial analyst accuracy by exploring these concepts, well known and analysed in other fields, but new in the context of financial analysts.

We believe that the identification of the drivers which affect forecast accuracy is important for at least three reasons. First, investors should benefit from being able to identify more accurate forecasts (and forecasters) as a good knowledge of these drivers can help them to spot more reliable information sources. Second, as earnings forecast are part of the input to analysts' valuation methods and their stock recommendations, more accurate forecasters could issue more profitable recommendations. Third, forecast accuracy is also important to brokerage houses and investment banks as it enhances the quality of their output. Trading commissions and portfolio performance, which are strongly based on analyst ability, in fact generate part of their revenues. Thus, forecast accuracy should be important in turn to analysts, who can be rewarded by their brokerage houses according to their accuracy. Finally, the results of this study could also contribute to the organisation of the research operations and financial research departments of investment banks and brokerage houses. The identification of a link between hubs of expertise and financial analyst performance could induce a change in the structure of financial research, from being country- or industry-based to expertise hub-based.

The paper is organised as follows: Section 2 explains the literature on financial analyst accuracy; Section 3 reports the methodology adopted to measure proximity to hubs of expertise; Section 4 describes the data and research design; Section 5 reports the main results; Section 6 illustrates the conclusions of this research, suggesting new patterns of analysis; and the Appendix reports the technical details of the procedure used to identify the hubs of expertise.

2. Literature Review

Many papers have investigated the link between geography and information asymmetries in a number of financial and economic contexts.

Focusing on investors, for example, it is well known that they are biased towards their home country. The literature explaining this home bias is extensive, but still far from having obtained conclusive results. An important stream has underlined the differences in information available to domestic and foreign investors. Early papers focused on this area, for example Gehrig (1993) and Kang and Stulz (1997). A number of papers attempted to identify more directly whether foreign investors have an informational disadvantage. Hau (2001) investigated trading data for professional traders and showed that local investors perform better than foreign traders. Choe et al. (2005) and Dvorak (2005) found that foreigners trade at worse prices in Korea and Indonesia, respectively.

On the institutional investors' side, Grinblatt and Keloharju (2000) and Seasholes (2000) argued that better resources and access to expertise allows foreign institutions to perform better than domestic institutions. Grinblatt and Keloharju (2000) found that in the Finnish market over a two-year period, foreign and domestic financial corporations bought more stocks which performed well over the next 120 trading days than domestic individual investors. Seasholes (2000) found that foreign investors buy (sell) ahead of good (bad) earnings announcements in Taiwan, while domestic investors do the opposite. Froot, O'Connell and Seasholes (2001) and Froot and Ramadorai (2001) used flow data to show that foreign investors have some ability to predict returns. These papers are consistent with the better information and greater sophistication on the part of foreign investors.

However, evidence on the performance of foreign investors is mixed. For instance, Shukla and van Inwegen (1995) showed that UK money managers underperform in comparison with their American counterparts when picking US stocks. Using 18 years of annual data, Kang and Stulz (1997) found no evidence that foreign investors outperformed domestic investors in Japan.

A new strand of literature looked at the impact of distance on portfolio choice within countries. Coval and Moskowitz (1999), using only U.S. stock returns, provided evidence that investor location matters, in that mutual fund managers overweight the stocks of firms located closer to them. In another paper, the same authors (2001) found that mutual fund managers are better at picking stocks of firms which are closer to where they are located than those from a more distant location. Huberman (2001) found local concentration in the ownership of the Baby Bells in the US. Ivković and Weisbrenner (2005) used data from a large discount brokerage house and found the striking result that one out of six US individuals in their sample only invested in companies headquartered within 250 miles of their household. Recent papers show that social networks are also important for stock holdings. Hong, Kubik and Stein (2005) showed that mutual fund managers were more likely to hold a particular stock when other managers in the same city held the same stock.

With specific regard to financial analysts, a number of papers have investigated how the geography of security analysts can affect their forecast performance. Some have analysed whether the country-related features have an impact on financial analyst accuracy. Chang, Khanna and Palepu (2000) and others documented considerable variation across countries in the accuracy of analyst forecasts, depending on specific country characteristics. However, the international evidence seems mixed and inconclusive. For instance, while Chang, Khanna and Palepu (2000) found evidence that a country's legal system helps us to understand the accuracy of analysts, Ang and Ciccone (2001) reached the opposite conclusion. Hope (2003) found that the enforcement of accounting standards and firm-level disclosure were important determinants of forecast accuracy.

Only a handful of papers investigate, directly or indirectly, how analysts' physical distance to evaluated companies affects the accuracy of their forecasts. Malloy (2005), for instance, found evidence that, US analysts located close to the evaluated firm are more accurate than those who are further away. He argued that the ability of local analysts to make house calls rather than conference calls and the opportunity to meet CEOs and survey company operations directly provide them with an opportunity to obtain valuable private information. Following this logic, geographic proximity is a proxy for the quality of analyst information.

In an international setting, analysts cover those countries which are open to foreign investors. Bae et al. (2008) showed that the financial opening of a country[2] is followed by the increasing interest of foreign analysts. Bacmann and Bolliger (2001) directly examined the relative performance of analysts from local and foreign brokerage houses in seven Latin American stock markets. They concluded that foreign analysts outperformed local analysts in their study which focused on seven Latin American countries. When they compared the mean difference in forecast error between local and foreign analysts, it was not significantly positive for all of the countries in their sample, with the exceptions of Mexico and Colombia. In contrast, Orpurt (2004) found evidence of a significant local advantage in a sample of seven European countries. In his study, local analysts were defined as resident in that country. He found that his evidence was driven by Germany. However, while Bolliger (2004) focused on local versus foreign brokerage houses (not analysts) and found an advantage for local brokerage houses in Europe, Orpurt (2004) did not find this type of local advantage. Conroy, Fukuda, and Harris (1997) also found a local brokerage house advantage in Japan. Finally, Chang (2003) compared specifically the stock recommendations of foreign and expatriate analysts covering Taiwanese firms. He found that there was a local advantage, as expatriate analysts outperformed foreign analysts. He also found that expatriate analysts outperformed local analysts working for domestic firms. This result is consistent with the hypothesis that local analysts working for foreign institutions have the advantage of belonging to more sophisticated and resourceful organisations. Bae et al. (2008) showed that there is a significant local advantage for analysts in a sample of 32 countries. This local analyst advantage holds after having controlled for analyst characteristics as well as firm characteristics. However, it varies substantially between countries. The local analyst advantage is stronger in countries where disclosure is weaker, institutional investors are less important and corporate ownership is more concentrated.

These results are very interesting, consistent with and complement another part of the literature, which analyses the information needs of financial analysts. Previts et al. (1994), for instance, performing a content analysis of 479 sell-side analyst reports, showed that analyst information needs to extend beyond that contained in financial reports and include softer, more subjective information. Breton and Tafler (2001) presented a content analysis of 105 analyst reports in order to assess the information used by analysts. Non-financial information seemed to be equally important as financial information. The financial analysts were particularly interested in non-financial information about management and strategy, as well as the trading environment of the firm. According to the distance-related literature, this information is probably easily gathered if the

[2] By financial opening the authors mean the opening of the country to foreign investors.

analysts are closer to the firm being evaluated. This is also supported by Barker's results (1998). Performing a survey, he found that analysts considered personal contact to be more important than earnings announcements and financial statements. Since proximity can help analysts to keep frequent personal contact, according to the local information advantage hypothesis, the accuracy of analysts who are close to the evaluated company should improve. The author provided four possible reasons underlying this evidence. First, personal contact can improve the timeliness of the disclosure of information. Second, analysts can question company managers directly. Third, it helps analysts to have comparative advantage over their peers, and, fourth, they can focus on strategic and forward-looking issues.

In summary, prior research has documented significant variation in the quality of analysts' forecasts, with some being more accurate others. According to previous results, Brown et al. (1985) and Brown (1993), it is possible to conclude that the accuracy of earnings forecasts depends on the difficulty or complexity of the task. Proximity to a source of informational advantage can help and simplify the complexity of the valuation task, thus improving the forecasting accuracy. Empirical evidence is inconclusive on this issue, but there is some evidence that geographical distance between the analyst and the followed company is an important factor in forecast accuracy. Other authors argue that local analysts may gather better quality and more timely information about the company, thereby gaining an informational advantage over their peers, the so-called local information advantage.

We do not fully agree with this theory. In fact, in such a globalised context, where physical presence can be easily substituted by virtual contact and distances are shortened by technology which facilitates communication, we argue that the physical proximity of analysts to firms is not associated with an informational advantage. Therefore, in our hypothesis, the information advantage derives from another form of geographical proximity which is more industry knowledge-related.

We therefore provide a new concept of proximity related to distance from centres of knowledge, which we define as hubs of expertise. While Malloy (2005) measured proximity as the number of kilometres between analysts and firms and Bae et al. (2008) defined an analyst as local if he or she was located in the same country of the followed firm, in our study, the distinction between local and foreign analysts is based on the analyst's location with respect to the hubs, which are identified by looking at the industrial specialisation of countries.[3]

[3] See also Section 3 and the Appendix to the paper.

According to the comparative advantage theory (Ricardo (1963)), each nation tends to shift its resources to its more productive industries, while increasing trade for goods in their less productive industries. So, each nation tends to have a specialisation in a specific industry. This is often associated with the development of industrial clusters. In the sphere of financial services, for instance, previous research has shown that large, medium and small financial service companies have a tendency to cluster in metropolitan areas because of the need to access a large pool of specialist and support services (e.g. accounting, actuarial, legal etc.), be in close proximity to the markets, benefit from agglomeration economies, reduce transactions costs, develop and innovate intrinsic skills through the sharing of knowledge and practice (e.g. Davies, 1990).

