Does Correlated Analyst Coverage Explain Excess Comovement?

Ryan D. Israelsen*
Kelley School of Business
Indiana University.

First Draft: March 2009
This Draft: November 1, 2010†

Abstract

Many studies have documented stock return comovement above and beyond that predicted by standard asset pricing models. For example, when stocks are added to an index, their betas with respect to that index tend to increase. I find that much of this excess comovement can be explained by correlated information. If individual analysts’ earnings forecast errors are correlated across stocks, stock return correlations should be higher than fundamental correlations. I develop a measure of correlated analyst coverage to test this hypothesis and find: (1) Stocks with similar analysts tend to exhibit more excess comovement, controlling for industry, (2) This effect is strongest among “star analysts”, (3) On average, when a stock enters the S&P 500 index, the same analysts that cover other S&P 500 stocks begin to cover the new stock, (4) Changes in excess comovement are larger for stocks with larger increases in correlated analyst coverage around this event, and (5) Changes in the measure around analyst turnover following brokerage firm mergers are correlated with changes in excess comovement. These effects are robust after controlling for industry, for correlations in mutual fund and institutional holdings, and for correlations in unexpected earnings and earnings revisions. Similar but weaker results exist using analysts from the same firm or analysts who previously worked for the same firm.

* Finance Department, Kelley School of Business, Indiana University, 1309 E Tenth St, Bloomington, IN 47405-1701 Tel: (812) 855-8435, fax: (812) 855-5875, and e-mail: risraels@indiana.edu.
† Thanks to Utpal Bhattacharya, Dennis Capozza, Jonathan Cohn, Andrew Ellul, Eitan Goldman, Raffi Indjejikian, Francisco Palomino, Paolo Pasquariello, Uday Rajan, Tyler Shumway, Irina Stefanescu, and Lu Zhang and participants in seminars at Indiana University, at the University of Michigan, and at the Financial Management Association meetings in New York City. I thank Hong Kee Sul and Jordan Winkler for help gathering star analyst data. This paper is adapted from Chapter 2 of my dissertation.
1 Introduction

Return comovement is a fundamental part of modern asset pricing theory. Without exposure to systematic risk, all assets should earn the risk free rate of return. However, many recent studies have documented the existence of “excess comovement” — or comovement above and beyond that which is predicted by standard models. These papers typically fall into one of two categories: (a) those which identify groups of assets whose returns are correlated after controlling for risk (e.g., Karolyi and Stulz, 1996, and Pasquariello and Kallberg, 2008) and (b) those which find increases in comovement after an event takes place which is not necessarily associated with a change in fundamentals, such as the addition of a stock to an index (e.g., Barberis, et al., 2005, and Boyer, 2009). In this paper, I use both approaches to show that correlated information is another important source of comovement and can be used in the baseline defining “excess”.

If news about the fundamentals of two stocks is both informative and correlated, the returns of the two stocks will tend to be correlated. Because there is scant data on individual investors’ information sets, information-driven comovement has received relatively less attention than alternate sources of comovement. Instead of directly focusing on those consuming information (the investors), I focus my attention on those producing information (the sell-side analysts). Analyst forecast data are readily available for a wide range of stocks. Based on a model which ties earnings and earnings forecasts to stock return correlations, I develop an easy-to-calculate, intuitive measure of correlated analyst coverage,

\[ N_{ij} / \sqrt{N_i N_j}, \]  

where \( N_i \) and \( N_j \), are the number of analysts covering stocks \( i \), and \( j \), respectively, and \( N_{ij} \) is the number of common analysts between the two stocks. I show this measure to be a theoretical proxy for earnings forecast error correlations, and to have practical power in explaining excess comovement. In particular, as correlated analyst coverage increases, comovement increases — both in theory and in practice.

Using method (a) — controlling for risk based on standard asset pricing models — I show that increases in correlated analyst coverage are associated with increases in excess comovement in SP500 stocks from 1995–2007. A 1-percent increase in the correlated information measure is associated with an increase in excess comovement of about 0.4 percent. Analysts tend to cover stocks within the same industry. To the extent that industry is a proxy for risk exposure, the correlated analyst
coverage variable could be proxying for risk not captured by the asset pricing models. To rule out this alternative hypothesis, I perform several tests. First, I include dummy variables which control for GICS industry definitions and find that the effect is still there. These results are robust to the inclusion of correlations in mutual fund ownership and institutional investor holdings, suggesting that the analyst coverage variable is not simply a proxy for the trading activity of style investors. Second, I find that the effect is strongest among the subset of coverage defined by “star analysts” – i.e., analysts whose forecasts are more likely to make an impact. This suggests that the effect is related to information and not to risk proxies. Third, I find that stocks that are covered by different analysts from the same firm tend to exhibit excess comovement. When making forecasts, they may be using the same model or the same inputs in their models. They may also share information with one another. To the extent that firms hire the same types of analysts, or that analyst self select, they may simply be interpreting information similarly. All of these reasons may lead these analysts’ forecast errors to be correlated. Fourth, I develop a measure of social connectedness that is orthogonal to the analyst and brokerage-firm-level measures. I find that pairs of stocks covered by analysts that worked for the same firm in the past tend to comove together. This measure seems to be capturing training, information sharing via social networks, and/or similarity based on screening or self selection. Finally, the correlated analyst coverage variable does not seem to be proxying for correlations in unexpected earnings, or in analyst earnings forecasts revisions — which have both been shown to be related to abnormal returns.

In a second test of the model, using method (b), I choose two events which result in changes in analyst coverage, but which are not necessarily associated with changes in the firms’ fundamentals. First, I examine the addition of stocks to the S&P500 index. I show that there is a fundamental change in the information structure associated with the newly added firms following the event. The number of analysts covering the firms increases on average, and, in particular, these new analysts tend to be those already covering other S&P500 stocks. With respect to the stocks already in the index, the average correlated analyst coverage measure increases significantly for the newly added firms — about 17% within the first quarter after the event. Furthermore, I find a positive cross-sectional relationship between the change in correlated analyst coverage and the change in excess comovement. Pairs of stocks whose sets of analysts become more similar tend to see increases in excess comovement. While the magnitude of the effect is not large enough to explain all of the change in comovement around these events, the effect itself is consistent with the hypothesis that
correlations in information are responsible for some level of excess comovement.

Second, I examine the turnover of analysts following 14 mergers of brokerage firms. These events have been used as a natural experiment for examining the effects of competition on the bias of analysts' forecasts (see Hong and Kacperczyk, 2010). Analyst coverage naturally changes following mergers as analysts are made redundant. Focusing on the subset of S&P500 stocks covered by the merging firms, I find a positive and significant relationship between the changes in correlated analyst coverage and in excess comovement which is not accounted for by changes in mutual fund holdings or in institutional ownership around these events. The coefficients are similar in magnitude to those associated with additions to the S&P500 index.

In both the standard neoclassical investment model and in standard consumption based asset pricing models, when there is perfect information about the current states of the economy, the correlation between any two firms' stocks is based solely on the fundamental risk characteristics of the firm, and expectations about risk factors. However, when noisy signals about future states of the world are introduced, expected returns, covariances, and correlations may be affected. In particular, when the errors in signals about future earnings are positively correlated, the expected conditional correlations of stock returns increase. Expected variances and covariances are functions of prior beliefs about the states and the noisy signals. Rational Bayesian updating leads to an increase in comovement when error correlations increase. Even though the errors are known to be correlated, the effects of this correlation cannot be fully erased when forming beliefs.

Analysts' forecasts are not perfect. Analysts often use models or standardized methodologies when making projections about earnings. Many of the inputs used in their models, such as projected GDP- or industry-growth may be used across multiple stocks, and among analysts within the same brokerage house. They may also simply interpret information in the same manner for multiple stocks. To the extent that each analyst uses the same methodology, model, or inputs across stocks we might expect any systematic errors to filter through to the earnings forecasts among the stocks for which they make projections. These correlations may be common knowledge to investors. However, even a fully rational investor using Bayesian updating will not be able to full reverse the correlations in forecast errors when updating beliefs about future cash flows.

Some level of excess comovement may be driven by the trading behavior of mutual funds, hedge funds, or other institutional investors. Insofar as analysts cater to these investors, the analyst coverage variable may be serving as a proxy for this behavior. Using mutual funds
and institutional holdings data, I find that correlated mutual fund and institutional holdings are important determinants of excess comovement. However, their effects are mostly orthogonal to that of correlated analyst coverage, which remains a significant factor in explaining comovement.

I find that the effect of correlated analyst coverage on excess comovement increases after the adoption of Regulation FD. Because analysts are privy to less private firm-specific information, more of the private information revealed by their forecasts is likely to be systematic or industry-wide information. If this information is common to the stocks the analyst is covering, any error in this information will also be common. This result is consistent with an information story.

I focus on the covariance structure of forecast error. However, the forecast itself is an estimate of the sum of idiosyncratic and systematic components. An unexpectedly high forecast could mean that either the systematic state is higher than was expected, the idiosyncratic state is higher, or both. So the signals reveal relevant information about systematic factors. Because this information will affect all stocks, this channel can also drive comovement. Veldkamp (2006) develops a model of excess comovement in a rational expectation equilibrium framework which has a similar aggregated information component. In her model, investors have information about a subset of assets. This generates comovement much the same way that the release of GDP or unemployment forecasts affects the prices of a wide range of assets. The covariance channel used in this paper provides a much stronger, direct effect than the aggregated information nature of earnings forecasts.

Chan, Jegadeesh and Lakonishok (1996), and Chordia and Shivakumar (2006) have documented a positive relation between standardized unexpected earnings and abnormal returns. To test whether the correlated analyst measure is capturing this effect, I include both in a model of excess comovement. Correlations in unexpected earnings are a significant source of excess comovement, but do not trump correlated analyst coverage as an important factor.

Finally, I examine whether the analyst variable is simply capturing the effect of correlations in analyst earnings forecast revisions. Da and Warachka (2009) show that a beta calculated using analyst forecast error revisions can explain much of the value premium, the size premium, and the long-term reversal puzzle. This variable is meant to capture the sensitivity of a stock to changes in market-wide revisions in expected cash flow. If stocks that are covered by similar analysts are also sensitive to systematic earnings news innovations, both measures might be capturing the same thing. However, while I do find that correlated earnings revisions does have explanatory power in the cross section of excess comovement, the importance of the analyst variable is not affected by
its inclusion in the regression.

The paper proceeds as follows. In Section 2, I develop the hypotheses to be tested and explain the intuition. In Section 3, I develop the correlated analyst forecast measure which is a proxy for information correlation and describe the data and the other measured used in the paper. In Section 4, I test the hypotheses that correlated information is associated with excess comovement using a variety of approaches. In Section 5, I conclude.

2 Hypotheses

To test whether the structure of the information production environment has a significant effect on the comovement of asset prices, I develop and test five closely related hypotheses. Under rational Bayesian updating, correlations across forecast errors lead to correlations in expectations. This is true even if there is no correlation in the underlying earnings. This is illustrated with a simple model in the appendix, however the intuition is straightforward. If there is a positive correlation between the consensus earnings forecasts for stocks A and B, both forecasts will likely be too high or too low, but a rational Bayesian agent will not know which is true. As long as the signals are informative about the means, both will be used to update expectations. The agent cannot fully undo the correlation in the signals, and as a result, the updated expectations will be subject to some level of correlation. For example, I may know that analysts are likely too optimistic or too pessimistic for both A and B, but because there is some level of precision in their forecasts, I am willing to incorporate them (along with the associated correlation) into my updated expectation of earnings. This correlation in expected earnings for the two firms will affect stock prices and returns:

**Hypothesis 1** Pairs of stocks with higher correlated analyst coverage, \( \rho_{ij} \), will exhibit more excess comovement.

In the appendix, I develop and calibrate a simple neoclassical model with earnings forecasts. According to the calibration, the upper bound on the marginal effect of increasing correlated analyst coverage should be around 0.40.

The obvious alternative to the correlated information hypothesis is the idea that analysts are better at picking similar stocks than are our asset pricing models. If this is the case, then the correlated analyst coverage variable will be a measure of how similar stocks are in terms of exposure to systematic risk. If our models fail to capture these risk exposures or if we do not measure
risk correctly, then we will find this association between correlated analyst coverage and “excess” comovement. To rule out this alternative, I control for GICS industry and develop several more hypotheses.

If the effect is driven by correlated information as opposed to unmeasured risk, then it ought to be the strongest where forecasts are given the most weight by investors. A natural measure of information precision is whether the analyst making the forecasts is considered a “star analyst” as measured by Institutional Investor magazine. If the effect is driven by unmeasured risk, then there should be no difference across the two types of analysts. If it is driven by information, then it should be stronger among the subset of star analysts:

**Hypothesis 2** The relationship between correlated analyst coverage and excess comovement will be strongest within star analysts.

Individual analysts tend to cover a small percentage of the entire universe of stocks. Any given brokerage firm, however, will employ analysts covering a much larger percentage of the S&P500 firms. To the extent that analysts working for the same firm use similar methodology, information, or process information in the same manner, we might expect their forecast errors to be correlated. The correlated broker coverage variable, $\rho_{ij}^{br}$, may provide additional power in explaining excess comovement. Additionally, because of the wider coverage at the firm level, mismeasured industry related risk factors are less likely to be an issue:

**Hypothesis 3** Pairs of stocks with higher correlated brokerage coverage, $\rho_{ij}^{br}$, will exhibit more excess comovement.