Since clusters provide knowledge-rich environments which are associated with innovation, knowledge spillover, the building of relationships and synergies, proximity to these centres of specialisation may allow financial analysts to improve their knowledge of value-relevant factors and use them to their advantage in the evaluation of companies in that industry. Therefore, in our opinion, the geographical proximity of financial analysts to hubs of expertise improves the quality of industry knowledge and allows analysts to develop unique expertise and skills, resulting in an informational advantage and greater forecast accuracy.

3. Hypothesis development and research design

3.1. Modelling the analyst accuracy

The set of information that analysts use to evaluate a company can be divided in two groups. A first group composed by commercial and structured information, easily collected by all analysts and a second group of soft (tacit) information that can be privileged and produced by the environment in which analysts work. Therefore, an analyst has an information advantage if it has access to the soft information, derived by his context.

Our basic hypothesis is that analysts located close to sources of soft knowledge have an information advantage. Therefore, the primary aim of this research is to test whether the accuracy of financial analysts depends of their proximity to hubs of expertise, generating soft knowledge.

The conceptual model used is therefore:

Analyst Forecast Accuracy = f (analyst stock of knowledge deriving from the proximity to the hubs, other control variables)

We adopt two estimation techniques in order to investigate the accuracy of financial analysts and in both cases we employ the Newey-West procedure[4] in order to provide consistent inferences on the estimated coefficients.

The former is a classic OLS regression, assuming a linear relationship between the analyst's accuracy, which is our dependent variable, and all of the independent variables.
The latter is a fixed effects model based on the within transformation. This model allows us to take into account the differences between the firms covered which are not controlled by our independent variables, thereby allowing us to manage time-series observations and cross-sectional units at the same time. We stress that the assumption underlying the fixed effect model is that the relationship between the explained and the explanatory variables is assumed to be constant both cross-sectionally and over time.

As a measure of relative forecast accuracy, we initially made two different definitions of accuracy: a simple and a more sophisticated one. The simplest one (*AFE*) is the Absolute Forecast Error calculated as:

$$AFE_{ijt} = \frac{ACTUAL_{jt} - FORECAST_{ijt}}{ACTUAL_{jt}} \tag{1}$$

where *ACTUAL* indicates the actual earnings per share for the company j in the fiscal year t and *FORECAST* is the forecast of earnings per share, issued by the analysts i for the company j in fiscal year t, no more than 100 days before the announcement date. As previous research has proved, this measure is too naïve and can be biased.[5] We also defined another measure, the Proportional Mean Absolute Forecast Error (*PMAFE*), calculated as:

$$PMAFE_{ijt} = \frac{AFE_{ijt} - AAFE_{jt}}{AAFE_{jt}} \ (-1) \tag{2}$$

This measures the difference between the absolute forecast error (*AFE*) of analyst i forecasting earnings for firm j in the fiscal year t and the average absolute forecast error across all analyst forecasts of firm j's fiscal year t earnings, expressed as a fraction of the average absolute forecast error across all analyst forecasts of firm j's fiscal year t earnings. *PMAFE* controls for firm-year effects by subtracting the mean absolute forecast error, *AAFE*, from the analyst's absolute forecast error. Deflating by *AAFE* reduces heteroskedasticity in forecast error distributions across firms (Clement (1999)) and multiplying by -1 ensures that higher values for *PMAFE* correspond to higher

[4]Brooks (2008) explains that the Newey-West procedure implies "HAC" (Heteroscedasticity and Autocorrelation Consistent) standard errors. This adjustment allows us to deal with the coefficients' standard errors since it produces a variance-covariance estimator which is consistent in the presence of both heteroscedasticity and autocorrelation.
[5] Clement (1998) documented that *PMAFE* improves the chances of identifying the differences in individual analyst forecast accuracy. Jacob et al. (1999) discussed these benefits in more detail.

levels of accuracy.[6]

Jacob et al. (1999) explain that the *PMAFE* variable is a relative measure of forecasting accuracy which is not affected by inter-temporal changes and cross-sectional differences in the price-to-earnings ratios. We rely on this variable in order to compare data from different firms and different years which could therefore allow us to provide more interesting figures on the relationship between knowledge and analyst accuracy.

3.2. Modelling analysts' stock of knowledge and other control variables

In order to assess whether the analyst's location with respect to hubs of expertise influences the quality of their knowledge and enhances the accuracy of their forecasts, we first identify the hubs, where the spill-overs of knowledge originate. Secondly, we test whether the accuracy of financial analysts' forecasts depends on their proximity to the source of spill-overs (the hubs) identified.

Since empirical measurement of knowledge spill-overs would be impossible because "knowledge flows are invisible, they leave no paper trail by which they may be measured and track[ed]" (Krugman, 1991, 125), we draw on concepts from industrial and international economics to find a proxy.

Specifically, we assume that there are three alternative methods for the study of knowledge generation: cluster-, sector- and network-based approaches. All of these three approaches are based on the basic assumption that the intensity of knowledge generated by a sector is related to its level of production, but they offer different ways to measure it.

The first approach is based on the idea that the knowledge derives from intensive and privileged exchanges amongst industries which are strongly related in agglomerates of sectors (clusters). In this approach, the structural relationships among sectors which characterise a cluster produce privileged knowledge. Therefore, even though a sector may be small, it is part of a cluster, and therefore generates an amount of knowledge dependent on the cluster of which it is part. Cluster literature explains how clusters retain privileged knowledge which can be spread amongst their members.

Since Marshall's (1920) seminal discussion about highly localised districts in the UK, a new perspective has been developed about the geographical clustering of firms from similar industries.[7]

---

[6] As in all the regressions these latter accuracy measures were the best, and consistently with previous literature (see Clement (1999 and 1998) and Jacob et al. (1999)), we report only the results obtained using this accuracy measure.

[7] See Storper (1997) for a review.

More recently, some pieces of research have conceptualised clusters as sources of enhanced knowledge creation (e.g. Lawson & Lorenz, 1999; Lorenzen & Maskell, 2004; and Malmberg & Maskell, 2002). In this regard, participating in a cluster would increase the spill-over effects of new technologies, knowledge and innovations. For instance, Forni and Paba (2001) show how strong linkages induce a relatively fast diffusion of knowledge and new technologies. Cluster analysis provides a possible solution to the identification of strongly interrelated sectors. By dividing the economic system into clusters of interrelated sectors, the clusters show exactly which sectors are closely related to each other.

Therefore, we associate this kind of 'clustered' knowledge with our concept of hubs of expertise, the source of shared knowledge.

The very basic definition of an industrial cluster is "geographical concentrations of industries that gain performance advantages through co-location" (Doeringer and Terkla (1995), page 225). This definition of clusters is similar to that of agglomeration economies, but in fact it is within industrial clusters that agglomeration economies are likely to be observed. Beyond the basic definition, however, there is little consensus on how to define an industrial cluster. Michael Porter extended the concept of industrial clusters in his book, *The Competitive Advantage of Nations* (1990) and developed the 'Diamond of Advantage', four factors which create a competitive advantage for firms. The four corners of the diamond include factor conditions, demand conditions, industrial strategy, and related and supporting industries. He used this diamond to determine which firms and industries had competitive advantage. A more in-depth discussion of the different definitions of industry clusters was presented by Jacobs and DeMan (1996) and Rosenfeld (1996, 1997).[8] They expanded on the definitions of vertical and horizontal industry clusters in order to identify the key dimensions which can be used to define clusters. These include the geographic or spatial clusters of economic activity, the horizontal and vertical relationships between industry sectors, the use of common technologies, the presence of a central actor (e.g., a large firm, research centre, etc.), the quality of the firm's network and its level of co-operation (Jacobs and DeMan (1996)). In addition to vertical and horizontal relationships, Rosenfeld (1997) included criteria for defining a cluster, including its size, economic or strategic importance, the range of products produced or services used and the use of common inputs. He did not define clusters exclusively by the size of the industries or the scale of employment.

We assume that analysts have access to enhanced and privileged knowledge on the basis of their geographical proximity to clusters and also benefit from a cultural information advantage that

[8] Jacobs and DeMan (1996, p. 425) argue that "there is not one correct definition of the cluster concept…different dimensions are of interest."

improves forecasts issued for companies (local or not) which belong to the same sectors of the proximate cluster. To be close to a cluster is in fact a source of informative advantage.

The second perspective that we use is a sector-based approach. The theory behind this approach is that if a sector is very productive in terms of output, it also has a strong competitive position with respect to other sectors. This advantaged position within the national economy generates specialisation and greater knowledge generation than other sectors less relevant for the economy. To be close to relevant sectors in terms of output can be a source of informational advantage.

Finally, the third approach is based on network logic. The basic assumption is that knowledge is generated by sharing and is the effect of cross-fertilisation between sectors, composing a network., The intensity of the knowledge therefore depends on the exchanges between the sectors of the network. In this case, proximity to the most relevant nodes of a network could be a source of informational advantage.

Our operational framework is therefore based on three steps in order to assess three proxies of knowledge intensity.

The first step is to define the hubs as industrial clusters. The empirical identification of clusters is not a straightforward procedure and the related literature shows how tricky it can be. There are no conclusive solutions for this.

Economic theory suggests several methods for identifying clusters. However, Hoen (2002), after describing how cluster analysis contributes to the study of linkages among sectors, shows that the cluster identification method based on a block diagonal matrix,[9] called the *diagonalisation method*, gives the best results. For this reason, we use this latter method in our analysis. According to this approach, we start from the input-output (I-O) matrix of each country. First, we calculate an I-O matrix of only the intermediate consumption[10] of different industries of a country. Therefore, the main diagonal elements, which represent the intermediate consumption of the same industry, are zeros. The off-diagonal elements are expressed as a percentage of the largest intermediate consumption between two industries, the benchmark for which has been set as equal to 100%. As per the literature, we also set a minimum threshold for input and output entries for being part of the matrix at 2%. After setting all of the elements that do not satisfy these restrictions to zero, putting the matrix in the block diagonal form shows which sectors belong to which clusters.[11] Each off-

[9] A block diagonal matrix can be split up in parts that have no connection with each other. By rearranging sectors appropriately (details of this method are reported in Appendix A), the matrix would look like blocks of matrices along the main diagonal.

[10] The intermediate consumption is an economic concept that represents the monetary value of goods and services consumed or used as inputs in production by firms of a sector in a country.

[11] There are several possible algorithms for making the block diagonal matrix by rearranging sectors. Appendix A describes an algorithm which does not involve complex computations and is easy to program. An algorithm based on

diagonal value (S_{ij}) in the same block (cluster) indicates the intermediate consumption between two sectors that is greater than the selected threshold. According to Hoen (2002), S_{ij} represents the strength of the link between two industries, *i* and *j,* belonging to the same cluster.

Once the different clusters (hubs) have been identified, we need to attribute a value for the knowledge spill-overs coming from each one. Assuming that the level of knowledge spill-over depends on the level of total production achieved by related sectors belonging to the same cluster, as defined by Hoen's procedure, we define a first proxy (*CLUSTER*). It is measured as the log transformation of the sum of S_{ij} of each cluster. In more formal terms, *CLUSTER* is:

$$CLUSTER_{zx} = Ln\sum_{ij} S_{ij} \tag{3}$$

where *z* indicates the country, and *i* and *j* two of the sectors composing the cluster *x*.

In other words, *CLUSTER* is a proxy of the level of information spill-over of which local analysts can take advantage of and it is based on the relevance of a sector depending on the cluster (approximated by the total production) of which it is part. Therefore, focusing on analysts' geographical location, we associate with each of them the value of the *CLUSTER* variable, depending on their location.

Let us assume, therefore, that a UK-based analyst evaluates (UK or foreign) companies in the oil industry. According to our framework, we will attribute to this analyst the stock of knowledge measured by the *CLUSTER* variable assigned to the UK of the cluster containing the oil sector. Should a specific sector not be contained in any of the identified UK clusters, the value of the variable will be forced to zero. Therefore, analysts located in different countries will benefit from different stocks of knowledge assigned to the clusters identified in their own country.

Analysts close to the most important clusters will show higher *CLUSTER* values, indicating higher spill-overs and informational advantages, which help them to issue more accurate forecasts in relation to national or international companies in industries belonging to that cluster. Thus, we expect this variable to have a positive impact on the accuracy of earnings forecasts.

The second approach when dealing with the hub identification issue is sector-based. Input-output tables are a useful tool used in the literature for studying the linkages between industries as they allow the measurement of the effect of a specific sector on the other sectors or the effect of each sector on the economic system as a whole. Therefore, by using the input-output tables, we can measure the importance of a sector in a country in terms of its production and level of specialisation. We can then assume that the value of production of a sector is a proxy for the

---

eigenvalues, which has the advantage of ordering clusters according to the strength of their linkages, can be found in Dietzenbacher (1996).

knowledge produced. According to this framework, each sector represents a hub, but hubs with higher values of production contain more important sectors for the overall economy of a country.
We start with a Use table[12] and calculate the variable *OUTPUT* by industry, which is defined as the sector output at basic prices (without considering relationships with other sectors). The reason for doing this is to measure the total output value produced by each sector. This variable is measured as the log transformation of the sum of the intermediate consumption and the value added of each sector, scaled for the country's power purchase parity (*PPP*), in order to compare the same variables across different countries. In more formal terms:

$$OUTPUT_{zi} = Ln\left[\frac{(intermediate\ consumption_{zi} + value\ added_{zi})}{PPP_z}\right] \qquad (4)$$

where z indicates the country, while i the industry. The informative value of *OUTPUT* is that it shows how much a sector is relevant in terms of production for the economy.
Therefore, since our framework production is associated with knowledge, a higher value of the variable with respect to a certain sector i are in general related to higher levels of knowledge spill-over spreading to that sector. Hence, similarly to the *CLUSTER* variable, we predict this variable to have a positive impact on the accuracy of local analysts' forecasts as a higher level indicates a greater informative advantage for them. For example, let us assume that a UK-based analyst evaluates a (UK or foreign) bank. According to our framework, we attribute to this analyst the stock of knowledge, measured by the *OUTPUT* variable, assigned to the financial sector in UK. Therefore, analysts located in different countries benefit from different stocks of knowledge depending on the sector's relevance, in terms of output, for the country.
Finally, we also apply a third approach in order to identify hubs, based on methods from social network analysis. We assume that the economy of a country can be represented as a network of sectors (nodes) which are more or less interrelated. The ties among the nodes measure the strength of their relationships.
Similarly to clusters, networks can also produce spill-overs of knowledge. The extensive literature on this field has generated a wide set of techniques and related measurements for capturing the many facets of information embedded in the network structure.
One of the primary aims of social network analysis is to identify the 'important' actors in the network. The concepts of centrality and prestige have been introduced in the network field in order to quantify an individual actor's prominence within a network by summarising structural

[12] Please see Appendix A for a detailed general definition of this table.

relationships among the g nodes. We draw from this literature to assess the prominence of economic centres (hubs) across countries, using the tools which it suggests.

In order to do this, we replicate the procedure proposed by Cetorelli and Peristiani (2009). The authors, using methodologies developed in social network analysis, elaborate measures to rank the relative degree of dominance of financial centres around the world. With such measures, they were able to assess more effectively whether US financial markets have lost their position of global leadership and the extent to which competition from other centres may have strengthened over time. The most complete measure which they implemented was the 'prestige index'. In network analysis, indices of prestige allow for the measurement of the dispersion or inequality of the prominence of all of the actors. Formally, the prestige index (*Pr*) for a node (in our case, a sector) i (ni) is calculated as:

$$Pr\ (n_i) = x_{1i}\, P(n_1) + x_{2i}\, P(n_2) + \ldots + x_{Ni}\, P(n_N) \qquad (5)$$

where the weights are represented by the flows from each of the nodes of the network onto n_i.

Therefore, by adapting and applying Cetorelli and Prestiani (2009)'s procedure for financial centres, we assess a prestige index for each sector (node) of all the countries (networks) of the dataset.

In order to represent the network, its nodes and the ties between nodes, we assume that the production value between sectors and within the same sector can be a proxy of the links of the network and indicate its knowledge intensity. Therefore, we measure flows between nodes through tables of intermediate consumption and the values added of each country. Specifically, the production of each country is represented by a matrix which exhibits the flow of intermediate consumption between each pair of sectors on the off-diagonal entries, while the main diagonal shows the sum of the intermediate consumption flow within each sector and the sector value added.

We apply the algorithm (5) proposed in the literature by Cetorelli and Peristiani (2009) to the whole network. Therefore, we have N equations in N unknowns for each network.

As shown by Katz (1953), this system has a finite solution if one first standardises the original network matrix. For this reason, firstly, we divide each column of the matrix by the column's sum.

After this standardisation, the system of equations becomes a more common matrix-characteristic equation, where the solution (that is, the vector of prestige indicators) is the eigenvector associated with the largest eigenvalue of the standardized matrix (SM). Since we do not use any specific mathematical software to calculate the eigenvalue of the matrix, we apply the 'power method'.[13] This is an iterative method and thus does not require any specific software to solve the problem. In order to apply this method, we raise each cell value to the *n*th power until the matrix converges into a table of equal vectors (by column) so that we can find out the eigenvector associated with the largest Eigen value of the matrix.[14] This eigenvector contains the index of prestige associated with each sector. We call this eigenvector a *NET* variable and it is dependent on the flows exchanged (approximated to the intermediate consumption) between sectors in the same country. In more formal terms, for a country/network *z*, the variable *NET* can be calculated as:

$$NETz = eigenvector\ of\ SMz \tag{6}$$

A node (sector) will thus have high prestige if it is chosen, in terms of flows, by a low number of highly prestigious other nodes or by a high number of other nodes with lower index value.

NET is therefore also our proxy for the extent of knowledge spill-overs of different sectors in a specific country. Its informative value is that it allows the knowledge of how much a sector is relevant in terms of exchanged flows of knowledge (approximated by the intermediate consumption) between the different sectors of the network. A greater value of this variable is associated with higher levels of knowledge spill-overs spreading from the specific sector *i*. Each analyst will be associated with a *NET* value, depending on their location and the sector which they are evaluating. For instance, a UK-based analyst evaluating a (UK or foreign) bank has a stock of knowledge measured by the *NET* variable, which is assigned to the financial sector in UK in relation to the 'prestige' of this sector with respect to others in the same country. Therefore, analysts located in different countries will benefit from different stocks of knowledge depending on the sector's relevance to the country.

Hence, similarly to the aforementioned proxies, we predict that this variable has a positive impact on the accuracy of local analysts' forecasts as a higher level indicates a greater informative advantage for them.