Social networks have been used to explain aspects of CEO investment behavior (Fracassi, 2009), to examine the role of boards in choosing and monitoring executives (Adams and Ferreira, 2007 and Subrahmanyam, 2008), and to investigate the ability of connected analysts to gather superior information (Frazzini et al., 2009). To the extent that information and advice is transferred via social networks, social connections may affect analyst forecasts, and ultimately return comovement. To test this, I create a connected analyst variable:

$$\rho_{ij}^{ca} \equiv N_{ij}^{ca} / \sqrt{N_i N_j}, \quad (2)$$

where $N_i$ and $N_j$ are the number of analysts covering stocks $i$, and $j$, respectively, and $N_{ij}^{ca}$ is the number of connected pairs of analysts. A pair of analysts is connected if the two analysts currently work for different firms, but each worked for the same firm in the past.
This connected analyst measure may capture a number of effects, all of which should lead to correlated earnings forecasts. First, analysts who worked with each other at the same firm may have remained in contact with each other and may be sharing information. Second, analysts at the same firm may have learned similar valuation techniques or information gathering and processing techniques which they both may have retained to some degree. Third, to the extent that individual firms choose to hire specific types of analysts, employment history is a proxy for analyst “type”. If similar “types” of analysts interpret information in the same way, this should filter through to the forecasts. Finally, along the same lines, there may be some self selection on the part of analysts when choosing employers – similar types may be attracted to the same firm. Because this measure excludes analysts who currently work from the same firm, it does not overlap with the correlated analyst or correlated broker measures.

**Hypothesis 4** Pairs of stocks with higher connected analyst coverage, $\rho_{ij}^{co}$, will exhibit more excess comovement.

The most ideal way to reject the alternative hypothesis of unmeasured risk would be to exogenously change the structure of analyst coverage without changing the fundamental risk in the underlying stocks and examine the corresponding changes in excess comovement. If the structure of information production affects comovement, there should be predictable changes in excess comovement:

**Hypothesis 5** Excess comovement will change with exogenous changes in analyst coverage.

To test this final hypothesis, I examine two separate types of events that are associated with changes in analyst coverage, but not necessarily changes in fundamentals: The addition of a stock to the S&P500 index and the merger of brokerage houses and the resulting turnover in analysts and analyst coverage. Both of these events have been used in the literature as natural experiments in explaining the behavior of traders or analysts.

### 3 Data Sources and Variable Construction

I use the unique analyst and broker id numbers from the I/B/E/S (Institutional Brokers’ Estimate System) database to determine which analysts are making earnings per share forecasts. I gather stock return and (six-digit) GICS industry codes data from the CRSP (Center for Research in Security Prices) database. Data for the 3 Fama French factors and the Carhart momentum factor
along with data on additions to the S&P500 index are from Wharton Research Data Services. Mutual funds holdings data are from the Thomson-Reuters Mutual Fund Holdings database and institutional investor holdings are from the Thomson-Reuters Institutional Holdings (13F) Database. Star analyst data are from the October issue of Institutional Investor magazine. Unless otherwise noted, the sample covers the period from 1995 through 2007. The beginning of the sample is limited by the availability of GICS industry codes. GICS codes from 1995 to 1998 are backfilled. These industry definitions have been shown to be the most relevant in explaining industry classifications by analysts. The use of SIC codes and a longer sample do not qualitatively change results.

3.1 Correlated Analyst Coverage and Forecast Errors Measure

I am interested in the covariance structure of the earnings forecast errors. It is possible to estimate this matrix directly using realized earnings forecast errors. However, given the staggered nature of both earnings announcement dates and analyst forecast dates, this is a difficult task in practice. Instead of estimating the covariance matrix directly, I take an indirect approach using observable analyst coverage data. Analysts often use models or standardized methodologies when making projections about earnings. Many of the inputs used in their models, such as projected GDP- or industry-growth may be used across multiple stocks, and among analysts within the same firm. Additionally, they may also interpret information in the same manner for multiple stocks. To the extent that each analyst uses the same methodology, model, or inputs, across stocks we might expect any systematic errors to filter through to the earnings forecasts across stocks for which they make projections.

With this in mind, assume for simplicity that the signal, or consensus, of the \( i \)th firm is the simple average of the forecasts of \( N_i \) analysts. Also, assume that forecast errors across analysts are uncorrelated, but that each individual analyst’s forecast errors are perfectly correlated across stocks and have a variance of \( \sigma_a^2 \) for any stock.\(^1\) Then, the variance of the consensus forecast error, \( e_i \), for the \( i \)th stock is equal to \( \sigma_a^2/N_i \). Let \( N_{i,j} \) be the number of shared analysts between the \( i \)th

\(^1\)The assumption that individual analyst forecast errors are perfectly correlated is made for simplicity. Allowing an imperfect correlation across stocks will change the definition of \( \rho_{ij}^a \), but the comparative statics used in developing the hypotheses and tests will not change. For example, if we assume that forecast errors at the individual analyst level across stocks have a correlation of \( \rho_{ij}^a > 0 \), then it is easy to show that \( \rho_{ij}^a = \alpha \frac{N_{i,j}}{\sqrt{N_i N_j}} \). Changing \( \alpha \) will change the magnitude of the effect, but not the sign or the nature of the relationship.
and \( j \)th stocks. Then, the covariance of the forecast errors for stocks \( i \) and \( j \) is:

\[
Cov(e_i, e_j) = \begin{cases} 
0 & \text{if } N_{ij} = 0 \\
\frac{\sigma^2}{N_iN_j} & \text{if } N_{ij} = 1 \\
\frac{2\sigma^2}{N_iN_j} & \text{if } N_{ij} = 2 \\
\vdots & \vdots \\
\frac{\min(N_i, N_j)\sigma^2}{N_iN_j} & \text{if } N_{ij} = \min(N_i, N_j)
\end{cases} \tag{3}
\]

and thus the correlation of the forecast errors is:

\[
\rho_{ij}^{an} \equiv Corr(e_i, e_j) = \frac{N_{ij}}{\sqrt{N_iN_j}}. \tag{5}
\]

This coefficient which ranges between 0 and 1 is similar to measures which have been used to quantify the “similarity” of sets.\(^2\) Thus, this quantity is a practical measure of the similarity of the sets of analysts covering the two stocks and a theoretical measure of the correlation between the forecast errors for the two stocks. Using the I/B/E/S database, this can be easily calculated for any pair of stocks.

Each calendar year, for every pair of stocks listed on the S&P 500 index from 1995 through 2007, I calculate the correlated analyst coverage measure, \( \rho_{ij}^{an} \), by counting the number of unique analysts making earnings forecasts in the I/B/E/S database, \( N_i \) and \( N_j \), and the number of common analysts within these two groups, \( N_{ij} \). There are about 1 million stock-pair observations over the 13 year period. The first line of Panel A in Table 1 shows the summary statistics for \( \rho_{ij}^{an} \). On average, this variable is about 2.7%. This number is relatively low. However, the standard deviation is about 8.6%. Analysts tend to specialize in a particular industry. Pairs of firms within the same industry generally have relatively high coefficients.

To test the second hypothesis, I create two new variables, correlated star analyst coverage, \( \rho_{ij}^{st} \), and correlated non-star analyst coverage, \( \rho_{ij}^{nst} \). These variables are defined much the same as is the the correlated analyst coverage measure, except that the term in the numerator, \( N_{ij} \), is replaced with the number of common star analysts or the number of common non-star analysts, respectively. An analysts is considered to be a “star analyst” if he or she is ranked 1, 2, or 3 in one of the ”All-American Research” teams in the October issue of Institutional Investor magazine.

---

\(^2\)Two popular measures of similarity are Sørensen’s (1948) similarity coefficient, \( 2N_{ij}/(N_i + N_j) \) and the Jaccard (1901) similarity coefficient, \( N_{ij}/(N_i + N_j - N_{ij}) \). The measure in this paper differs from Sørensen’s measure only in the sense that a geometric average is used in lieu of an arithmetic average.
from the previous year. Because star analysts make up only a small subset of all analysts, it is not surprising that the average value of this variable is much smaller than the general measure – at 0.2%. The standard deviation is 1.6%. The non-star analyst variable is 2.5% on average and has a standard deviation of 7.5%. The average of the sum of the two is equal to the average of the basic analyst coverage variable. This is to be expected, however, because \( \rho_{ij}^{st} + \rho_{ij}^{-st} = \rho_{ij}^{an} \) by construction.

To the extent that analysts working for the same firm use the same methodology or share the same information, they will make similar mistakes in projecting a firm’s earnings. With this in mind, I repeat the previous exercise at the brokerage firm level and generate a correlated broker coverage measure, \( \rho_{ij}^{br} \). The second line of Panel A presents the summary statistics for this variable. Not surprisingly, the mean of this coefficient is much larger than the analyst coefficient at about 49%. The standard deviation is about 17%. While individual analysts focus on industries, brokerage firms take a more broad approach. The correlation coefficient between the two measures is 0.34.

As described in the previous section, I develop a measure of “social connectedness” that is orthogonal to the correlated analyst and correlated broker coverage variables. The connected analyst coverage variable, \( \rho_{ij}^{ca} \), is defined as \( \rho_{ij}^{ca} = N_{ij}^{ca} / \sqrt{N_i N_j} \), where \( N_{ij}^{ca} \) is the number of social connections between analysts covering stocks \( i \) and \( j \), and \( N_i \) and \( N_j \) are the number of analysts covering each stock. A pair of analysts is considered to be socially connected if both analysts worked at the same firm at some point in the past, but no longer work with each other at the same firm. The final row of Panel A in Table 1 shows that this variable has a mean of 32.7% and a standard deviation of 19%. Because there are potentially \( N_i \times N_j \) social connections, the variable is not constrained at 100%.

### 3.2 Comovement Measures

In this paper, I define comovement as the correlation of returns or alphas between two stocks. The pairwise methodology has been used in the literature to measure and explain excess comovement (e.g., Kallberg & Pasquariello, 2008). The alternative approach typically involves estimating betas relative to a market index or a subset of similar stocks. The analyst coverage measure in this study does not easily lend itself to the latter method. Because the analyst coverage variable is defined between a pair of stocks and not with respect to an index, the pairwise variable is a more natural measure when using panel data. Both measures involve scaled covariances of pairs of stocks – a beta is simply a weighted average of individual asset covariances divided by the variance of the
entire index or portfolio of assets.

Excess comovement, $\rho_{ij}^{ret}$, is defined as the correlation coefficient between two firms’ realized alphas using an asset pricing model:

$$\rho_{ij}^{ret} = \frac{\sum_{t=1}^{T} e_{it}e_{jt}}{\sqrt{\left(\sum_{t=1}^{T} e_{it}^2\right) \left(\sum_{t=1}^{T} e_{jt}^2\right)}}$$

(6)

where $e_{it}$ and $e_{jt}$ are the residuals from the asset pricing equation:

$$r_{it} - r_{ft} = \alpha_i + \beta_{i1}f_{it} + \beta_{i2}f_{1t} + \ldots + \beta_{iN}f_{Nt} + e_{it}.$$  

(7)

Every calendar year from 1995 through 2007, I find all stocks which are part of the S&P500 index during that entire year and the previous year. The factor loadings, $\beta_{i1}, \beta_{12}, \ldots \beta_{Ni}$, are estimated using data from the previous calendar year. Three models are used: the CAPM, the Fama French 3 factor model (FF3 henceforth), which includes the market return, the size factor and the growth factor, and a 4 factor model (FF4 henceforth) which adds the Carhart momentum factor to the Fama French factors. For comparison purposes, I also calculate $\rho_{ij}^{ret}$ using raw returns in lieu of residuals.

Panel B of Table 1 proved summary statistics for these estimates using daily returns data. There are an average of 390 stocks per year that are part of the index both the current and previous year and have been assigned GICS codes. That leads to 985,484 yearly correlations over the course of the sample, or about 76,000 per year. The average correlation using raw returns is about 0.239. Excess comovement among these S&P500 stocks based on the CAPM is about 0.07 on average, and about 0.097 and 0.118 using the FF3 and FF4 models, respectively. There is some negative excess comovement in some cases, and the maximum for each variable exceeds 96%.

Because daily returns may be subject to microstructure issues like nonsynchronous trading, I have also estimated comovement using weekly returns according to the same procedure as with the daily returns. Though not reported here, the results of the paper are qualitatively unchanged when using excess comovement estimated at weekly frequencies.

### 3.3 Other Measures

To control for other potential sources of excess comovement, I create four additional measures. Two are related to trading behavior of sophisticated investors and two are related to unexpected earnings and earnings revisions.
Every calendar year, I determine the holdings of mutual funds from the Thomson-Reuters Mutual Fund Holdings database. For every pair of stocks, I define correlated mutual fund holdings, $\rho_{ij}^{mf}$ as:

$$\rho_{ij}^{mf} \equiv \frac{N_{ij}^{mf}}{\sqrt{N_i^{mf}N_j^{mf}}}$$ (8)

where $N_i^{mf}$ is the number of mutual funds who own shares of stock $i$, and $N_{ij}^{mf}$ is the number of mutual funds that hold both stocks $i$ and $j$ at time $t$. If a pair of stocks is held by the same set of mutual funds, they are more likely to receive (near) simultaneous price pressure when these funds experience inflows or outflows of capital. Using the Thomson-Reuters Institutional Holdings (13F) Database, I define correlated institutional holdings, $\rho_{ij}^{in}$, in an analogous manner. Panel C of Table 1 shows that the average value of the correlated mutual fund and institutional holdings variables are 0.41 and 0.57, respectively. So, on average the S&P500 stocks included in my sample share about 41% of their mutual fund investors and 57% of their institutional investors. These numbers are as low as 7.6% and as high as 93% in some cases.

In order to determine whether my analyst coverage variable is simply capturing the effect of common exposure to earnings momentum or other common sources of news innovations, I create two measures which are based on measures used to explain these effects in the literature. The first, correlated unexpected earnings, $\rho_{ij}^{ue}$, is the sample correlation of standardized unexpected earnings between stocks $i$ and $j$. The second, correlated analyst earnings forecast revisions, $\rho_{ij}^{ef}$, is the sample correlation coefficient of monthly consensus analyst earnings forecast revisions for the annual earnings between stocks $i$ and $j$. These two measures are estimated using 5 years worth of either quarterly or monthly data and are described in more detail in Section 4.8.

4 Results

4.1 Analyst Coverage and Excess Comovement

To test the hypothesis that increases in correlated analyst coverage are associated with increases in comovement, I regress excess comovement on the correlated analyst coverage variable along with year fixed effects. The first columns of Panels A, B, C and D in Table 2 present the results of these regressions using each of the four methods to estimate excess comovement. Year fixed effects are included but not reported. In all four cases, the coefficients are positive and highly significant, ranging between 0.35 and 0.43. These are roughly the same as the theoretical value of
0.41 from the calibration exercise in the appendix. A one percent increase in correlated analyst coverage is associated with an increase in excess comovement of roughly 0.4 percent. The standard errors, which are reported in parentheses, show that the coefficients are all statistically significant at the 1% level. To adjust for the bias in standard errors resulting from the use of a panel data set, I follow the advice in Peterson (2009) and Thompson (2009) and cluster by year and firm.

In economic terms, a one-standard-deviation increase in correlated analyst coverage is associated with an increase in excess comovement of about 3 to 4 percent. The Adjusted R-Square ranges between 0.33 in the case of raw returns and 0.59 when the four-factor model is used in defining excess comovement.

**Industries**

An alternative hypothesis that is consistent with these results is that our asset pricing models are imperfect and analysts tend to cover firms with the same risk characteristics. If this is the case, analyst coverage may provide information about risk above and beyond that provided by our models. Individual analysts tend to focus on firms within the same industry. In order to test whether this is what’s being captured by my measure, I perform several tests. In the first, most direct test, I include an industry related dummy variable in the equation. I create the variable GICS, which is an indicator variable set to 1 if the two firms share the same six-digit Global Industry Classification Standard (GICS) code, and 0 otherwise. Boni and Womack (2006) show that the partitions generated using GICS classification industries are a good proxy for the way in which analysts define themselves by their coverage choices.

The second column of each panel presents the results when controlling for industry affiliation. Common GICS means an increase in excess comovement of between 0.10 and 0.14, so there does seem to be some level of industry-related comovement not captured by the models. However, in each case, the coefficient on the correlated analyst coverage variable remains positive and significant – though magnitudes are smaller, ranging between 0.22 and 0.27. These numbers are smaller than the theoretical value from the calibration exercise in the appendix. However, this may not be surprising. The assumption which was used in linking correlated analyst coverage and correlated earnings forecast errors provides an upper bound. If an individual analyst’s forecast errors across stocks are not perfectly correlated, this number should be lower. To the extent that these industries control for unmeasured risk, correlated analyst coverage increases comovement above and beyond

---

3 Clustering by year, firm-pair and firm does not significantly change standard errors.
that explained by common industry.

4.2 Star Analysts

If the excess comovement associated with analyst coverage is driven by the structure of the information production process, then it ought to be the most pronounced among the set of analysts whose forecasts are given more attention by investors. To test this second hypothesis, I regress excess comovement on the common star analyst measure along with the industry indicator variable and year fixed effects. Results are in the fourth column. The marginal effect of an increase in correlated star analyst coverage is 3 to 4 times as large as the general analyst coverage variable – ranging between 0.80 and 0.99. All variables are statistically significant. On the margin, increasing correlated star analyst coverage has a strong effect on excess comovement. A 1-percent increase in coverage is associated with a 0.8 to 1-percent increase in excess comovement.

In the final column of each panel, I add the non-star analyst variable to the regression. With the addition of this variable, the coefficient on the star analyst measure drops to about 0.46 – 0.60, which is about twice as large as the coefficient on the aggregate analyst measure. The coefficient on the non-star analyst variable is less than the coefficient on the general measure, but still statistically significant – between 0.19 and 0.22.

As a formal test of the difference of the effects of star analysts, and non-star analysts, I test the restriction that the coefficients of the two variables, $\beta_{st}$ and $\beta_{-st}$, are equal. The final three rows of each panel provide the differences in these two coefficients, the F-statistics for the restriction that the two are equal and the p-values. The differences in marginal effects range from 0.270 to 0.385. Moreover, using all four measures of excess comovement, the difference in effects between the two types of analysts are statistically significant with F-statistics greater than 400. Separating the general variable into the two subsets yields results consistent with the hypothesis that the excess comovement associated with common analyst coverage is driven by correlated information, and not simply by unmeasured risk.

4.3 Brokerage Firms

Individual analysts tend to cover a small percentage of the entire universe of stocks – typically an industry. Any given brokerage firm, however, may have analysts covering a much larger percentage of the S&P500 firms. To the extent that analysts working for the same firm use similar methodology, information, predictions, or process information in the same manner, we might expect their forecast
errors to be correlated. The correlated broker coverage variable, $\rho_{ij}^{br}$, may provide additional power in explaining excess comovement. Additionally, because of the wider coverage, mismeasured industry-related risk factors are less likely to be an issue.

To test this third hypothesis I repeat the regression in Table 2, replacing the analyst level variable with the brokerage level measure. These results are in the first and second columns of the four panels in Table 3. The coefficients on the brokerage coverage variable are all positive and significant, ranging from about 0.07 to 0.08 without accounting for GICS code, and from about 0.04 to 0.05 after controlling for industry. Not surprisingly, the effects of this indirect measure of correlated information are smaller than the direct effect associated with the analyst variable. However, because brokerage houses cover a much more broad set of stocks than do individual analysts, this is consistent with an information-driven channel driving some level of excess comovement, and not an unmeasured risk story.

4.4 Social Networks and Excess Comovement

Social networks have been used to explain aspects of CEO investment behavior (Fracassi, 2009), to examine the role of boards in choosing and monitoring executives (Adams and Ferreira, 2007 and Subrahmanyam, 2008), and to investigate the ability of connected analysts to gather superior information (Frazzini et al., 2009). To the extent that information and advice is transferred via social networks, social connections may affect analyst forecasts, and ultimately return comovement. To test this, I create a connected analyst variable:

$$\rho_{ij}^{ca} \equiv N_{ij}^{ca} / \sqrt{N_i N_j},$$

where $N_i$ and $N_j$ are the number of analysts covering stocks $i$ and $j$, respectively, and $N_{ij}^{ca}$ is the number of connected pairs of analysts. A pair of analysts is connected if the two analysts currently work for different firms, but each worked for the same firm at the same time in the past.

This connected analyst measure may capture a number of effects, all of which should lead to correlated earnings forecasts. First, analysts who worked with each other at the same firm may have remained in contact with each other and may be sharing information. Second, analysts at the same firm may have learned similar valuation techniques or information gathering and processing techniques which they both may have retained to some degree. Third, to the extent that individual firms choose to hire specific types of analysts, employment history is a proxy for analyst “type”. If
similar “types” of analysts interpret information in the same way, this should filter through to the forecasts. Finally, along the same lines, there may be some self selection on the part of analysts when choosing employers – similar types may be attracted to the same firm. Because this measure excludes analysts who currently work from the same firm, it does not overlap with the correlated analyst and correlated broker measures.

I repeat the exercise in the previous section, but now use the connected analyst measure, $\rho_{ij}^{ca}$, as an independent variable. Columns three and four of the panels in Table 3 present the results of these regressions. In every specification, the coefficient on $\rho_{ij}^{ca}$ is positive and significant, 1.7% to 2.6% and statistically significant. These coefficients are much smaller than those for both the analyst level measure and the broker level measure. However, given the degree of separation between the analysts behind this measure, this positive, significant result provides additional evidence that some degree of comovement is driven by the information structure.

### 4.5 Mutual Funds and Institutional Investors

Comovement in stock returns may be partially driven by price pressure from style investors or mutual fund managers tracking an index. Boyer and Zheng (2009) find a positive and significant correlation between mutual fund flows and returns. Coval and Stafford (2007) find that funds that experience large outflows or inflows tend to adjust their positions accordingly which may cause price pressure on the stocks in their portfolios. Barberis et al. (2005) attribute the increase in comovement in stocks that are added to the S&P 500 index to the trading behavior of style investors, index funds, or other institutional investors. To the extent that analysts cater to particular institutional investors, the positive coefficient on $\rho_{ij}^{mn}$ may simply be proxying for the trading behavior of institutional investors. To control for this possible source of comovement and to gauge its importance, I develop a measure that is similar in spirit to the correlated analyst coverage measure, but is based on stock ownership data.

Every calendar year, I determine the holdings of mutual funds from the Thomson-Reuters Mutual Fund Holdings database. For every pair of stocks, I define correlated mutual fund holdings, $\rho_{ij}^{mf}$ as:

$$\rho_{ij}^{mf} \equiv \frac{N_{ij}^{mf}}{\sqrt{N_i^{mf}N_j^{mf}}}$$

where $N_{ij}^{mf}$ is the number of mutual funds who own shares of stock $i$, and $N_i^{mf}$ is the number of
mutual funds that own shares in both stocks $i$ and $j$ at time $t$. If a pair of stocks is held by the same set of mutual funds, they are more likely to receive (near) simultaneous price pressure when these funds experience inflows or outflows of capital. Using the Thomson-Reuters Institutional Holdings (13F) Database, I define correlated institutional holdings, $\rho_{ij}^{m}$, in an analogous manner.

Table 4 presents the results of the regressions of excess comovement on combinations of correlated analyst coverage and correlated mutual fund holdings. Table 5 presents the same regressions for the institutional holdings variable. Common mutual fund holdings are associated with higher levels of comovement, suggesting that either (a) mutual funds tend to hold similar types of stocks, or (b) buying and selling pressure by these investors causes excess comovement in stock returns. The first column of each of the four panels of Table 4 shows coefficients for the regression of excess comovement on correlated mutual fund holdings and year fixed effects. The coefficient on $\rho_{ij}^{m}$ is about 0.26 using raw returns and around 0.45 using alphas from the asset pricing models. Adding the GICS industry dummy decreases the coefficients by about 0.05. In columns three and four, the analyst coverage variable is added to the regression without and with the GICS dummy. Adjusted R square ranges between 0.31 and 0.62. All variables are statistically significant. Accounting for common mutual fund holdings does not significantly reduce the coefficient on correlated analyst coverage which can be seen by comparing to the first two columns of Table 2.

The results for the institutional holdings variable are roughly the same as the those for mutual funds. The main difference is the smaller magnitudes of the coefficient on the institutional ownership variables - between 0.19 and 0.32. However, the mean of the correlated mutual fund measure is smaller (0.41 vs. 0.57), so the average effect is roughly the same. Adjusted R square is as high as 0.60. The coefficient on the analyst coverage variable drops only slightly after including correlated institutional holdings. Thus, trading pressure as measured by these variables does not seem to negate the ability of correlated analyst coverage in explaining some level of excess comovement.

4.6 Additions to the S&P500 Index

Barberis, Shleifer and Wurgler (2005) find that stocks’ betas with respect to the S&P500 index tend to increase when they added to that index. They argue that nothing fundamental changes with the stock in the short window over which they estimate these betas and ascribe the increased correlation to the trading behavior of “style investors” – investors who trade stocks based on characteristics or classifications. “Style investors” and institutional investors buying and selling stocks in order to track the S&P500 index are likely to cause an increase in a firm’s measured beta. However, there
may be another source of comovement. If the information structure changes, then we might expect the comovement to change.

Hegde and McDermott (2003) show that additions to the S&P500 index are associated with increases in analyst coverage. If these new analysts are also those covering other S&P500 stocks, the forecast errors may become more correlated. With this in mind, I calculate $\rho^{\text{com}}_{ij}$ each quarter from 12 quarters before to 12 quarters after the event. To avoid cases in which firms were simply not initially in the sample, I use data from the 386 firms from the period 1982 through 2007 which had at least one analyst in the I/B/E/S database before the event. Table 6 shows the average values before and after the event, plus the percentage increases, differences, t-statistics and p-values. On average, these measures of correlated information increase when a stock is added to the index. There is a 17% increase in $\rho^{\text{com}}_{ij}$ for these new S&P500 stocks from the quarter before to the quarter after the addition. From 12 quarters before to 12 quarters after, there is a 140% increase. Figure 1 plots this variable in event time. There is a distinct break visible at the event date. So on average, the correlated analyst coverage increases with respect to the S&P500 stocks when these stocks are added to the index.