Therefore, in more formal terms, the model that we test is:

---

[13] In mathematics, the power iteration is an algorithm: given a matrix A, the algorithm will produce a number λ (the eigenvalue) and a non-zero vector v (the eigenvector), such that $Av = \lambda v$. The algorithm is also known as the Von Mises iteration.

[14] Appendix B reports further technical details.

$$ACCURACY_i = \alpha + \beta_1\, STOCK\ OF\ KNOWLEDGE_i + \beta_2\, CONTROL\ VARIABLES_i + \varepsilon_i \qquad (7)$$

where the *ACCURACY* variable is defined in Section 3.1 and the *STOCK OF KNOWLEDGE* is alternatively measured by the *CLUSTER*, *OUTPUT* and *NET* variables.

With regard to the *CONTROL VARIABLES*, we include in the model a limited number of control variables because of the small size of the sample. Specifically, we insert the variable *AGE*, measuring the number of days from the date of the release of the report to the end of the fiscal year and *VOL*, which is a control variable measuring the coefficient of the variation in the firm's quarterly EPS over the past three years. We hypothesise that the greater the variability of the actual EPS over time, the greater the complexity of the analysts' forecasts. Finally, we employ dummy variables in order to control for the inter-temporal changes in analyst accuracy.[15] We do not expect to find any significant results as the *PMAFE* variable provides an adjustment for the inter-temporal variability of the analysed topic itself (see Section 3.1.).

4. Data

The sample construction started with a rich dataset of observations on analyst forecasts collected from Factsect over four fiscal years, from 2005 to 2008.

For each earnings forecast, we have the research date, the recommendation issued, the previous research date and forecast by the same analyst, and the type of report issued. We also collected information about analyst characteristics, such as their full name, brokerage house and office telephone number. The latter allowed me to infer their geographical location. As we had some missing data with regard to the last piece of information, we collected some of them by hand from *Nelson's Directory of Investment Research*, which provides extensive information about analysts, which companies they follow and brokerage houses. Each volume of *Nelson's Directory* in year *t* is based on the analysts' information from year *t-1*. As the 2008 volume is not available, we used the 2007 volume as a proxy for the 2007 analysts' missing data. As we did not find any clear information on some analysts, we excluded these missing observations.

These raw data needed to be filtered in order to match the restrictions based on the aim of our research.

Firstly, since the computation of the knowledge variable based either on the cluster or the sector

[15] We include *D05*, which is a dummy variable equal to 1 if the observation obtained was in 2005, 0 otherwise; *D06* is a dummy variable equal to 1 if the observation obtained was in 2006, 0 otherwise; and *D07*, which is equal to 1 if the observation obtained was in 2007, 0 otherwise.

concept is only available for specific countries, i.e. Italy, Germany, the United Kingdom and France, we eliminate all of the observations associated with analysts who are not located in these countries. Moreover, we cancel out all of the observations produced by teams of analysts placed in different countries. Following these adjustments, we can manage a dataset which satisfies our assumptions about the relevance of the proximity between the location of analysts and the hubs of expertise.

Secondly, we identify the end date of the fiscal year and eliminate all of the analyst reports released more than one hundred days before this reference point. We adjust the data in this way in order to have homogenous annual EPS forecasts.

Furthermore, Jacob et al. (1999) point out that each analyst benefits from both public information released by firms and previous information released by other analysts. In order to control for these sources of information, we compute a control variable which represents the age of the forecast, assuming that more recent reports benefit from the information released in earlier firm reports and from new public information.
The *CLUSTER* variable is adjusted to the data from the 2002 input-output tables because of a lack of data availability. Moreover, the *OUTPUT* variable is also based on these 2002 input-output tables. The *NET* variable is entirely based on data from 2000 and 1995 for the UK. We expect that this is not a big issue as our variables capture the structural relationships amongst the sectors which should not constantly change over time.

The final dataset is composed of 205 observations related to 33 firms, from 2005 to 2008.

The stated variables can be summarised by their descriptive statistics in Table 1.

Insert Table 1.

On the basis of these statistics, we could assert that, on average, analysts provide accurate forecasts but with relevant outliers and high variability amongst them. Moreover, in looking at the coefficients of variation, we can appreciate how the analyst accuracy variable exhibits a coefficient which should require more explanatory variables in order to be almost completely explained. However, since this research aims to focus on the role of knowledge in analyst accuracy and the sample size is not large, we only focus on the aforementioned explanatory variables. Table 2 reports the correlation matrix among variables.

Insert Table 2.

Furthermore, we notice that the literature on the argument provides regressions in which the adjusted R-squared hits a value of approximately 0.15, which is an argument for this type of research in order to shed more light on the topic.

The firms which comprise the final sample belong to ten different sectors and seven countries, as summarised in Table 3.

Insert Table 3.

The dataset does not have many control variables but we assert that this provision could be sufficient since this attempt only represents the preliminary stage of the research on the relationship between knowledge and analyst accuracy.

5. Results

We start the empirical analysis by applying the OLS technique, running different models in order to examine the impact of each knowledge variable on *PMAFE*, which represents analyst accuracy.

Insert Table 4.

All of the models are somewhat poor in explaining analyst accuracy. Firstly, all of the values of the F-statistic only allow us to argue that all of the coefficients are not significantly different from zero. Similarly, all of the t-statistics of the knowledge variables only allow us to argue that each knowledge variable is not significantly different from zero. Finally, the adjuster R-squared is negligible for all of the models; therefore the dependent variable could be better explained by looking at its mean.

From this discussion we could appreciate the apparent relevance of the *AGE* coefficient, which confirms our expectation that more recent reports would benefit from past analyst reports and incremental public information.

We explain these preliminary results by recognising that the OLS estimate does not account for the differences between firms. In order to control for these unexplained dynamics, we should analyse the analysts' accuracy conditionally on the firm's identity. Indeed, by using this perspective, we can appreciate the impact of the independent variables in explaining the variability of the analysts' accuracy without the requirement of control variables at the firm level. Moreover, these firm level control variables could be regarded as omitted variables, and thus their absence could debase the OLS results.

Since we only have the *VOL* variable as a control variable at the firm level, we employ a within transformation at firm-level in order to tackle this issue. We label the new variables with the suffix '*D*' and eliminate the constant term on the basis of the within transformation.

Insert Table 5.

After controlling for company identity, we notice that *AGE* is still significant, which is consistent with our expectation. Indeed, we can confirm that an analyst gains in accuracy when he or she provides his or her report close to the actual EPS issue. This figure could be motivated by our assumption of the benefits of incremental public information and the information contained in reports previously released by other analysts.

The control variable *VOL* is not significant. On the basis of this result, we cannot confirm our expectation of a negative relationship between the coefficient of variation of the historical actual EPS and analyst accuracy. A plausible reason for this result is that the literature refers to the variability of the actual EPS only in order to explain the distribution of the analyst forecasts, not the topic of analyst accuracy.

We also note that the dummy variables do not provide any contribution to the explanation of the dependent variable. This is probably due to the formulation of the *PMAFE* variable, since it accounts for inter-temporal changes in analyst accuracy.

This introduction allows us to focus on the explanatory power of our knowledge variables. First of all, we notice that *CLUSTER_D* is not significantly different from zero (Model 2). This result could derive from the aforementioned difficulties in measuring cluster boundaries and the knowledge contained therein. Therefore, this inconsistency could be caused by the drawbacks in the procedure of identification of clusters and in the representation of cluster knowledge.

Following this argument, we use the *NET_D* variable, which recognises hubs of expertise at a sector level rather than at a cluster level (Model 3). From this setting, we report a coefficient that is significantly different from zero. Focusing on knowledge at a sector level allows us to confirm our expectations on the role of proximity in increasing analysts' stock of knowledge. The positive sign of the coefficient means that the proximity between the analyst and the hub of expertise represents a source of analyst accuracy. Moreover, the adjusted R-squared of Model 3 increases significantly after the inclusion of the *NET_D* variable. Above all, if we multiply the coefficient of *NET_D* by this standard deviation, we can evaluate the variable impact on analyst accuracy between -8% and 8%.

This result is a preliminary confirmation of the relevance of analyst proximity to hubs of expertise. It is obtained by elaborating on the input-output tables on the basis of network analysis. As explained above, we measure sector knowledge on the basis of an index which represents the sector prestige recognised by all of the other sectors of the national economy. In order to check the usefulness of the network analysis, we define a third variable which considers sector output as a proxy of the stock of knowledge within the sector. In reality, this choice of proxy is not arbitrary since it represents the variable which we have split in the network information matrix and then elaborated in order to obtain the *NET_D* variable. If we obtained the same results, we could assert that all of the information on sector knowledge is contained in the sector attributes. Therefore, the analysis of sector ties should not provide incremental information.

We notice that the coefficient of the *OUTPUT* variable (Model 4) is not significantly different from zero. On the basis of this result, we confirm the benefits of exploiting network analysis in order to trace the availability of knowledge amongst units of analysis.

To sum up, network analysis synthesises network interactions, thereby providing a holistic analysis of knowledge amongst sectors within a national economy. Moreover, thanks to this approach, we demonstrate the relevance of knowledge about production to explain analyst accuracy on the basis of proximity to centres of knowledge.

6. Conclusions

This research aims to provide new insights on the issue of analyst accuracy, by developing a set of variables which should represent part of the stock of knowledge owned by analysts and help them in their task.

First of all, we point out that the analysis of analyst accuracy is essential in order to increase employers' reputations. Investment banks and brokerage houses would offer the services of analysts for free in order to benefit from analysts' reputations, a fact which is recognised by the financial markets.

Secondly, we argue that prior research on analyst accuracy has been more about analysts' characteristics rather than on the knowledge which is available to them. We recognise the utility of the first approach but try to develop the knowledge framework in order to provide a new stream of research.

We point out that each analyst has a certain stock of knowledge available: the firm's public information and the information contained in previous reports. The definition of our knowledge variables refers to the analyst's personal knowledge. We assume that the concept of proximity is essential for the detection of this source of personal expertise.

The results of this research confirm our main expectation since we find some evidence of greater accuracy associated with forecasts issued by analysts who are close to so-called hubs of expertise. This result is not based on the concept of cluster since the empirical identification of clusters is not straightforward. We ground our results by considering that hubs of expertise represent knowledge associated with single sectors of a national economy. Using this perspective, we report on the relevant role of sector knowledge on production for local analysts, even if they cover firms which are established abroad.

We conclude this research by suggesting plausible steps in order to improve the analysis. Firstly, further improvements are needed in terms of cluster identification and the knowledge available to the analyst. We have suggested network analysis as a reliable algorithm which could be used to detect concentrations of sectors within the input-output framework, thus providing a new approach to the evaluation of the stock of knowledge within the clusters. We suggest the development of this network analysis in order to measure the stock of knowledge within each unit of analysis. In the next steps of this work, each analyst will represent a unit of analysis.

Secondly, we suggest increasing the number of observations since the size of this sample is only reliable for preliminary results and insights. The plausible direction is to expand the group of countries analysed, thereby providing a more comprehensive picture of the European continent. Furthermore, it would be useful to collect firms' quarterly EPS's in order to increase the number of observations over time.

Finally, we propose the merger of these knowledge variables with the explanatory variables based on analyst characteristics. Using this, we could verify the relationships between these two classes of explanatory variables in order to improve our understanding of analyst accuracy.

Tables

Table 1: Summary statistics of the main variables of the dataset.

Statistics	*PMAFE*	*AGE*	*VOL*	*CLUSTER*	*NET*	*OUTPUT*
Mean	0.0204	50.85	0.2462	6.75	0.0379	11.31
Median	0.0277	49.00	0.2102	7.90	0.0276	11.36
Max.	0.9947	100.00	0.9141	10.41	0.2108	12.65
Min.	-2.0789	2.00	00.0267	0.00	0.0023	9.94
Std. Dev.	0.4667	24.27	0.1928	3.17	0.0433	0.5932
Coeff. of variation	22.87	0.47	0.78	0.46	1.14	0.05

Notes: This table reports the main descriptives of the model variables. The *PMAFE*, representing the analyst accuracy, while *CLUSTER*, *NET* and *OUTPUT* are defined above and represent alternatively measures of different analysts knowledge. *VOL* and *AGE* are 2 control variables indicating, respectively, the company earnings volatility and the age of the analysts forecast.

Table 2. The correlation matrix among variables

Panel A. The Pearson's correlation.

	pmafe	age	lncluster	net	lnoutput	vol
pmafe	1					
age	-0.1489*	1				
lncluster	-0.0623	-0.1864*	1			
net	0.0451	0.0458	0.2062*	1		
lnoutput	-0.0482	0.0334	0.5081*	0.5157*	1	
vol	-0.0025	0.0315	-0.0601	-0.1307*	0.044	1

This table reports the correlation matrix of the different model specification variables. It is based on the Spearman's correlation definition.

* denotes significance at the 10%.

Table 2. The correlation matrix among variables

Panel B. The Spearman's correlation.

	pmafe	age	lncluster	net	lnoutput	vol
pmafe	1					
age	-0.1980*	1				
lncluster	-0.1288*	-0.0872	1			
net	0.0923	-0.0793	0.0218	1		
lnoutput	-0.1431*	0.0071	0.6989*	0.2171*	1	
vol	-0.1187*	0.0693	-0.0766	-0.076	0.0222	1

Notes. This table reports the correlation matrix of the different model specification variables. It is based on the Spearman's correlation definition.

* denotes significance at the 10%

Table 3: Sector and country weights in the dataset.

Sector weight	%	Country weight	%
Banks	24.24%	**Finland**	0.00%
Insurance	12.12%	**France**	18.18%
Telecommunications Services	15.15%	**Germany**	18.18%
Technology	9.09%	**Italy**	9.09%
Non-Cyclical Consumer Goods & Services	9.09%	**Netherlands**	18.18%
Energy	6.03%	**Spain**	6.06%
Pharmaceuticals	12.12%	**Sweden**	0.00%
Utilities	3.03%	**Switzerland**	9.09%
Healthcare	6.06%	**United Kingdom**	21.21%
Basic Materials	0.00%		
Cyclical Consumer Goods & Services	3.03%		

Notes. This table reports the weights of each industry and country in the whole dataset.

Table 4: The effect of different knowledge variables on the analysts' accuracy – OLS estimation

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	0.247***	0.267***	0.225**	0.527	0.986
AGE	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
VOL	-0.023	-0.023	-0.007	-0.019	0.015
CLUSTER		-0.003			-0.0001
NET			0.697		12.013
OUTPUT				-0.025	-0.069
D05	-0.094	-0.088	-0.087	-0.093	-0.081
D06	-0.104	-0.098	-0.106	-0.099	-0.093
D07	-0.108	-0.100	-0.124	-0.103	-0.121
R-squared	0.031	0.032	0.035	0.032	0.041
Adj. R-squared	0.007	0.002	0.006	0.003	0.002
Durbin Watson	2.055	2.052	2.056	2.05	2.061
Prob (F-stat)	0.266	0.366	0.299	0.357	0.397

OLS Estimates; *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

Notes: This table describes the main results obtain by OLS estimations of different model specifications. The independent variable is always *PMAFE*, representing the analyst's accuracy, while *CLUSTER*, *NET* and *OUTPUT* are defined above and represent alternatively measures of different analysts knowledge. *VOL* and *AGE* are 2 control variables indicating, respectively, the company earnings volatility and the age of the analysts forecast. Finally, *D05*, *D06* and *D07* are dummy variables controlling for a time effect.

Table 5: The effect of different knowledge variables on the analysts' accuracy – Fixed effect estimation

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
AGE_D	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***
VOL_D	0.060	0.068	0.166	-0.005	0.16
CLUSTER_D		-0.026			-0.023
NET_D			9.59*		9.40*
OUTPUT_D				-0.164	-0.009
D05	0.0008	0.001	0.007	-0.0005	0.007
D06	0.0013	0.002	0.016	0.006	0.017
D07	0.0002	-0.0008	-0.016	-0.003	-0.017
R-squared	0.041	0.047	0.069	0.045	0.075
Adjusted R-squared	0.021	0.024	0.045	0.022	0.042
Durbin-Watson stat	2.084	2.072	2.10	2.08	2.09

Fixed-effect Estimates; *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

Notes: This table describes the main results obtain by Fixed effect estimations of different model specifications. The independent variable is always *PMAFE*, representing the analyst's accuracy, while *CLUSTER*, *NET* and *OUTPUT* are defined above and represent alternatively measures of different analysts knowledge. *VOL* and *AGE* are 2 control variables indicating, respectively, the company earnings volatility and the age of the analysts forecast. Finally, *D05*, *D06* and *D07* are dummy variables controlling for a time effect.

APPENDIX A

This appendix illustrates how we identify and define the hubs of expertise following a cluster-based approach.

The key point is the definition of an index allowing us to rank the European sectors by knowledge intensity. The methodology, based on the literature about cluster theory, is for the identification of two different proxies of hub of expertise.

Firstly, in order to identify the boundaries and size of hubs of expertise, we apply the Hoen algorithm (Hoen, 2002), based on symmetric input-output tables.[16] Hoen asserts that any sector needs linkages with other sectors in order to develop its own business. The symmetric input-output tables of different countries should represent these relations. Therefore, by analysing the input-output tables, we identify the strongest linkages between sectors and thus the clusters, i.e. aggregations of sectors within a national economy.

Input-output analysis can be used to evaluate the impact of different policies on macroeconomic variables, such as gross domestic product, employment, consumption, productivity, competitiveness, etc, as well as the environment. In the 1930s, the economist Wassily Leontief described the inter-industry relations in the economy from which it had developed. The structure of each sector's production activity was represented by appropriate structural coefficients, which described in quantitative terms the relationships between the inputs that it absorbs and the output that it produces. The input-output framework was based on three types of table: supply tables, use tables and symmetric input-output tables. A synthetic description of each table is given below.

Eurostat defines the supply table as a product-by-industry-based table, in which products are placed in the rows and industries and imports in the columns. A simplified illustration can be represented in the following way:

[16] The OECD defines an input-output table as a tool for the presentation of a detailed analysis of the process of production and the use of goods and services (products), and the income generated in that production for any European country.

<table>
<tr><td rowspan="2">Industries
Product</td><td colspan="3">Industries</td><td rowspan="2">Import</td><td rowspan="2">Total</td></tr>
<tr><td>Agriculture</td><td>Industry</td><td>Services Activities</td></tr>
<tr><td>Agricultural products</td><td colspan="3" rowspan="3">Output by product and by industry</td><td rowspan="3">Import by product</td><td rowspan="3">Total supply by product</td></tr>
<tr><td>Industrial products</td></tr>
<tr><td>Services</td></tr>
<tr><td>Total</td><td colspan="3">Total output by industry</td><td>Total imports</td><td>Total supply</td></tr>
</table>

Table A.1. A simplified supply table. Source: Eurostat (Eurostat, 18)

The supply table's rows exhibit the supply of goods and services to sectors by the type of product, differentiating between domestic supply and imports. The columns indicate the domestic output of industries by product.