If the increase in correlated information is responsible for some of the increase in return comovement, we would expect this increase to be highest where the increase in correlated analyst coverage is greatest. I regress the changes in excess comovement defined by the four measures on changes in the analyst coverage variable from the calendar years before and after the event. Table 7 presents the results using 386 firms that are added to the index from 1982 through 2007 for a total of 136,291 observations. In the simple regression of $\Delta \rho^{\text{com}}_{ij}$ on $\Delta \rho^{\text{com}}_{ij}$, the coefficients using all 4 comovement definitions all between about 5% and 6% and are highly statistically significant. The overall fit of the model, however is fairly low. Most of the total variation in changes in comovement with these firms is not explained by changes in analyst coverage alone. Once year fixed effects are added, however, the total fit does increase, especially in the raw returns case. This is probably capturing some of the changes in market returns during this period. Controlling for the average increase in comovement for each firm, the correlated analyst coverage variable still exhibits a positive relation with comovement. This is shown in the third column of the table.

Overall, the effects of changes in the correlated analyst coverage on excess comovement are consistently positive and significant around the addition date. Across all specifications which include $\Delta \rho^{\text{com}}_{ij}$, its coefficient ranges between 4.3 and 8.7 percent, suggesting that increases in analyst
coverage do lead to an increase in comovement that is distinct from changes in fundamental risk characteristics.

4.7 Turnover around Brokerage House Mergers

Mergers of brokerage houses have been used in the literature as a natural experiment which do not affect stocks covered by the associated analysts. Hong and Kacperczyk (2010) use exogenous analyst turnover around 15 mergers to examine the effect of competition on analyst earnings forecast bias. As analysts become redundant following a merger, analyst coverage changes to accommodate more analysts, or as a result of analyst turnover at the resulting firm. Because nothing is likely to have changed in terms of the fundamental risk characteristics of the underlying firms covered by these brokerage houses, these events provide an ideal way to test whether analyst coverage leads to increased excess comovement, and is not simply capturing unmeasured risk.

Using 14 of the 15 mergers in Table III of Hong and Kacperczyk4, I calculate correlated analyst coverage, correlated mutual fund holdings, correlated institutional holdings, and excess comovement before and after the events for the subset of S&P500 stocks covered by either or both firms before or after the event. Then, I calculate changes in these variables (denoted with a delta, Δ). Factor loadings are calculated using data from the one-year period beginning two years prior to the merger. Panel A of Table 8 presents summary statistics for these changes around the 14 mergers. There are 387,181 stock-pair observations. Correlated analyst coverage increases by an average of 0.02 in the stocks covered by the two merging firms. The standard deviations for the change in analyst coverage is about 8%. Both the mutual fund and institutional holdings variables also increase around this window - about 0.017 and 0.021, respectively. Raw return correlations increase around the event by 0.028 on average. After controlling for risk, excess comovement decreases on average by about 2 to 4 percent. However, the standard deviation of the change in excess comovement relatively large – 13 to 16 percent.

To examine whether exogenous changes in analyst coverage lead to increases in excess comovement, I run a cross sectional regression of the change in (excess) comovement on changes in correlated analyst coverage, Δρ^an_{ij}. The first column of of Panels B through E presents the coefficients, standard errors and overall fit of these regressions using each of the four measures of comovement. Panel B shows that increases in correlated analyst coverage that result from the

---

merger are associated with an increase in raw return correlation of 2.6% on the margin. Panels C, D and E show that after controlling for risk, this affect is stronger – 3.5% using CAPM alphas, 4.6% using FF3 alphas, and 8.1% using FF4 alphas. All of these coefficients are statistically significant and are roughly the same magnitude as those estimated using additions to the S&P500 index.

If the brokerage house or individual analysts cater to institutional investors, it could be the case that the change in the analyst coverage variable is picking up the effect of a change in style investors. With this motivation, I add the change in correlated mutual fund holdings, \( \Delta \rho_{ij}^{mf} \), in the regression in the second column, and the change in correlated institutional holdings, \( \Delta \rho_{ij}^{in} \), in the regression in the third column. None of these variables is statistically significant in explaining the cross section of changes in comovement around these mergers. Furthermore, the coefficient on the change in the analyst coverage variable is not reduced by the addition of these variables. Thus, as was the case when examining changes around additions to the S&P500 index, there is a positive and significant relationship between correlated analyst coverage and excess comovement around brokerage house mergers, suggesting that the information structure itself is responsible for some level of excess comovement.

### 4.8 Unexpected Earnings and Earnings Forecast Revisions

Studies by Chan, Jegadeesh and Lakonishok (1996), and Chordia and Shivakumar (2006) have documented a positive relation between standardized unexpected earnings (SUE, see Foster, Olsen, and Shevlin, 1984) and abnormal returns. SUE is calculated as

\[
SUE_{it} = \frac{eps_{iq} - eps_{iq-4}}{\sigma_{it}},
\]

where \( eps_{iq} \) is the most recent quarterly earnings per share announcement and \( \sigma_{it} \) is the estimated standard deviation of unexpected earnings, \( eps_{iq} - eps_{iq-4} \), over the previous 8 quarters. To the extent that similar stocks are followed by similar analysts and experience similar patterns of unexpected earnings, the analyst coverage measure may be simply proxying for the correlation in unexpected earnings. To test this, I estimate the correlation in unexpected earnings, \( \rho_{ij}^{ue} \), for every pair of S&P500 stocks for which sufficient data are available and examine the effect of this variable on the significance of the coefficients of \( \rho_{ij}^{on} \) in explaining excess comovement.

For each calendar year from 1990 to 2007, I estimate \( \rho_{ij}^{ue} \) using the previous five years worth of quarterly earnings data. I restrict the sample to those S&P500 stocks that are listed on the
index during the entire period and for which there are no missing earnings data. For every pair of stocks that meet these criteria, I estimate correlation coefficients using a 16-quarter time series of unexpected earnings, $\text{eps}_{iq} - \text{eps}_{i,q-4}$. Because I focus on correlations, standardizing the unexpected earnings is less important. During this 13-year period, an average of 259 stocks per year are included in the analysis. That leads to 436,139 estimates of $\rho_{ij}^{ue}$. Requiring a firm to have a 5 years of quarterly data on earnings reduces the sample size considerably. Panel C of Table 1 shows the summary statistics. This variable has a positive mean of about 6.4%, a standard deviation of 36% and a very wide range. The variable itself has a correlations of 4.9% with $\rho_{ij}^{en}$. While this number is not large in magnitude, it is highly statistically significant.

The first two columns of the four panels in Table 9 present the results of regressing excess comovement on correlated analyst coverage without and with correlated unexpected earnings along with other control variables. The same sample is used in each regression. As shown in the second column of each panel, the inclusion of $\rho_{ij}^{ue}$ in the regression does not significantly affect the magnitude of the coefficient of $\rho_{ij}^{en}$, which is about 24% using raw returns, and about 27% using alphas from each of the three asset pricing models. The coefficient on the unexpected earnings variable is between 1.6% and 1.8% using the risk adjusted returns and about 2.4% using raw returns. All coefficients are statistically significant. Adjusted R square increases slightly with the inclusion of the unexpected earnings variable. Thus, the analyst coverage variable does not seem to be simply a proxy for common risk exposure to earnings innovations as measured by unexpected earnings.

Da and Warachka (2009) show that a beta calculated using analyst forecast error revisions can explain much of the value premium, the size premium, and the long-term reversal puzzle. This variable is meant to capture the sensitivity of a stock to changes in market-wide revisions in expected cash flow. It is possible that my analyst coverage variable is indirectly capturing some of this effect. To rule this out, I create a correlated analyst earnings forecast revisions variable, $\rho_{ij}^{er}$. Using the I/B/E/S database for each stock in the S&P500 index, I calculate monthly consensus annual earnings-per-share forecasts for the next fiscal year. Next, I use the previous five-year timeseries of monthly data to calculate sample correlations between every pair of stocks. In Panel C of Table 1, we see that this variable is about 10.3% on average, with a standard deviation of 48.5% and ranges between -98% and 99.7%.

The results of regressing excess comovement on correlated analyst coverage without and with
correlated earnings forecast revisions are in the third and fourth columns of the panels in Table 9. As was the case with the unexpected earnings measure, the coefficients on correlated earnings revisions are statistically significant, but do not significantly affect the coefficient of the analyst coverage variable. Using the same sample, the coefficient on correlated analyst coverage drops only slightly from 23.7% to 23.5% using raw returns and from 27.1% to 26.9% using CAPM alphas and from 27.0% to 26.8% using the FF3 and FF4 models. So, the inclusion of each of these two earnings based measures of risk does not significantly affect the ability of the correlated analyst coverage variable in explaining excess comovement.

4.9 Regulation Fair Disclosure

In August of 2000, the SEC implemented Regulation Fair Disclosure (See Heflin, et al, 2003 for a review) which changed the way analysts collect and process data. Public firms were now required to disclose information to all investors at the same time. Because this regulation limited the extent of analysts’ informational advantages, they were forced to rely less on private, firm-specific information, and more on their own abilities to process public information. Because firm-specific private information becomes more difficult to obtain, the relative amount of firm-specific private info relative in forecasts drops. Thus, correlations across individual analysts’ forecasts should increase. As a result, their forecasts across stocks are more likely to be correlated with macro level information. This should lead to an increase in the effect of $\rho_{ij}^\text{on}$ on return correlations.

To test this hypothesis, I create the variable $\text{REGFD}$ which indicates whether the observation takes place after the implementation of Regulation FD. This variable takes on the value 1 for the years 2001–2007 and zero otherwise. Because observations from 2000 are not included in the regression, the number of observations used drops slightly to 922,294. The regressions with GICS industry dummies are repeated and results are in Table 10. Using all 4 asset pricing models, there is an average increase in excess comovement ranging from 3% to 10% after the adoption of REG FD. However, these coefficients are not statistically significant. However, the coefficient on the interaction between this dummy variable and the correlated analyst coverage variable ranges between 10% and 53%, and is statistically significant when using the three asset pricing equations to determine excess comovement. So, after Regulation Fair Disclosure is implemented, correlated analyst coverage becomes much more important in determining excess comovement in the cross section of returns. This is consistent with individual analysts who are forced to rely relatively more on public or aggregate information making common mistakes across stocks.
5 Conclusion

When relevant information is correlated across stocks, the stocks’ returns will tend to comove more than predicted by standard asset pricing models. This paper develops a new, easy to calculate measure of correlated information based on analyst coverage. A one percent increase in this measure leads to about a 0.4 percent increase in excess comovement before controlling for industry, and a 0.25 percent increase after controlling for industry. This measure is strongest when measured using a subset of “star analysts” and does not seem to be simply a proxy for risk exposure, for style investors, or for correlations in unexpected earnings or earnings revisions. The effect increases after the implementation of REG FD and is significant around additions to the S&P500 index or around brokerage firms mergers, where there is a change in the information structure, but not necessarily in the risk exposure of the stock. The effect is weaker, but still present at the brokerage firm level, and when using a social networking measure based on past working relationships between analysts. Combined with the standard asset pricing models, this measure helps to provide a benchmark level of comovement against which to define “excess”.


Fracassi, C, 2009, Corporate finance policies and social networks. working paper.


Sørensen, Thorvald, 1948, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab* 5, 1–34.


Thompson, S., 2009, Simple formulas for standard errors that cluster by both firm and time. Working paper, Harvard University.


A Appendix

A.1 Model

To illustrate how correlated earnings forecast errors will lead to excess comovement in the presence of rational Bayesian updating, I use a neoclassical investment model. Because these models link prices and returns to earnings, they provide a natural framework in which to incorporate earnings forecasts. The mechanism is simple enough, however, that a Rational Expectations Equilibrium model in the spirit of Admati (1985) will generate similar results. First, I derive stock return correlations in the model with no signals about earnings, and then I incorporate earnings forecasts and examine the effect of these signals on the correlations.

Firms maximize the present value of future cash flows given the pricing kernel as in the standard neoclassical investment model. At time 0, given the current capital, $k_0$, the firm chooses the amount of investment $i_0$ and capital $k_1$ which will produce output, $\pi_1$ at time 1. After the output is produced, it is consumed along with the capital after depreciation, $(1 - \delta)k_1$. Capital at time 1 is equal to capital stock at time 0 plus investment minus depreciation:

$$k_1 = k_0(1 - \delta) + i_0.$$  \hspace{1cm} (12)

For simplicity, there are no adjustment costs. Firms maximize the market value of equity with respect to the exogenous pricing kernel, $m_{0,1}$:

$$\max \{ \pi(k_0, x_0, y_0) - i_0 + E_0 [m_{0,1} [\pi(k_1, x_1, y_1) + (1 - \delta)k_1]] \}$$  \hspace{1cm} (13)

The first order condition is:

$$1 = E_0 [m_{0,1} [\pi_1(k_1, x_1, y_1) + (1 - \delta)]].$$  \hspace{1cm} (14)

As in the standard case, the following relations hold:

$$E_t [m_{0,1} r^f_{0,1}] = 1$$  \hspace{1cm} (15)

where $r^f_{0,1}$ is the investment return between periods 0 and 1 and is equal to

$$r^f_{0,1} = \frac{\pi_1(k_1, x_1, y_1) + (1 - \delta)}{1}.$$  \hspace{1cm} (16)

Furthermore,

$$P_0 = E_0 [m_{0,1} [\pi(k_1, x_1, y_1) + (1 - \delta)k_1]],$$  \hspace{1cm} (17)

where $P_0$ is the ex-dividend value of equity at period 0. When $\pi_t$ is linearly homogeneous, we can use the first order condition, (14) to get: $P_0 = k_1$. The stock return over this period is:

$$r^f_{0,1} = \frac{\pi(k_1, x_1, y_1) + (1 - \delta)k_1}{E_t [m_{0,1} [\pi(k_1, x_1, y_1) + (1 - \delta)k_1]]}.$$  \hspace{1cm} (18)

Substituting the first order condition, we get the standard results equating stock returns and investment returns over the period 0 to 1. Let’s impose some structural form on the functions. Let
the production function for the \( i \)th be as follows:

\[
\pi_{it} = \pi(k_{it}, x_t, y_{it}) = (x_t + y_{it})k_{it},
\]

where the systematic productivity, \( x_t \), evolves according to:

\[
x_{t+1} = \varpi(1 - \rho_x) + \rho_x x_t + \sigma_x \varepsilon_{x,t+1}^x,
\]

where \( \varepsilon_{x,t+1}^x \) is an i.i.d. standard normal variable, and the idiosyncratic productivity for firm \( i \), \( y_{it} \), evolves according to:

\[
y_{it+1} = \bar{y}_i(1 - \rho_{y_i}) + \rho_{y_i} y_{iu} + \sigma_{y_i} \varepsilon_{y,t+1}^y.
\]

The implicit assumption behind the structure of this production function is that all firms have the same exposure to the systematic productivity variable. Generalizing this function to allow the risk exposures to vary across stocks or including multiple sources of systematic risk is straightforward. However, for the sake of simplicity, and because I am focusing on “excess” comovement, and will be holding fundamentals constant, I use this simple form. The variable \( x_t \) has the following conditional moments:

\[
\mu_{x_1|x_0} \equiv E[x_1 \mid x_0] = \varpi(1 - \rho_x) + \rho_x x_0
\]

\[
\sigma_{x_1|x_0}^2 \equiv Var(x_1 \mid x_0) = \sigma_x^2
\]

and similar for \( y_{it} \). The pricing kernel is assumed to be of the following form:

\[
\log m_{t,t+\tau} = \log \eta^\tau + \gamma(x_t - x_{t+\tau})
\]

**Correlations without Signals**

The variance of the stock return of any firm \( i \) from time 0 to time 1 is equal to

\[
Var_0(r_{0,1}^i) = Var_0 \left( \frac{P_{i1}}{P_{i0}} \right) = \frac{1}{k_{i1}^2} Var_0(P_{i1}) = Var_0(x_1 + y_{i1})
\]

The covariance of the stock returns of firm \( i \) and firm \( j \) can be written:

\[
Cov_0(r_{0,1}^i, r_{0,1}^j) = \frac{1}{k_{i1}k_{j1}} Cov_0(P_{i1}, P_{j1}) = Cov_0(x_1 + y_{i1}, x_1 + y_{j1})
\]
Thus, the correlation between the two stock returns is

$$
\rho_{ij} \equiv \text{Corr}_0(r_{0,1}^i, r_{0,1}^j) = \frac{\text{Cov}_0(r_{0,1}^i, r_{0,1}^j)}{\sqrt{\text{Var}_0(r_{0,1}^i)} \sqrt{\text{Var}_0(r_{0,1}^j)}}
$$

(31)

$$
\rho_{ij} = \frac{\text{Cov}_0(x_1 + y_{i1}, x_1 + y_{j1})}{\sqrt{\text{Var}_0(x_1 + y_{i1})} \sqrt{\text{Var}_0(x_1 + y_{j1})}}
$$

(32)

$$
\rho_{ij} = \frac{\sigma_x^2}{\sqrt{(\sigma_x^2 + \sigma_y^2)(\sigma_x^2 + \sigma_y^2)}}
$$

(33)

When the conditional volatility of the systematic productivity variable is large relative to the idiosyncratic volatilities, this correlation will be close to one. When there is a large amount of firm-specific-productivity uncertainty, this value will be relatively small. If the state variables, $x_1$, $y_{i1}$ and $y_{j1}$ are partially inferred from some sort of a signal, the correlations may not be the same as in the above case, as is described in the next subsection.

## A.2 Information

To examine the effects of correlated information on comovement, I introduce correlated signals into the model. At time $0^+$, immediately after the firms make their investment decisions, earnings forecasts are released. No investment or production takes place at time $0^+$. The only other difference between times 0 and $0^+$ is the information set. None of the state variables differ (e.g., $x_0 = x_{0^+}$). The state $x_0$, $y_{i0}$, $y_{j0}$ and $y_{0}$ are known with certainty. At times 0 and $0^+$. Also, assume that at time 1, the states $x_1$, $y_{i1}$, $y_{j1}$ and $y_{11}$ will be also known with certainty. The earnings forecasts, $\tilde{e}_i$, are of the form $\tilde{e}_i = \pi + e_i k_{i1} = (x_1 + y_{i1} + e_i)k_{i1}$, where the forecast error, $e_i$, is a normal random variable which has a variance $\sigma_e^2$ and a covariance $\sigma_{eij}$. Because the levels of capital are known at time $0^+$ – one period in advance – an equivalent forecast is $z_i = x_1 + y_{i1} + e_i$. Using these earnings forecasts at time $0^+$, the representative agent prices all stocks based on the known states at time $0^+$, and the best estimate of the future states given all of the earnings forecasts.

### Correlations with Signals

As already shown, before the earnings forecasts are released (or in the case without earnings forecasts) the market value of equity for firm $i$, $P_{i0}$ is equal to the optimal capital stock, $k_{i1}$. At time 1, the value, $P_{i1}$, is equal to the cash flow plus the non depreciated portion of the capital stock:

$$
P_{i1} = (x_1 + y_{i1})k_{i1} + (1 - \delta)k_{i1}
$$

(34)

As soon as the earnings forecasts are released, the price will depend on the signals (earnings forecasts) of all firms:

$$
P_{i0^+} = E_{0^+}[m_{0^+,1}P_{i1}],
$$

(35)
We can solve for this based on the state variables at time 0 and the signals. However, because we are focusing on correlations, we need not do so. Similar to the case with no earnings forecasts,

\[
Var_{0+}(r_{0+1}) = \frac{1}{P_{i0}^+} Var_{0+}(P_{i1}) = \frac{k_{i1}^2}{P^2_{i0+}} Var_{0+}(x_1 + y_1)
\]

(36)

\[
Cov_{0+}(r_{0+1}, r_{0+1}) = \frac{k_{i1}k_{j1}}{P^2_{i0+}P^2_{j0+}} Cov_{0+}(x_1 + y_1, x_1 + y_1).
\]

(37)

So,

\[
\rho_{ij}^* = Corr_{0+}(r_{0+1}, r_{0+1}) = \frac{Cov_{0+}(x_1 + y_1, x_1 + y_1)}{\sqrt{Var_{0+}(x_1 + y_1)Var_{0+}(x_1 + y_1)}}.
\]

(38)

This equation differs from (33) only in the information set. If the signals are sufficiently noisy, the two will be close to the same. However, if the signals are informative, the two may differ. The conditional stock return correlation will be a function of the parameters of the state variables and of the signals, and can be determined by solving for the components of (38).

### A.3 Kalman Filter

Due to the linear structure of the model, the Kalman filter algorithm or linear least squares method provides the best estimate of the state variables and their covariances at time 1, given the signals at time 0+. The Kalman Filter can be applied as follows. the state process, \( s_t \)

\[
s_1 = As_0 + B\bar{z} + \varepsilon_t
\]

(39)

with measurements that are

\[
z_{0+} = Cs_1 + e_1.
\]

(40)

In our case, the state variable vector \( s_t \in \mathbb{R}^{n+1} \) is \( s_t = (x_t, y_{1t}, y_{2t}, \ldots, y_{nt})' \), where \( x_t \) represents the systematic level of productivity and \( y_{it} \) is the idiosyncratic component of productivity for firm \( i \) and \( \bar{z} = (\bar{x}, \bar{y}^1, \bar{y}^2, \ldots, \bar{y}^n) \). The matrices \( A, B, \) and \( C \) are as follows:

\[
A = \begin{bmatrix}
\rho_x & 0 & 0 & \cdots & 0 \\
0 & \rho_y^1 & 0 & \cdots & 0 \\
0 & 0 & \rho_y^2 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \rho_y^n
\end{bmatrix},
\]

(41)

\[
B = \begin{bmatrix}
(1 - \rho_x) & 0 & 0 & \cdots & 0 \\
0 & (1 - \rho_y^1) & 0 & \cdots & 0 \\
0 & 0 & (1 - \rho_y^2) & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & (1 - \rho_y^n)
\end{bmatrix},
\]

(42)

---

the following section, I calibrate the model and show numerically that return comovement increases introducing a measure of correlated analyst coverage which proxies for correlated forecast errors in forecast errors. See the appendix for a solution using two firms. After describing the data and show that the stock return correlation between any two stocks increases with the correlation in and thus comparative statics are hard to generalize. For this reason, I calibrate the model and inverting a matrix, when the number of stocks is large, it becomes difficult to solve symbolically

Applying the Kalman filter algorithm to estimate the states, \( \hat{s}_1 \) and their variance-covariance matrix, \( \Sigma \):

\[
\hat{s}_1 = E[s_1 \mid I_{0+}] = A\hat{s}_0 + B\bar{\pi} + \Sigma C'N^{-1}(z_{t+1} - C(A\hat{s}_t + B\bar{\pi}))
\]

\[
\Sigma = E[(s_1 - \hat{s}_1)(s_1 - \hat{s}_1)' \mid I_{0+}] = M - MC'(CMC' + N)^{-1}CM.
\]

where \( M \) is the \( n+1 \times n+1 \) diagonal variance-covariance matrix

\[
M = \begin{bmatrix}
\sigma_x^2 & 0 & 0 & \cdots & 0 \\
0 & \sigma_y^1 & 0 & \cdots & 0 \\
0 & 0 & \sigma_y^2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \sigma_y^n
\end{bmatrix}
\]

and \( N \) is the \( n \times n \) diagonal variance-covariance matrix

\[
N = \begin{bmatrix}
\sigma_{e1}^2 & \sigma_{e1,2} & \sigma_{e1,3} & \cdots & \sigma_{e1,n} \\
\sigma_{e1,2} & \sigma_{e2}^2 & \sigma_{e2,3} & \cdots & \sigma_{e2,n} \\
\sigma_{e1,3} & \sigma_{e2,3} & \sigma_{e3}^2 & \cdots & \sigma_{e3,n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sigma_{e1,n} & \sigma_{e2,n} & \sigma_{e3,n} & \cdots & \sigma_{en}^2
\end{bmatrix}
\]

Given the \( \Sigma \) matrix, we can solve for the predicted conditional stock return correlation at time \( 0^+ \): Given the signals at time \( 0^+ \), The variances, covariances, and correlations of the stock returns are

\[
\rho_{12}^+ \equiv Corr_0^+(r_{0^+1}, r_{0^+2}) = \frac{Cov_0^+(x_1 + y_{11}, x_1 + y_{21})}{\sqrt{Var_0^+(x_1 + y_{11})Var_0^+(x_1 + y_{21})}},
\]

where

\[
Cov_0^+(x_1 + y_{11}, x_1 + y_{11}) = \Sigma_{1,1} + \Sigma_{1,i+1} + \Sigma_{1,j+1} + \Sigma_{i+1,j+1}
\]

\[
Var_0^+(x_1 + y_{11}) = \Sigma_{1,1} + 2\Sigma_{1,i+1} + \Sigma_{i+1,i+1} + \Sigma_{i+1,j+1} + \Sigma_{j+1,i+1} + \Sigma_{j+1,j+1}
\]

\[
Var_0^+(x_1 + y_{11}) = \Sigma_{1,1} + 2\Sigma_{1,i+1} + \Sigma_{j+1,i+1} + \Sigma_{j+1,j+1}
\]

and \( \Sigma_{k,l} \) denotes the element in the \( k \)th row and \( l \)th column of \( \Sigma \). Because \( \Sigma \) is calculated by inverting a matrix, when the number of stocks is large, it becomes difficult to solve symbolically and thus comparative statics are hard to generalize. For this reason, I calibrate the model and show that the stock return correlation between any two stocks increases with the correlation in forecast errors. See the appendix for a solution using two firms. After describing the data and introducing a measure of correlated analyst coverage which proxies for correlated forecast errors in the following section, I calibrate the model and show numerically that return comovement increases
with increases in correlated forecast errors.