The use table is a product-by-industry-based table with products and components of value added in the rows and industries, categories of final use and imports in the columns. A use table shows the use of goods and services by product and by type of use, i.e. as intermediate consumption by industry, final consumption, gross capital formation or export. A simplified illustration is the following:

<table>
<tr><td rowspan="2">Industries
Product</td><td colspan="3">Industries</td><td colspan="3">Final uses</td><td rowspan="2">Total</td></tr>
<tr><td>Agriculture</td><td>Industry</td><td>Services Activities</td><td>Final consumption</td><td>Gross capital formation</td><td>Exports</td></tr>
<tr><td>Agricultural products
Industrial products
Services</td><td colspan="3">Intermediate consumption by product and by industry</td><td colspan="3">Final uses by product and by category</td><td>Total use by product</td></tr>
<tr><td>Value added</td><td colspan="3">Value added by component and by industry</td><td colspan="3"></td><td>Value added</td></tr>
<tr><td>Total</td><td colspan="3">Total output by industry</td><td colspan="3">Total final uses by category</td><td></td></tr>
</table>

Table A.2. A simplified use table. Source: Eurostat (Eurostat, 20)

The symmetric input-output tables are analytical tables derived from the supply and use system. An input-output table is a quantitative economic tool which represents the interdependencies between different branches of the national economy or different, even competing, economies. The transformation procedure converts the product-by-industry system of the supply and use tables into a product-by-product system or industry-by-industry system. Input-output tables are used to identify economically-related industry clusters and also so-called ‘key’ or ‘target’ industries of a specified economy.

<table>
<tr><td rowspan="2">Products
Products</td><td colspan="3">Homogeneous units of production</td><td colspan="3">Final uses</td><td rowspan="2">Total use</td></tr>
<tr><td>Agricultural products</td><td>Industrial products</td><td>Services</td><td>Final consumption</td><td>Gross capital formation</td><td>Exports</td></tr>
<tr><td>Agricultural products</td><td colspan="3" rowspan="3">Intermediate consumption by product and by homogeneous units of production</td><td colspan="3">Final uses by product and by category</td><td rowspan="3">Total use by product</td></tr>
<tr><td>Industrial products</td><td colspan="3"></td></tr>
<tr><td>Services</td><td colspan="3">Final uses by product and by category</td></tr>
<tr><td>Value added</td><td colspan="3">Value added by component and by homogeneous units of production</td><td colspan="3"></td><td></td></tr>
<tr><td>Imports for similar products</td><td colspan="3">Total imports by product</td><td colspan="3"></td><td></td></tr>
<tr><td>Supply</td><td colspan="3">Total supply by homogeneous units of production</td><td colspan="3">Total final uses by category</td><td></td></tr>
</table>

Figure A.3. A simplified input-output table. Source: Eurostat (Eurostat, 25)

Input-output tables often contain an enormous amount of detailed data. In order to deal with these data, it is necessary to aggregate the data. One possibility is to search for clusters of sectors with strong linkages. The clusters then denote how the sectors may be aggregated (Aroche-Reyes, 2001).

Hoen (2002) developed an algorithm based on these symmetric input-output tables. His algorithm aggregates sectors into clusters after the following rule: two sectors compose a cluster if their relations, the so-called linkages, to economic growth, are large, compared to the whole economic system.

This algorithm is based on the matrices of intermediate consumption across industries. Then, to identify a cluster empirically, the author uses the block diagonal matrix method[17].
As suggested by Hoen, to apply his procedure we first have to set a threshold of significance level for the elements of the input-otput matrix that we use. We use a cut-off point of 2%. Then, we have to select all elements that belong to the 2% of largest elements. Elements that do not satisfy this restriction are put to zero.

Then we have to select, check if the intermediate consumption matrix is decomposable and rearrange the sectors so that the elements are given in blocks.
A block diagonal matrix can be split up into parts with no connections to each other. The algorithm reported below rearranges the sectors appropriately. All of the elements between sectors which are not in the same block are zero. Hence, all off-diagonal blocks would consist entirely of zeros. The zeros denote the boundaries of the clusters, while each block of matrix represents a cluster.

According to Hoen, (Hoen, 2002, 25), the algorithm to use for rearranging sectors and dividing them into clusters is the following one:
Step 1. Start at the upper-left part of the input-output table, with the element in the first column and the first row. The sector belonging to this element is the first temporary cluster.
Step 2. Move to the sector in the next row. Compute the sum of the deliveries from this sector to all sectors of the temporary cluster and the deliveries from all sectors of the temporary cluster to this sector. If this number is zero, go to step 3. Otherwise, add this sector to the temporary cluster and repeat step 2.
Step 3. Move to the next sector and compute the sum of the deliveries from this sector to all sectors of the temporary cluster and the deliveries from all sectors of the temporary cluster to this sector. If this number is zero, go to step 4. Otherwise, repeat step 3. If the last sector is reached, go to step 5.
Step 4. Swap the sector just found with the first sector right below the last sector of the temporary cluster. (For example, if the temporary cluster consists of the sectors 1, 2, and 3, and sectors 4 and 5 have no linkages with the first three sectors whereas sector 6 does, swap sectors 4 and 6). Swap the rows and the columns. Next, add the sector just found and swapped to the temporary cluster (in the example, add sector 6 to the temporary cluster). Continue with the last sector of the temporary cluster (in the example, let's say, sector 6, which is now the fourth row (and column) of the new matrix) and move to step 2.

[17] Hoen (2002) shows that this method brings to same results also selecting other input-output tables.

Step 5. The temporary cluster is now a definitive cluster. Go to the first sector directly beneath the last sector of this cluster. This sector is the starting sector of the new temporary cluster. Move to step 2.

Hoen's algorithm allows us to identify hubs by the intermediate consumption flow in a national perspective.

APPENDIX B

This appendix illustrates how we identify and define the hubs of expertise following a network-based approach.

We base our methodology on Cetorelli and Presotiani (2009)'s approach, which adopted network analysis to deal with a comparison of stock exchanges in a global perspective. Following their procedure, we determine the so-called prestige index which allows us to compare hubs of expertise from different countries. We regard each country as a network and the sectors of tha country as nodes of the network. The production patterns are indicates ties between nodes.

As in the approaches used previously, we start from an input-output matrix and we build a network matrix. Each element of the matrix is considered as a bidirectional flow.

Figure B.1 describes a typical network matrix, used in our framework. The row entries represent the origin of the flow, while the column entries present the destination of it. In this way, the main diagonal accounts for flows due to the sector activity (measured by the sum of intermediate consumption and value added of each sector) and off-diagonal entries represent interactions between different nodes. For instance, I.C.$_{11}$+V.A.$_{11}$ indicates the flow produced and accumulated by industry 1 itself, I.C.$_{12}$ indicates the flow of intermediate consumption from industry 1 (origin) to industry 2 (destination), while I.C.$_{21}$ is the flow of intermediate consumption from industry 2 to industry 1.

	Industry 1	**Industry 2**	**Industry 3**	**Industry 4**	**Industry 5**
Industry 1	I.C.$_{11}$+V.A.$_{11}$	I.C.$_{12}$	I.C.$_{13}$	I.C.$_{14}$	I.C.$_{15}$
Industry 2	I.C.$_{21}$	I.C.+V.A.	0	0	0
Industry 3	I.C.$_{31}$	I.C.$_{32}$	I.C.+V.A.	I.C.$_{34}$	0
Industry 4	I.C.$_{41}$	0	0	I.C.+V.A.	0
Industry 5	I.C.$_{51}$	0	0	0	I.C.+V.A.

Figure B.1. A network matrix example

By analysing the matrix by row, we can identify the intensity of the interaction of each unit towards other destination nodes. This indicator is called the out-degree index and is calculated as the row sum, excluding the main diagonal entry. In examining the matrix by column, it is possible to compute the so-called in-degree index, which represents the ability to influence the origin of flows. Neither index offers details about where flows are coming from.

In order to consider the out-degree and in-degree indices simultaneously, Cetorelli and Peristiani (2009) suggested using the betweenness index, which exploits network ties and captures the uniqueness of a given node in a network. Let $m_{jk}(n_i)$ be the maximum flow between nodes (*nj ,nk*) which goes through node n_i. Aggregate across all possible pairs of nodes in the network, other than n_i, and obtain the overall betweenness of node n_i as $\sum_j \sum_k m_{jk}(n_i)$. In order to allow for comparison over time, normalisation is recommended, so that the betweenness index of node n_i is:

$$P(ni) = \sum_j \sum_k \frac{m_{jk}(n_i)}{m_{jk}} \qquad \text{(eq. B1)}$$

Therefore, the prestige index of node n_i is:

$$Pr\ (n_i) = x_{1i}\, P(n_1) + x_{2i}\, P(n_2) + \ldots + x_{Ni}\, P(n_N) \qquad \text{(eq.B2)}$$

where the weights are represented by the flows from each of the nodes onto n_i . We have N equations in N unknowns for each network.

This sophisticated and standardised index allows for the judgement of the importance of each node in a network, fully exploiting the information contained in the entire network structure.

This metric allows us to normalise the data from symmetric input-output tables and identify an international ranking for hubs of expertise. This index is a proxy for the knowledge level of every industry in each country. The greater values in this index are associated with the greater influence of the sector in the production of goods and services for the whole economy.