**A.4 Kalman Filter: 2 Stock Example**

When the number of stocks is large—as is the case in practice—it is difficult to solve for the variance-covariance matrix of the states symbolically. To give some intuition as to what such a solution might look like, I solve for the two-stock case in this section. In this case, the signals for stocks $i$ and $j$ are $z_i = x_1 + y_{i1} + \sigma_{ei} w_i$ & $z_j = x_1 + y_{j1} + \sigma_{ej} w_j$, which are revealed at time $0^+$. We want to solve for

$$
\Sigma = M - MC' (CMC' + N)^{-1} CM.
$$

(52)

In this case, we have:

$$
M = \begin{bmatrix}
\sigma_x^2 & 0 & 0 \\
0 & \sigma_{y_1}^2 & 0 \\
0 & 0 & \sigma_y^2
\end{bmatrix}
$$

(53)

$$
N = \begin{bmatrix}
\sigma_{e1}^2 & \sigma_{e1,2} \\
\sigma_{e1,2} & \sigma_{e2}^2
\end{bmatrix}
$$

(54)

$$
C = \begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1
\end{bmatrix}.
$$

(55)

Thus,

$$
CM = \begin{bmatrix}
\sigma_x^2 & \sigma_{y_1}^2 & 0 \\
\sigma_x^2 & 0 & \sigma_y^2
\end{bmatrix},
$$

(56)

$$
MC' = \begin{bmatrix}
\sigma_x^2 & \sigma_x^2 \\
\sigma_{y_1}^2 & 0 \\
0 & \sigma_y^2
\end{bmatrix},
$$

(57)

and

$$
(CMC' + N)^{-1} = \frac{1}{D} \begin{bmatrix}
\sigma_x^2 + \sigma_{y_2}^2 + \sigma_{e2}^2 & -(\sigma_x^2 + \sigma_{e1,2}) \\
-(\sigma_x^2 + \sigma_{e1,2}) & \sigma_x^2 + \sigma_{y_1}^2 + \sigma_{e1}^2
\end{bmatrix},
$$

(58)

where

$$
D = (\sigma_x^2 + \sigma_{y_1}^2 + \sigma_{e1}^2)(\sigma_x^2 + \sigma_{y_2}^2 + \sigma_{e2}^2) - (\sigma_x^2 + \sigma_{e1,2})^2.
$$

(59)
Combining these equations, we can solve for $\Sigma$:

$$
\Sigma_{11} = Var_{0^+}(x_1) = \frac{\sigma_x^2 \left[ (\sigma_{y_1}^2 + \sigma_{e_1}^2)(\sigma_{y_2}^2 + \sigma_{e_2}^2) - (\sigma_{e_{1,2}})^2 \right]}{D}
$$

$$
\Sigma_{22} = Var_{0^+}(y_1) = \frac{\sigma_{y_1}^2 \left[ (\sigma_x^2 + \sigma_{e_1}^2)(\sigma_{y_2}^2 + \sigma_{e_2}^2) + \sigma_x^2 \sigma_{e_1}^2 - \sigma_{e_{1,2}}(\sigma_{e_{1,2}} + 2\sigma_x^2) \right]}{D}
$$

$$
\Sigma_{33} = Var_{0^+}(y_2) = \frac{\sigma_{y_2}^2 \left[ (\sigma_x^2 + \sigma_{e_2}^2)(\sigma_{y_1}^2 + \sigma_{e_1}^2) + \sigma_x^2 \sigma_{e_2}^2 - \sigma_{e_{1,2}}(\sigma_{e_{1,2}} + 2\sigma_x^2) \right]}{D}
$$

$$
\Sigma_{12} = Cov_{0^+}(x_1, y_1) = -\frac{\sigma_x^2 \sigma_{y_1}^2 (\sigma_{y_2}^2 + \sigma_{e_2}^2 - \sigma_{e_{1,2}})}{D}
$$

$$
\Sigma_{13} = Cov_{0^+}(x_1, y_2) = -\frac{\sigma_x^2 \sigma_{y_2}^2 (\sigma_{y_1}^2 + \sigma_{e_1}^2 - \sigma_{e_{1,2}})}{D}
$$

$$
\Sigma_{23} = Cov_{0^+}(y_1, y_2) = \frac{\sigma_{y_1}^2 \sigma_{y_2}^2 (\sigma_x^2 + \sigma_{e_{1,2}})}{D}
$$

where $\Sigma_{ij}$ is the element in the $i$th row and $j$th column of $\Sigma$.

Given the signals at time $0^+$, The variances, covariances, and correlations of the stock returns are

$$
\rho_{12}^+ = Corr_{0^+}(r_{0^+,1}^+, r_{0^+,1}^2) = \frac{Cov_{0^+}(x_1 + y_1, x_1 + y_2)}{\sqrt{Var_{0^+}(x_1 + y_1)Var_{0^+}(x_1 + y_2)}},
$$

where

$$
Cov_{0^+}(x_1 + y_1, x_1 + y_2) = \Sigma_{11} + \Sigma_{12} + \Sigma_{13} + \Sigma_{23}
$$

$$
Var_{0^+}(x_1 + y_1) = \Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}
$$

$$
Var_{0^+}(x_1 + y_2) = \Sigma_{11} + 2\Sigma_{13} + \Sigma_{33}.
$$

### A.5 Calibration

In principal, the predicted stock return correlation for any two stocks from the model, $Corr_{0^+}(r_{0^+,1}^+, r_{0^+,1}^j)$, can be solved symbolically as a function of the underlying parameters. However, when the number of signals is large, this becomes intractable as it involves inverting a matrix whose size increases with the number of stocks. So solving for partial derivatives when there are 500 stocks becomes infeasible. To avoid this, I calibrate the relevant model parameters, $\sigma_x^2$, $\sigma_{y^1}^2$, $\sigma_{e_1}^2$ and $\sigma_{e_{1,2}}$, and show numerically that the stock return correlation between any two stocks increases with the covariance of the stocks’ earning forecast errors.

Because the earnings cycle is typically quarterly, I focus on that frequency. I set the conditional volatility, $\sigma_x$, of the systematic productivity process to 0.007, and the conditional volatility, $\sigma_{y^1}$, of the firm-specific productivities to 0.3. These quarterly parameters are consistent with Cooley and Prescott (1995) and Zhang (2005). I estimate the standard deviation, $\sigma_{e_i}$ of the forecast error using the realized errors from the quarterly consensus earnings-per-share forecasts for the sample period from 1982 through 2007 for all S&P500 stocks in the I/B/E/S database. The standard deviation of the quarterly earnings per share forecast error during this period is 0.423. Because of the assumption that earnings are a linear function of the capital stock, $K$, this error needs to be scaled by capital-per-share. Once this adjustment is made, the realized standard deviation is 0.373,
which I use as an estimate of $\sigma_{e_t}$. Because firms report earnings at different moments in time, the earnings forecast error covariance parameter, $\sigma_{e_{ij}}$, is difficult to estimate directly. However because I am interested in the theoretical effect of changes in this parameter on stock return correlations, I vary this parameter for stocks $i$ and $j$ and fix it for all other pairs of stocks and examine the changes in $\rho^{ij}$ using equation (38). I set the error covariances for all other pairs of the 500 stocks used in the procedure such that the error correlations are 0.026 — the sample average value of $\rho_{ij}^{an}$.

Figure 2 shows the results of varying the forecast error correlation on the stock return comovement for two stocks using these assumptions. Panel A shows that comovement increases from close to zero all the way towards one as the error correlation increases. Panel B plots the slope of the curve in Panel A. This is the marginal effect of increasing the error correlation on comovement. This marginal effect is between 0.4 and 0.5 until $\rho_{ij}^{an}$ is increased to above 0.35. When $\rho_{ij}^{an}$ is equal to 0.027, its sample average, $\partial\rho_{ij}^{ret}/\partial\rho_{ij}^{an}$ equals about 0.40.

These results provide the following testable hypothesis: firms with high correlated analyst coverage, $\rho_{ij}^{an}$, have high comovement $\rho_{ij}^{ret}$. Specifically, in a regression of comovement on correlated analyst coverage, the coefficient should be positive. According to the calibration, around our sample mean of 0.027, this coefficient should be around 0.40. I test this hypothesis in the body of the paper. If individual analysts’ earnings forecasts are not perfectly correlated across stocks, then the effect will be smaller.
Figure 1: Correlated Analyst Coverage - Additions to S&P500

These panels present average correlated analyst coverage, $\rho_{ij}^a$, between firms added to the S&P500 index from 1982 through 2007 and all other S&P500 firms by quarter from the 12th quarter before the addition date to the 12th quarter after the date. Analyst data are from the I/B/E/S database. S&P500 constituent data are from CRSP. The values plotted in these graphs are found in Table 6.

Panel A: Correlated Analyst Coverage
Figure 2: Comovement and Correlated Forecast Errors

These panels show the effect of changing the forecast error correlation on return comovement in the model. Panel A plots the error correlation variable, $\rho_{ij}^{an}$ against the return correlation $\rho_{ij}^{ret}$. Panel B plots the slope of this line, $\frac{\partial \rho_{ij}^{ret}}{\partial \rho_{ij}^{an}}$. In the calibration, $\sigma_e$ is set to 0.007, $\sigma_\eta$ to 0.3, $\sigma_{e,t}$ to 0.373, and $\sigma_{e,t}$ to 0.028 for all stocks. Then the $\sigma_{e,t}$ is varied between 0 and 0.373 for two stocks and $\rho_{ij}^{ret}$ is then calculated for these two stocks.

Panel A: Return Correlations and Correlated Forecast Errors

Panel B: Changes in Return Correlations and Correlated Forecast Errors
Table 1: Summary Statistics

This table reports summary statistics for the correlated analyst coverage variables and the stock return correlation coefficients for pairs of S&P500 stocks from the period 1995–2007. Panel A presents the correlated analyst coverage, correlated star analyst coverage, correlated non-star analyst coverage, correlated broker coverage, and connected analyst coverage. The correlation coefficients used in Panels B are estimated using raw returns, and residuals from the CAPM, the Fama French 3 factor model, and the Fama French 3 factor model plus the Carhart momentum factor. The three models are fit using one year’s worth of daily data and are applied to the following year’s data to estimate the correlation coefficients. Panel C presents statistics for correlated mutual fund holdings and correlated institutional holdings along with correlation coefficients for unexpected earnings and analyst earnings forecast revisions. Analyst data are from the I/B/E/S database and Institutional Investor magazine, stock returns are from CRSP, the Fama French and Carhart factor data are from WRDS, mutual fund and institutional investor holdings data are from Thomson-Reuters, and earnings and earnings forecasts are from I/B/E/S. Means, standard deviations, minimums and maximums are expressed in percentage terms.

### Panel A: Analysts and Brokers

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^a$</td>
<td>Analyst Coverage</td>
<td>985484</td>
<td>2.7</td>
<td>8.6</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\rho_{ij}^s$</td>
<td>Star Analysts</td>
<td>985484</td>
<td>0.2</td>
<td>1.6</td>
<td>0.0</td>
<td>57.2</td>
</tr>
<tr>
<td>$\rho_{ij}^{\text{non}^*}$</td>
<td>Non-Star Analysts</td>
<td>985484</td>
<td>2.5</td>
<td>7.5</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\rho_{ij}^b$</td>
<td>Brokerage Coverage</td>
<td>985484</td>
<td>48.6</td>
<td>17.3</td>
<td>0.0</td>
<td>189.7</td>
</tr>
<tr>
<td>$\rho_{ij}^{ca}$</td>
<td>Connected Analysts</td>
<td>985484</td>
<td>32.7</td>
<td>19.0</td>
<td>0.0</td>
<td>201.2</td>
</tr>
</tbody>
</table>

### Panel B: Stock Return Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{\text{ret}}$</td>
<td>Raw Returns</td>
<td>985484</td>
<td>23.9</td>
<td>14.2</td>
<td>-34.6</td>
<td>96.2</td>
</tr>
<tr>
<td>$\rho_{ij}^{\text{capm}}$</td>
<td>CAPM Alphas</td>
<td>985484</td>
<td>7.0</td>
<td>16.9</td>
<td>-55.2</td>
<td>97.5</td>
</tr>
<tr>
<td>$\rho_{ij}^{\text{ff3}}$</td>
<td>FF Alphas</td>
<td>985484</td>
<td>9.7</td>
<td>18.2</td>
<td>-51.8</td>
<td>97.5</td>
</tr>
<tr>
<td>$\rho_{ij}^{\text{ff4}}$</td>
<td>FF+Carhart Alphas</td>
<td>985484</td>
<td>11.8</td>
<td>20.8</td>
<td>-58.6</td>
<td>98.7</td>
</tr>
</tbody>
</table>

### Panel C: Other Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{mf}$</td>
<td>Mutual Fund Holdings</td>
<td>985484</td>
<td>40.9</td>
<td>8.2</td>
<td>12.3</td>
<td>83.0</td>
</tr>
<tr>
<td>$\rho_{ij}^{in}$</td>
<td>Institutional Holdings</td>
<td>985484</td>
<td>56.5</td>
<td>6.3</td>
<td>7.6</td>
<td>92.8</td>
</tr>
<tr>
<td>$\rho_{ij}^{ue}$</td>
<td>Unexpected Earnings</td>
<td>436139</td>
<td>6.4</td>
<td>35.9</td>
<td>-97.6</td>
<td>98.5</td>
</tr>
<tr>
<td>$\rho_{ij}^{er}$</td>
<td>Earnings Revisions</td>
<td>438651</td>
<td>10.3</td>
<td>48.5</td>
<td>-98.0</td>
<td>99.7</td>
</tr>
</tbody>
</table>
Table 2: Excess Comovement and Analyst Coverage

This table reports OLS regression coefficients and standard errors from regressing stock return correlations between pairs of stocks on combinations of correlated analyst coverage, $\rho_{ij}^{an}$, correlated star analyst coverage, $\rho_{ij}^{st}$, correlated non-star analyst coverage, $\rho_{ij}^{-st}$, and various control variables. Correlated analyst coverage is defined as $N_{ij}/\sqrt{N_iN_j}$, where $N_{ij}$ is the number of common analysts for stocks $i$ and $j$ and $N_i$ and $N_j$ are the total number of analysts covering stocks $i$ and $j$, respectively. Correlated star analyst coverage and non-star analyst coverage are defined similarly, with the numerator equal to the number of common star analysts ($N_{ij}^{st}$) or non-star analysts ($N_{ij}^{-st}$), respectively. The variable gics is an indicator variable that is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. Year fixed effects are included but not reported. The correlation coefficients are estimated yearly using daily raw returns (Panel A), or the daily residuals from the CAPM (Panel B), the Fama French 3-factor model (Panel C) and the Fama French 3-factor model augmented with the Carhart momentum factor (Panel D) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP and factor data from WRDS. Analyst data are from I/B/E/S, and star analyst data are from Institutional Investor magazine. Standard errors are clustered at the firm and year level and are listed in parentheses. For the regressions which include both $\rho_{ij}^{st}$ and $\rho_{ij}^{-st}$, the difference in these two coefficients is reported along with the F-statistic and p-value for the restriction $\beta_{st} = \beta_{-st}$. 985,484 observations are used from the period 1995–2007.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Raw Rets</th>
<th>Panel B: CAPM alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{an}$</td>
<td>0.350 (0.0258)</td>
<td>0.428 (0.0560)</td>
</tr>
<tr>
<td></td>
<td>0.225 (0.0342)</td>
<td>0.265 (0.0707)</td>
</tr>
<tr>
<td>$\rho_{ij}^{st}$</td>
<td>0.807 (0.0791)</td>
<td>0.993 (0.0989)</td>
</tr>
<tr>
<td></td>
<td>0.462 (0.0868)</td>
<td>0.602 (0.1971)</td>
</tr>
<tr>
<td>$\rho_{ij}^{-st}$</td>
<td>0.192 (0.0428)</td>
<td>0.217 (0.0989)</td>
</tr>
<tr>
<td>gics</td>
<td>0.106 (0.0217)</td>
<td>0.138 (0.0275)</td>
</tr>
<tr>
<td></td>
<td>0.136 (0.0141)</td>
<td>0.171 (0.0135)</td>
</tr>
<tr>
<td></td>
<td>0.102 (0.0193)</td>
<td>0.132 (0.0223)</td>
</tr>
<tr>
<td>Adj. R-sq.</td>
<td>0.335</td>
<td>0.369</td>
</tr>
<tr>
<td></td>
<td>0.343</td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td>0.339</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>0.344</td>
<td>0.380</td>
</tr>
<tr>
<td>$\beta_{st} - \beta_{-st}$</td>
<td>0.270</td>
<td>0.385</td>
</tr>
<tr>
<td>F-stat($\beta_{st} = \beta_{-st}$)</td>
<td>604</td>
<td>450</td>
</tr>
<tr>
<td>pr &gt; F</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF4 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{an}$</td>
<td>0.417 (0.0570)</td>
<td>0.405 (0.0634)</td>
</tr>
<tr>
<td></td>
<td>0.258 (0.0731)</td>
<td>0.256 (0.0752)</td>
</tr>
<tr>
<td>$\rho_{ij}^{st}$</td>
<td>0.960 (0.0934)</td>
<td>0.938 (0.0920)</td>
</tr>
<tr>
<td></td>
<td>0.574 (0.2056)</td>
<td>0.552 (0.1917)</td>
</tr>
<tr>
<td>$\rho_{ij}^{-st}$</td>
<td>0.214 (0.1044)</td>
<td>0.214 (0.1044)</td>
</tr>
<tr>
<td>gics</td>
<td>0.134 (0.0270)</td>
<td>0.127 (0.0231)</td>
</tr>
<tr>
<td></td>
<td>0.166 (0.0124)</td>
<td>0.159 (0.0136)</td>
</tr>
<tr>
<td></td>
<td>0.128 (0.0217)</td>
<td>0.121 (0.0184)</td>
</tr>
<tr>
<td>Adj. R-sq.</td>
<td>0.420</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>0.428</td>
<td>0.593</td>
</tr>
<tr>
<td></td>
<td>0.425</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>0.429</td>
<td>0.594</td>
</tr>
<tr>
<td>$\beta_{st} - \beta_{-st}$</td>
<td>0.360</td>
<td>0.338</td>
</tr>
<tr>
<td>F-stat($\beta_{st} = \beta_{-st}$)</td>
<td>424</td>
<td>454</td>
</tr>
<tr>
<td>pr &gt; F</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Table 3: Comovement and Analyst Coverage - Social Networks

This table reports OLS regression coefficients and standard errors from regressing stock return correlations between pairs of stocks on combinations of correlated broker coverage, $\rho_{ij}^{br}$, or correlated connected analyst coverage, $\rho_{ij}^{ca}$, and various control variables. Correlated broker coverage is defined as $N_{ij}^{br} / \sqrt{N_i^{br} N_j^{br}}$, where $N_{ij}^{br}$ is the number of common brokers for stocks $i$ and $j$ and $N_i^{br}$ and $N_j^{br}$ are the total number of brokers covering stocks $i$ and $j$, respectively. Correlated connected analyst coverage is defined as $N_{ij}^{ca} / \sqrt{N_i^{ca} N_j^{ca}}$, where $N_{ij}^{ca}$ is the number of connected pairs of analysts between the two stocks and $N_i^{ca}$ and $N_j^{ca}$ are the total number of analysts covering stocks $i$ and $j$, respectively. A pair of analysts is connected if the two analysts currently work for different firms, but each worked for the same firm in the past. The variable gics is an indicator variable that is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. Year fixed effects are included but not reported. The correlation coefficients are estimated yearly using daily raw returns (Panel A), or the daily residuals from the CAPM (Panel B), the Fama-French 3-factor model (Panel C) and the Fama-French 3-factor model augmented with the Carhart momentum factor (Panel D) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP and factor data from WRDS. Analyst data are from I/B/E/S. Standard errors are clustered at the firm and year level and are listed in parentheses. 985,484 observations are used from the period 1995–2007.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Raw Rets</th>
<th>Panel B: CAPM alphas</th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF4 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{br}$</td>
<td>0.071 (0.0113)</td>
<td>0.044 (0.0120)</td>
<td>0.075 (0.0097)</td>
<td>0.041 (0.0086)</td>
</tr>
<tr>
<td>$\rho_{ij}^{ca}$</td>
<td>0.026 (0.0081)</td>
<td>0.025 (0.0076)</td>
<td>0.026 (0.0079)</td>
<td>0.024 (0.0083)</td>
</tr>
<tr>
<td>gics</td>
<td>0.173 (0.0149)</td>
<td>0.181 (0.0143)</td>
<td>0.220 (0.0145)</td>
<td>0.227 (0.0152)</td>
</tr>
<tr>
<td>Adj R-sq</td>
<td>0.299</td>
<td>0.335</td>
<td>0.293</td>
<td>0.334</td>
</tr>
</tbody>
</table>

38
Table 4: Comovement and Analyst Coverage - Mutual Fund Holdings

This table reports OLS regression coefficients and standard errors from regressing stock return correlations between pairs of stocks on combinations of correlated analyst coverage, $\rho_{ij}^m$, correlated mutual fund holdings, $\rho_{ij}^{mf}$, and various control variables. Correlated analyst coverage is defined as $N_{ij}/\sqrt{N_iN_j}$, where $N_{ij}$ is the number of common analysts for stocks $i$ and $j$ and $N_i$ and $N_j$ are the total number of analysts covering stocks $i$ and $j$, respectively. Correlated mutual fund holdings is equal to $N_{ij}^{mf}/\sqrt{N_i^{mf}N_j^{mf}}$, where $N_{ij}^{mf}$ is the number of mutual funds who own shares of both stocks $i$ and $j$, and $N_i^{mf}$ and $N_j^{mf}$ are the number of mutual funds who own shares of stock $i$ and $j$, respectively. The variable gics is an indicator variable that is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. Year fixed effects are included but not reported. The correlation coefficients are estimated yearly using daily raw returns (Panel A), or the daily residuals from the CAPM (Panel B), the Fama French 3-factor model (Panel C) and the Fama French 3-factor model augmented with the Carhart momentum factor (Panel D) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP and factor data from WRDS. Analyst data are from I/B/E/S. Mutual fund holdings data are from the Thomson Reuters Mutual Fund Holdings database. Standard errors are clustered at the firm and year level and are listed in parentheses. 985,484 observations are used from the period 1995–2007.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Raw Rets</th>
<th>Panel B: CAPM alphas</th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF4 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^m$</td>
<td>0.329 (0.0219)</td>
<td>0.388 (0.0529)</td>
<td>0.377 (0.0549)</td>
<td>0.367 (0.0603)</td>
</tr>
<tr>
<td>$\rho_{ij}^{mf}$</td>
<td>0.257 (0.0276)</td>
<td>0.463 (0.0717)</td>
<td>0.460 (0.0810)</td>
<td>0.440 (0.0703)</td>
</tr>
<tr>
<td>gics</td>
<td>0.168 (0.0131)</td>
<td>0.202 (0.0144)</td>
<td>0.196 (0.0137)</td>
<td>0.189 (0.0158)</td>
</tr>
<tr>
<td>Adj. R sq.</td>
<td>0.312 0.347 0.350</td>
<td>0.370 0.405 0.407</td>
<td>0.423 0.451 0.453</td>
<td>0.589 0.609 0.610</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Table 5: Comovement and Analyst Coverage - Institutional Holdings

This table reports OLS regression coefficients and standard errors from regressing stock return correlations between pairs of stocks on combinations of correlated analyst coverage, $\rho_{ij}^{an}$, correlated institutional holdings, $\rho_{ij}^{in}$, and various control variables. Correlated analyst coverage is defined as $N_{ij}/\sqrt{N_i N_j}$, where $N_{ij}$ is the number of common analysts for stocks $i$ and $j$ and $N_i$ and $N_j$ are the total number of analysts covering stocks $i$ and $j$, respectively. Correlated institutional holdings is equal to $N_{ij}^{in}/\sqrt{N_{i}^{in}N_{j}^{in}}$, where $N_{ij}^{in}$ is the number of institutional investors who own shares of both stocks $i$ and $j$, and $N_{i}^{in}$ and $N_{j}^{in}$ are the number of institutional investors who own shares of stock $i$ and $j$, respectively. The variable gics is an indicator variable that is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. Year fixed effects are included but not reported. The correlation coefficients are estimated yearly using daily raw returns (Panel A), or the daily residuals from the CAPM (Panel B), the Fama French 3-factor model (Panel C) and the Fama French 3-factor model augmented with the Carhart momentum factor (Panel D) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP and factor data from WRDS. Analyst data are from I/B/E/S. Institutional holdings data are from the Thomson-Reuters Institutional Holdings (13F) Database. Standard errors are clustered at the firm and year level and are listed in parentheses. 985,484 observations are used from the period 1995–2007.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Raw Rets</th>
<th>Panel B: CAPM alphas</th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF4 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{an}$</td>
<td>0.332</td>
<td>0.207 (0.0250)</td>
<td>0.406 (0.0566)</td>
<td>0.394 (0.0576)</td>
</tr>
<tr>
<td>$\rho_{ij}^{in}$</td>
<td>0.248 (0.0195)</td>
<td>0.193 (0.0179)</td>
<td>0.303 (0.0329)</td>
<td>0.314 (0.0332)</td>
</tr>
<tr>
<td>gics</td>
<td>0.174 (0.0142)</td>
<td>0.106 (0.0209)</td>
<td>0.219 (0.0148)</td>
<td>0.212 (0.0135)</td>
</tr>
<tr>
<td>Adj. R sq.</td>
<td>0.304</td>
<td>0.341</td>
<td>0.378</td>
<td>0.395</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF4 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{an}$</td>
<td>0.236</td>
<td>0.236</td>
</tr>
<tr>
<td>$\rho_{ij}^{in}$</td>
<td>0.247</td>
<td>0.247</td>
</tr>
<tr>
<td>gics</td>
<td>0.134</td>
<td>0.203</td>
</tr>
<tr>
<td>Adj. R sq.</td>
<td>0.427</td>
<td>0.570</td>
</tr>
</tbody>
</table>
Table 6: Correlated Analyst Coverage Before and After Additions to the SP500

This table reports the average correlated analyst coverage measure for stocks before and after being added to the S&P500 index, plus the percentage change and difference in these means, and the t-statistic and p-value for these differences. The means are calculated for the 1st quarter before and after the event through the 12th quarter before and after. 386 stocks are used from 1982 through 2007.