In basic terms, we dispose the flows of intermediate consumption between every pair of sectors on the off-diagonal entries, while the main diagonal includes the sum of the intermediate consumption flows within every sector with the sector value added. This matrix represents all of the data on a country's production. We divide every column of the matrix by the column sum and apply the power method in order to calculate the eigenvector associated with the largest eigen value of the matrix. This eigenvector contains the index of prestige of any sector. Lastly, we identify the main sector of each firm and assign to each analyst covering that firm the index of prestige associated with that analyst's country location.

APPENDIX C

This appendix illustrates and summarises the application of the theoretical framework through an illustrative example: calculate the variables *CLUSTER*, *OUTPUT* and *NET* for the UK, assuming that this economy has just six sectors.

<u>The cluster-based approach:</u>

We start from the following input-output table[18] where the off-diagonal elements are the intermediate consumption between two industries. The main diagonal elements are the sum of the intermediate consumption and the value added of a sector:

Sectors	1	2	3	4	5	6
1	33000	200	57	4500	7000	5000
2	200	500	3000	20000	30000	100
3	57	3000	10000	300	1000	1500
4	4500	20000	300	2000	2500	3000
5	7000	30000	1000	2500	20000	6000
6	5000	100	1500	3000	6000	45000

Table C.1. Input-output tables among UK industries.

We calculate an I-O matrix of only the intermediate consumption between a country's different industries. The off-diagonal elements are expressed as a percentage of the largest intermediate consumption between two industries, the benchmark for which has been set as equal to 100% (in this case, 45,000 is set as 100%). The main diagonal elements, which represent the intermediate consumption of the same industry, are zeros. We also set a minimum threshold for input and output entries to be part of the matrix at 2%. Therefore we should delete the elements highlighted in yellow.

Sectors	1	2	3	4	5	6
1	0,00	0,44	0,13	10,00	15,56	11,11
2	0,44	0,00	6,67	44,44	66,67	0,22
3	0,13	6,67	0,00	0,67	2,22	3,33
4	10,00	44,44	0,67	0,00	5,56	6,67
5	15,56	66,67	2,22	5,56	0,00	13,33
6	11,11	0,22	3,33	6,67	13,33	0,00

Table C.2. Input-output tables of intermediate consumption among industries express as percentages.

[18] These numbers are hypothetical. This is just an exemplification.

Hence, the matrix becomes:

Sectors	1	2	3	4	5	6
1	0,00	0,00	0,00	10,00	15,56	11,11
2	0,00	0,00	6,67	44,44	66,67	0,22
3	0,00	6,67	0,00	0,00	2,22	3,33
4	10,00	44,44	0,00	0,00	5,56	6,67
5	15,56	66,67	2,22	5,56	0,00	13,33
6	11,11	0,22	3,33	6,67	13,33	0,00

Table C.3. Input-output tables of intermediate consumption among industries express as percentages and values greater than the threshold.

That can be expressed in absolute values as:

Sectors	1	2	3	4	5	6
1	0	0	0	4500	7000	5000
2	0	0	3000	20000	30000	100
3	0	3000	0	0	1000	1500
4	4500	20000	0	0	2500	3000
5	7000	30000	1000	2500	0	6000
6	5000	100	1500	3000	6000	0

Table C.4. Input-output tables of intermediate consumption among industries express in absolute values and values greater than the threshold.

Putting the matrix in the block diagonal form by rearranging sectors according to Hoen (2002)'s algorithm, the matrix shows which sectors belong to which clusters.

Let us assume that the *diagonalisation* technique applied to this example,[19] result in the identification of the two clusters coloured in the diagonal blocks below. The elements of the matrix are the intermediate consumptions.

Sectors	1	6	4	3	2	5
1	0	5000	4500	0	0	0
6	0	0	3000	0	0	0
4	0	0	0	0	0	0
3	0	0	0	0	3000	
2	0	0	0	0	0	30000
5	0	0	0	0	0	0

Table C.5. Input-output tables of intermediate consumption among industries rearranged by clusters.

[19] The diagonalisation procedure implemented by Hoen is reported step by step in the Appendix A. In this example, we are not following the indicated steps, because it is difficult to make it effective in this simplified example. Therefore, we are assuming its implementation and the results indicated in Table 5.

We can now calculate the *CLUSTER* value for each sector composing the cluster.

Clusters	Sector	Sum of Sij	CLUSTER=LN(Sum of Sij)
Cluster 1	1	12500	9.434
	6	12500	9.434
	4	12500	9.434
Cluster 2	3	33000	10.404
	2	33000	10.404
	5	33000	10.404

Table C.6. Input-output tables of intermediate consumption among industries rearranged by clusters in UK.

Cluster 1 is composed of sectors 1, 4 and 6, while cluster 2 contains sectors 2, 3 and 5. We associate the sum of the intermediate consumption of the corresponding cluster to each sector and by applying the natural log transformation, we obtain the variable *CLUSTER*.

The second step is to attribute the *CLUSTER* values to analysts. Following our hypothesis, analysts located in the UK who evaluate companies based either in the UK or in another country, and belonging to one of the sectors 1, 4 or 6, will have a stock of knowledge of about 9.43. The same analysts who evaluate companies belonging to sectors of cluster 2 (sectors 2, 3 or 5) will have a higher stock of knowledge of about 10.40.

Let us assume that another country, for example Italy, could have the same sectors agglomerated in a different way. The stock of knowledge (*CLUSTER*) values would be different. Let us suppose, for instance, that Italy has a cluster value for sector 1, 2 and 3 equal to 5 and for sectors 4, 5 and 6 equal to 12, as summarized in the following table:

Clusters	Sector	CLUSTER=LN(Sum of Sij)
Cluster 1	1	5
	2	5
	3	5
Cluster 2	4	12
	5	12
	6	12

Table C.7. Input-output tables of intermediate consumption among industries rearranged by clusters in Italy.

According to our framework, an Italian analyst evaluating a company belonging to sector 4, 5 or 6 would perform better than the UK analyst because he or she has a bigger stock of knowledge produced by the agglomeration of these sectors (12 vs 10.404), while the UK analyst will issue better forecasts for companies in sectors 1, 2 or 3 (9.434 vs 5).

The sector- based approach:

In order to apply the sector-based approach and calculate the variable *OUTPUT*, we start with a Use table (see Appendix A for a detailed definition)[20] using the same numbers used in the previous approach:

Sectors	1	2	3	4	5	6
1	33000	200	57	4500	7000	5000
2	200	500	3000	20000	30000	100
3	57	3000	10000	300	1000	1500
4	4500	20000	300	2000	2500	3000
5	7000	30000	1000	2500	20000	6000
6	5000	100	1500	3000	6000	45000
Intermediate consumption	33000	500	10000	2000	20000	45000
V.A.	100	150	300	50	20	500
Total output at basic price	**33100**	**650**	**10300**	**2050**	**20020**	**45500**

Table C.8. Use table of UK industries.

We then scale the total output values for the country power purchase parity (PPP) in order to compare the same variables across different countries.

If UK PPP is equal to 0.98, applying the formula of *OUTPUT*, our variable assumes the following values for each of the six sectors.

Sectors	1	2	3	4	5	6
OUTPUT	10,42749127	6,49717507	9,260101882	7,645797779	9,92468976	10,74567031

Table C.9. Output values for UK industries.

Therefore, we have to associate the *OUTPUT* variable for each analyst in the dataset.

According to our framework and to these numbers, an analyst located in UK evaluating a firm from sector 1 will have a bigger stock of knowledge (10.42) than a colleague evaluating firms belonging to sector 2 (a stock of knowledge equal to 6.49), regardless to the company's location.

At the same time, if France, for instance, has different *OUTPUT* values, all else being equal, analysts located in that country will have a different informational advantage in evaluating the same companies. It depends on the stock of knowledge produced by France in relation to the six sectors.

[20] A Use table is a product-by-industry-based table with products and components of value added in the rows and industries, categories of final use and imports in the columns

The network-based approach:

Finally, in order to implement this procedure, we assume that the production value between sectors and within the same sector can be a proxy of the links of the network. Specifically, the production of each country is represented by a matrix which exhibits the flow of intermediate consumption between each pair of sectors on the off-diagonal entries, while the main diagonal shows the sum of the intermediate consumption flow within each sector and the sector value added.

Therefore, if the matrix we are looking for is the following one:

Sectors	1	2	3	4	5	6
1	33100	200	57	4500	7000	5000
2	200	650	3000	20000	30000	100
3	57	3000	10300	300	1000	1500
4	4500	20000	300	2050	2500	3000
5	7000	30000	1000	2500	20020	6000
6	5000	100	1500	3000	6000	45500

Table C.10. Table of the UK production.

We apply the algorithm (5) proposed in the literature by Cetorelli and Peristiani (2009) to the whole network.

As indicated above, we divide each column of the matrix by the column's sum.

Sectors	1	2	3	4	5	6
Sum per column	49857	53950	16157	32350	66520	61100

Table C.11. Sum of the UK production values by column (sector).

And we obtain the standardised matrix (SM):

Sectors	1	2	3	4	5	6
1	0,66	0,00	0,00	0,14	0,11	0,08
2	0,00	0,01	0,19	0,62	0,45	0,00
3	0,00	0,06	0,64	0,01	0,02	0,02
4	0,09	0,37	0,02	0,06	0,04	0,05
5	0,14	0,56	0,06	0,08	0,30	0,10
6	0,10	0,00	0,09	0,09	0,09	0,74

Table 12. The standardized matrix (SM) of UK production values.

After this standardisation, the system of equations becomes a more common matrix-characteristic equation, where the solution (that is, the vector of prestige indicators) is the eigenvector associated with the largest eigenvalue of the standardised matrix.