<table>
<thead>
<tr>
<th>Qtr</th>
<th>Before</th>
<th>After</th>
<th>%Change</th>
<th>Diff</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0098</td>
<td>0.0115</td>
<td>17.3</td>
<td>0.0017</td>
<td>3.39</td>
<td>&lt;0.0007</td>
</tr>
<tr>
<td>2</td>
<td>0.0095</td>
<td>0.0120</td>
<td>26.3</td>
<td>0.0025</td>
<td>4.64</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3</td>
<td>0.0088</td>
<td>0.0121</td>
<td>37.5</td>
<td>0.0033</td>
<td>6.27</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>4</td>
<td>0.0081</td>
<td>0.0125</td>
<td>54.3</td>
<td>0.0044</td>
<td>8.58</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>5</td>
<td>0.0081</td>
<td>0.0124</td>
<td>53.1</td>
<td>0.0042</td>
<td>8.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6</td>
<td>0.0079</td>
<td>0.0126</td>
<td>59.5</td>
<td>0.0047</td>
<td>8.70</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>7</td>
<td>0.0074</td>
<td>0.0125</td>
<td>68.9</td>
<td>0.0052</td>
<td>9.54</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>8</td>
<td>0.0068</td>
<td>0.0123</td>
<td>80.9</td>
<td>0.0055</td>
<td>10.82</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>9</td>
<td>0.0062</td>
<td>0.0126</td>
<td>103.2</td>
<td>0.0064</td>
<td>12.13</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>10</td>
<td>0.0064</td>
<td>0.0125</td>
<td>95.3</td>
<td>0.0061</td>
<td>11.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>11</td>
<td>0.0062</td>
<td>0.0124</td>
<td>100.0</td>
<td>0.0062</td>
<td>11.30</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>12</td>
<td>0.0052</td>
<td>0.0125</td>
<td>140.4</td>
<td>0.0073</td>
<td>13.95</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table 7: Excess Comovement: Additions to the SP 500 Index

This table reports OLS regression coefficients and t-statistics from regressing the change in stock return correlations between stocks that are added to the S&P500 index and all other stocks in the index on changes in correlated analyst coverage, $\Delta \rho_{ij}^{oa}$ before and after the addition date. Year and firm fixed effects are included where indicated. The correlated analyst coverage variable is calculated using the I/B/E/S database and equation (5). The correlation coefficients are estimated using daily raw returns (Panel A), or residuals from the CAPM (Panel B), the Fama French 3 factor model (Panel C) and a model using the Fama French factors plus the Carhart momentum factor (Panel D). Returns data are from CRSP and factor data from WRDS. 136,291 observations are used. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \rho_{ij}^{oa}$</td>
<td>0.056 0.0543 0.0602</td>
<td>0.0492 0.0473 0.0593</td>
<td>0.0551 0.0538 0.0566</td>
<td>0.0515 0.0510 0.0530</td>
</tr>
<tr>
<td></td>
<td>(0.0176)(0.0151)(0.0125)</td>
<td>(0.0129)(0.0131)(0.0152)</td>
<td>(0.0120)(0.0120)(0.0120)</td>
<td>(0.0120)(0.0119)(0.0120)</td>
</tr>
<tr>
<td>Year FE</td>
<td>N Y N</td>
<td>N Y N</td>
<td>N Y N</td>
<td>N Y N</td>
</tr>
<tr>
<td>Firm FE</td>
<td>N N Y</td>
<td>N N Y</td>
<td>N N Y</td>
<td>N N Y</td>
</tr>
<tr>
<td>Adj. R Sq.</td>
<td>0.0001 0.2817 0.5095</td>
<td>0.0001 0.0084 0.0554</td>
<td>0.0002 0.0018 0.0162</td>
<td>0.0001 0.0016 0.0132</td>
</tr>
</tbody>
</table>

42
Table 8: Excess Comovement: Brokerage House Mergers

This table reports summary statistics (Panel A) and OLS regression coefficients and standard errors from regressing the change in stock return correlations between stocks on changes in correlated analyst coverage, $\Delta \rho_{ij}^{an}$, correlated mutual fund holdings, $\Delta \rho_{ij}^{mf}$, and correlated institutional holdings, $\Delta \rho_{ij}^{in}$, before and after 14 brokerage house mergers between 1994 and 2005 (see Hong and Kacperczyk, 2010, Table III for a list). Merger fixed effects are included, but not reported. The correlated analyst coverage variable is calculated using the I/B/E/S database and equation (5). The correlation coefficients are estimated using daily raw returns (Panel B), or residuals from the CAPM (Panel C), the Fama French 3 factor model (Panel D) and a model using the Fama French factors plus the Carhart momentum factor (Panel E). Returns data are from CRSP and factor data from WRDS. Mutual fund and institutional holdings data are from the Thompson-Reuters databases as described in the data section. 387,181 observations are used. Standard errors are in parentheses.

### Panel A: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Change in</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \rho_{ij}^{an}$</td>
<td>Correlated Analyst Coverage</td>
<td>387181</td>
<td>0.021</td>
<td>0.079</td>
<td>-0.474</td>
<td>0.939</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{mf}$</td>
<td>Correlated Mutual Fund Holdings</td>
<td>387181</td>
<td>0.017</td>
<td>0.058</td>
<td>-0.250</td>
<td>0.539</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{in}$</td>
<td>Correlated Institutional Holdings</td>
<td>387181</td>
<td>0.021</td>
<td>0.109</td>
<td>-0.617</td>
<td>0.821</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{rt}$</td>
<td>Raw Return Correlations</td>
<td>387181</td>
<td>0.028</td>
<td>0.137</td>
<td>-0.624</td>
<td>0.611</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{capm}$</td>
<td>CAPM Alpha Correlations</td>
<td>387181</td>
<td>-0.021</td>
<td>0.130</td>
<td>-0.639</td>
<td>0.730</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{ff3}$</td>
<td>FF3 Alpha Correlations</td>
<td>387181</td>
<td>-0.035</td>
<td>0.139</td>
<td>-0.695</td>
<td>0.748</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{ff4}$</td>
<td>FF4 Alpha Correlations</td>
<td>387181</td>
<td>-0.043</td>
<td>0.156</td>
<td>-0.910</td>
<td>0.745</td>
</tr>
</tbody>
</table>

### Panel B: Raw Rets

<table>
<thead>
<tr>
<th>Variable</th>
<th>Change in</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \rho_{ij}^{an}$</td>
<td>Correlated Analyst Coverage</td>
<td>387181</td>
<td>0.024</td>
<td>0.029</td>
<td>0.029</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0027)</td>
<td>(0.0066)</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{mf}$</td>
<td>Correlated Mutual Fund Holdings</td>
<td>387181</td>
<td>-0.001</td>
<td>0.013</td>
<td>0.013</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0035)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{in}$</td>
<td>Correlated Institutional Holdings</td>
<td>387181</td>
<td>-0.026</td>
<td>-0.026</td>
<td>-0.032</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0018)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Adj R Sq</td>
<td></td>
<td>0.196</td>
<td>0.196</td>
<td>0.196</td>
<td>0.030</td>
<td>0.030</td>
</tr>
</tbody>
</table>

### Panel C: CAPM alphas

<table>
<thead>
<tr>
<th>Variable</th>
<th>Change in</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \rho_{ij}^{an}$</td>
<td>Correlated Analyst Coverage</td>
<td>387181</td>
<td>0.046</td>
<td>0.046</td>
<td>0.050</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0028)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{mf}$</td>
<td>Correlated Mutual Fund Holdings</td>
<td>387181</td>
<td>0.010</td>
<td>0.030</td>
<td>0.030</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0037)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td>$\Delta \rho_{ij}^{in}$</td>
<td>Correlated Institutional Holdings</td>
<td>387181</td>
<td>-0.026</td>
<td>-0.037</td>
<td>-0.037</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0019)</td>
<td>(0.01730)</td>
</tr>
<tr>
<td>Adj R Sq</td>
<td></td>
<td>0.137</td>
<td>0.137</td>
<td>0.137</td>
<td>0.161</td>
<td>0.161</td>
</tr>
</tbody>
</table>
Table 9: Excess Comovement, Unexpected Earnings and Earnings Revisions

This table reports OLS regression coefficients and standard errors from regressing stock return correlations between pairs of stocks on combinations of correlated analyst coverage, $\rho_{ij}^{an}$, correlated unexpected earnings, $\rho_{ij}^{ue}$, and correlated earnings forecast revisions, $\rho_{ij}^{ef}$, along with an industry indicator variable and year fixed effects. Correlated analyst coverage is defined as $N_{ij}/\sqrt{N_i N_j}$, where $N_{ij}$ is the number of common analysts for stocks $i$ and $j$ and $N_i$ and $N_j$ are the total number of analysts covering stocks $i$ and $j$, respectively. Correlated unexpected earnings is the correlation coefficient of “unexpected earnings” between stocks $i$ and $j$ estimated over the previous 5-year period using quarterly data. Unexpected earnings is defined as the change in quarterly earnings per share in a one-year period ($ep_{i,t} - ep_{i,t-4}$). Correlated earnings forecast revisions is the monthly correlation coefficient for consensus annual earnings forecasts between stocks $i$ and $j$ and is estimated using the previous 20 quarters worth of data. The variable gics is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. Year fixed effects are included but not reported. The correlation coefficients are estimated yearly using daily raw returns (Panel A), or the daily residuals from the CAPM (Panel B), the Fama French 3-factor model (Panel C) and the Fama French 3-factor model augmented with the Carhart momentum factor (Panel D) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP and factor data from WRDS. Analyst data along with earnings and earnings forecast data are from I/B/E/S. Standard errors are clustered at the firm and year level and are listed in parentheses. 438,651 observations are used for the regressions involving unexpected earnings, and 438,651 for the regressions involving earnings revisions. Both samples cover the period 1995–2007.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Raw Rets</th>
<th>Panel B: CAPM alphas</th>
<th>Panel C: FF3 alphas</th>
<th>Panel D: FF3 alphas</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{an}$</td>
<td>0.240</td>
<td>0.236</td>
<td>0.237</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>(0.0344)</td>
<td>(0.0342)</td>
<td>(0.0353)</td>
<td>(0.0353)</td>
</tr>
<tr>
<td>$\rho_{ij}^{ue}$</td>
<td>0.0242</td>
<td>0.0161</td>
<td>0.0129</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0035)</td>
<td>(0.0027)</td>
<td>(0.0029)</td>
</tr>
<tr>
<td>$\rho_{ij}^{ef}$</td>
<td></td>
<td>0.0129</td>
<td>0.0112</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0027)</td>
<td>(0.0026)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>gics</td>
<td>0.0728</td>
<td>0.0714</td>
<td>0.0735</td>
<td>0.0730</td>
</tr>
<tr>
<td></td>
<td>(0.0204)</td>
<td>(0.0201)</td>
<td>(0.0209)</td>
<td>(0.0210)</td>
</tr>
<tr>
<td>N obs</td>
<td>436,139</td>
<td>436,139</td>
<td>438,651</td>
<td>438,651</td>
</tr>
<tr>
<td>Adj. R sq.</td>
<td>0.380</td>
<td>0.384</td>
<td>0.380</td>
<td>0.382</td>
</tr>
</tbody>
</table>

44
Table 10: Regulation Fair Disclosure - Comovement and Analyst Coverage

This table reports OLS regression coefficients and t-statistics from regressing stock return correlations between pairs of stocks on correlated analyst coverage, $\rho_{ij}^{cn}$, along with the variable, REGFD, which is equal to 1 if the observation is from after Regulation Fair Disclosure was instituted in 2000 and a zero otherwise. Additionally, REGFD is interacted with the correlated analyst coverage variable. Correlated analyst coverage is defined as $N_{ij}/\sqrt{N_i N_j}$, where $N_{ij}$ is the number of common analysts for stocks $i$ and $j$ and $N_i$ and $N_j$ are the total number of analysts covering stocks $i$ and $j$, respectively. The variable gics is an indicator variable that is equal to 1 if the pair of stocks share the same 6-digit Global Industry Classification Standard (GICS) code and zero otherwise. The correlation coefficients are estimated yearly using daily raw returns (column 1), or the daily residuals from the CAPM (column 2), the Fama French 3-factor model (column 3) and the Fama French 3-factor model augmented with the Carhart momentum factor (column 4) for each pair of S&P 500 stocks that are in the index during the calendar and the previous year. Factor loadings for the asset pricing model are estimated using data from the previous year. Returns data are from CRSP, factor data from WRDS, and analyst data are from I/B/E/S. 922,294 observations are used from 1995 through 1999 and 2001 through 2007. Observations from 2000 are omitted. Standard errors are clustered by firm and year and are reported in parentheses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(Raw Ret)</th>
<th>(CAPM)</th>
<th>(FF3)</th>
<th>(FF+Carhart)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{ij}^{cn}$</td>
<td>0.0522</td>
<td>0.301</td>
<td>0.302</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td>(0.0882)</td>
<td>(0.0761)</td>
<td>(0.0866)</td>
<td>(0.1216)</td>
</tr>
<tr>
<td>REGFD</td>
<td>0.0997</td>
<td>0.0366</td>
<td>0.0274</td>
<td>0.0323</td>
</tr>
<tr>
<td></td>
<td>(0.0363)</td>
<td>(0.0367)</td>
<td>(0.0424)</td>
<td>(0.0601)</td>
</tr>
<tr>
<td>$\rho_{ij}^{cn} \times$ REGFD</td>
<td>0.107</td>
<td>0.329</td>
<td>0.376</td>
<td>0.532</td>
</tr>
<tr>
<td></td>
<td>(0.1015)</td>
<td>(0.1242)</td>
<td>(0.1383)</td>
<td>(0.2231)</td>
</tr>
<tr>
<td>GICS</td>
<td>0.134</td>
<td>0.0534</td>
<td>0.0386</td>
<td>-0.0158</td>
</tr>
<tr>
<td></td>
<td>(0.0368)</td>
<td>(0.0402)</td>
<td>(0.0450)</td>
<td>(0.0837)</td>
</tr>
<tr>
<td>Adj. R Sq</td>
<td>0.158</td>
<td>0.113</td>
<td>0.101</td>
<td>0.100</td>
</tr>
</tbody>
</table>