Since we do not use any specific mathematical software to calculate the eigenvalue of the matrix, we apply the ‘power method’.

Sectors	1	2	3	4	5	6
1	0,48	0,11	0,02	0,12	0,12	0,13
2	0,12	0,49	0,16	0,08	0,17	0,08
3	0,01	0,05	0,42	0,04	0,04	0,04
4	0,08	0,05	0,09	0,25	0,20	0,05
5	0,15	0,21	0,17	0,40	0,37	0,12
6	0,16	0,09	0,14	0,10	0,11	0,58
Sectors	1	2	3	4	5	6
1	0,29	0,15	0,09	0,16	0,16	0,17
2	0,16	0,31	0,20	0,16	0,19	0,13
3	0,03	0,06	0,20	0,06	0,05	0,05
4	0,10	0,10	0,11	0,16	0,15	0,08
5	0,21	0,23	0,22	0,30	0,29	0,18
6	0,21	0,15	0,18	0,16	0,16	0,39

Table 13. First two iterations to calculate the eigenvalue of the SM.

And after a number of iterations, the matrix converges to the following:

Sectors	1	2	3	4	5	6
1	0,011	0,011	0,011	0,011	0,011	0,011
2	0,004	0,004	0,004	0,004	0,004	0,004
3	0,001	0,001	0,001	0,001	0,001	0,001
4	0,001	0,001	0,001	0,001	0,001	0,001
5	0,020	0,020	0,020	0,020	0,020	0,020
6	0,000	0,000	0,000	0,000	0,000	0,000

Table 14. The eigenvectors of the SM.

Where all the equal columns are the eigenvectors of the SM.

NET for the UK is therefore equal to the following vector of values which represents the prestige index of each sector in that country:

Sector	**NET**
1	0,011
2	0,004
3	0,001
4	0,001
5	0,020
6	0,000

Table 15. NET values for UK sectors.

According to this approach, a UK analyst evaluating companies belonging to sector 1 has a stock of knowledge equal to 0.011, whereas whilst evaluating companies in sector 6 he or she has an informational advantage equal to zero. Each country (network) has a proper *NET* vector and, therefore, analysts located in different countries have different informational advantages deriving from the network to which they are closest.

References

[1] ANG, J. S., CICCONE S., (2001) "International differences in analyst forecast properties", Working paper, University of New Hampshire.

[2] BACMANN, J.F, BOLLIGER G., (2001) Who are the best? Local versus foreign analysts on the Latin American stock markets, Working paper, University of Neuchâtel, Neuchâtel, Switzerland.

[3] BAE K., STULZ R., 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*:*581–606.

[4] BARKER R. (1998) "The market for information: Evidence from Finance Director, Analysts and Fund Managers", *Accounting and Business Research*, 29, 3-20.

[5] BOLLIGER, G, (2004) "The characteristics of individual analysts' forecasts in Europe", *Journal of Banking and Finance* 28, 2283-2309

[6] BRETON G., TAFLER R. J. (2001): "Accounting information and analyst stock recommendation decisions: a content analysis approach", in *Accounting and Business Research*, 31 No 2, pp. 91-102.

[7] BROWN, L. (1993) "Earnings forecasting research: its implications for capital markets research", *International Journal of Forecasting*, 9, 295−320.

[8] BROWN P., G. FOSTER G., E. NOREEN (1985) "Security analyst multiyear earnings forecasts and the capital market", *Studies in Accounting Research*, 21, American Accounting Association.

[9] CETORELLI N., PERISTIANI S., (2009), Prestigious Stock Exchanges: A Network Analysis of International Financial Centers, *Staff Report*, Federal Reserve Bank of New York.

[10] CHANG, C. (2003) “Information footholds: expatriate analysts in an emerging market”, *Working Paper*, Haas School of Business, U. C. Berkeley .

[11] CHANG, J. J., KHANNA T., PALEPU K, (2000) “Analyst activity around the world”, *Working Paper*, Harvard Business School, Strategy Working Paper Series.

[12] CHOE H., KHO B.C., STULZ R.M. (2005) “Do domestic investors have an edge? The trading experience of foreign investors in Korea”, *Review of Financial Studies*, 18, pp. 795-829.

[13] CLEMENT M. B. (1998) “Some considerations in measuring analysts’ forecasting performance”, *Working Paper*, University of Texas, Austin.

[14] CLEMENT M. B. (1999): “Analyst forecast accuracy: do ability, resources, and portfolio complexity matter?”, *Journal of Accounting and Economics*, 27, pp. 285-303.

[15] COVAL, J. D., MOSKOWITZ T. J. (1999) Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045–2073.

[16] COVAL, J. D., MOSKOWITZ J. T. (2001) “The geography of investment: Informed trading and asset prices”, *Journal of Political Economy* 4, 811–841.

[17] DAVIS S., HENREKSON M. (2000) “Wage-setting Institutions as Industrial Policy”, SWOPEC, *Working Paper* n. 352.

[18] DIETZENBACHER, E., (1996), An Algorithm for Finding Block-Triangular Forms, *Applied Mathematics and Computation*, 76: 161-171.

[19] DOERINGER P., TERKLA D. (1995) “Business Strategy and cross-country clusters”, *Economic Development Quarterly*, 9: 225-237.

[20] DVORAK T. (2005) “Do domestic investors have an information advantage: evidence form Indonesia”, *Journal of Finance*, 60: 1461-1493

[21] FORNI, M., PABA S., (2001), *Knowledge Spillovers and the Growth of Local Industries*, CEPR Discussion Paper, no. 2934.

[22] FROOT, K., P. O’CONNELL, M. SEASHOLES (2001) “The Portfolio flows of international investors”, *Journal of Financial Economics* 59, 151-193.

[23] FROOT, K., T RAMADORAI (2001) “The information content of international portfolio flows”, NBER, *Working paper,* no. 8472.

[24] GEHRIG, T., (1993), “An information based explanation of the domestic bias in international equity investment”, *Scandinavian Journal of Economics* 95, 97-109.

[25] GRINBLATT, M., M. KELOHARJU, (2000) “The investment behaviour and performance of various investor types: A study of Finland’s unique data set”, *Journal of Financial Economics* 55, 43-67.

[26] HAU H., (2001) “Location matters: An examination of trading profits”, *Journal of Finance* 56, 1959-1983.

[27] HOEN A.R., (2002), “Identifying linkages with cluster based methodology”, *Economic System Research*, 14 (2): 131-146

[28] HONG, H, J. D. KUBIK, J. C. STEIN (2005) “Thy neighbour’s portfolio: word-of-mouth effects in the holdings and trades of money managers”, *Journal of Finance*, 60 (6).

[29] HUBERMAN, G. (2001), “Familiarity breeds investment”, *Review of Financial Studies* 14, 659–680.

[30] JACOB J., LYS T., NEALE M. (1999): “Expertise in forecasting performance of security analysts”, in *Journal of Accounting and Economics*, 28, pp. 51-82.

[31] JACOBS D., DEMANN A. (1996) “Cluster, Industrial Policy and Firm Strategy: a menu approach”, *Technology Analysis & Strategic Management*, 8 (4), pp. 425-437.

[32] KANG, J.., STULZ, R.M. (1997) “Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan”, *Journal of Financial Economics* 46, 3 – 28.

[33] KRUGMAN P., (1991), *Geography and Trade*, MIT Press, Cambridge

[34] LAWSON, C., LORENZ, C. (1999), “Collective learning, tacit knowledge and regional innovation capacity”, *Regional Studies,* 33: 305-317

[35] LORENZ, M., MASKELL, P (2004), “The cluster as a nexus of knowledge creation”, In P. Cooke &A. Piccaluga (Eds.), *Regional Economies as Knowledge Laboratories:* 77-92. London: Edward Elgar.

[36] MALLOY, C. J. (2005) “The Geography of equity analysis”, Journal of Finance 60, 719-755.

[37] MALBERG, A., & MASKELL, P. (2002). The elusive concept of localization economies: Towards a knowledge based theory of special clustering. *Environment and Planning A*, 34: 429-449

[38] MARSHALL, A. (1920). *Principles of economics*, 8th edn. London: McMillan

[39] ORPURT, S., (2004), “Local analyst earnings forecast advantages in Europe”, *Working Paper*, University of Chicago.

[40] PORTER M. (1990) *The Competitive Advantage of Nations*, London, MacMillan Press.

[41] PREVITS. G., BRICKER R., ROBINSON T. and YOUNG S. (1994), “ A content analysis of sell-side financial analyst company reports”, *Accounting Horizons*, 8 (2): 55-70.

[42] RICARDO, D. (1963), *The Principles of Political Economy and Taxation*, Irwin, Homewood, IL.

[43] ROSENFELD S. (1997) “Bringing Business Clusters into the Mainstream of Economic Development”, *European Planning Studies*, 5, n.1.

[44] ROSENFELD S. (1996) “Industrial Strength Strategies: Regional Business Clusters and Public Policy”, Aspen Institute Rural Economic Policy Program, *Best Practice Series*, Industrial Strength Strategies.

[45] SEASHOLES, M. (2000), “Smart foreign traders in emerging markets”, *Working Paper*, Harvard University.

[46] SHUKLA, RAVI AND GREGORY VAN INWEGEN, (1995) “Do Locals Perform Better Than Foreigners? An Analysis of U.K and U.S. Mutual Fund Managers”, *Journal of Economics and Business,* 47, 241-254.

[47] STORPER M. (1997), The Regional World, Guildford Press, New York