The Correlation Structure of the CDS Market: An Empirical Investigation

LARA CATHCART, LINA EL-JAHEL AND LEO EVANS *

July 30, 2010

ABSTRACT

Using an extensive data set of CDS spreads on eminent U.S. firms, we investigate the correlation structure of the CDS market. For comparative purposes, we also examine the correlation structure of their equity returns. We find that whilst industry affiliation plays a central role in CDS correlations, so too does rating classification above and below investment grade. By contrast, the correlation structure of equity returns is characterized by industry classification. Our results highlight differences in the organization of these markets, the salience of the investment-grade boundary and provide evidence of friction-based co-movement in credit markets.

JEL CLASSIFICATION CODES: G10, G24, C20.

KEY WORDS: Credit Default Swaps, Credit Correlation, Cluster Analysis, Principle Components Analysis, Investment-Grade Boundary, Friction-based Co-movement.

*All errors are ours. Correspondence should be addressed to L.Cathcart or L.El-Jahel or L.Evans at Imperial College Business School, South Kensington Campus, London SW7 2AZ, United Kingdom. Emails: l.cathcart@imperial.ac.uk, l.el-jahel@imperial.ac.uk, leonard.evans03@imperial.ac.uk.

One of the most prominent divides of classification in capital markets is that of the investment-grade ratings boundary; separating investment-grade from high-yield rated issuer based on a subjective assessment of credit risk made by credit rating agencies. It is often cited in regulation[1] and many institutional investors are restricted from holding high-yield debt[2]. The boundary is also referenced in many financial contracts; sell-side firms utilize the distinction when acting as dealers in OTC derivatives markets to determine acceptable counterparties and collateral levels for outstanding credit exposures. Although not subject to regulatory restrictions, bond mutual funds typically specialize in either investment-grade or high-yield debt and at the very least reference the boundary with respect to permissible asset concentrations[3].

In economies with frictions, irrational investors and limits to arbitrage, Barberis and Shleifer (2003) and Barberis, Shleifer and Wurgler (2005) demonstrate that a class of friction and sentiment-based co-movement theories arise. 'Style' investing is one such theory and somewhat behavioral in nature. The basis of which is that one of the clearest mechanisms of human thought is classification; we do so in order to simplify problems of choice and the ability to process vast quantities of information. Style investors first categorize risky assets (market capitalization, sector, rating etc.) and then allocate funds at the category level. Habitat investors, a related concept, can only trade in a particular subset of securities due to the presence of trading restrictions. Crucially, in these theories common factors exist in the returns of assets that happen to be classified in the same category, even if the assets' cash flows are themselves uncorrelated.

Given the importance of the investment-grade boundary, we look for evidence of common factors related to investment-grade or high-yield rating classification in Credit Default Swap (CDS) spreads[4], as suggested by the theories outlined above. A CDS is a bilateral contract which transfers a defined credit risk between counter-parties. The protection buyer pays a quarterly premium to the protection seller and receives a settlement equal to the difference between the par and market

[1]The Joint Forum (2009) report, 'Stocktaking on the use of credit ratings', details member authorities' usage of credit ratings in legislation, regulations and supervisory policies.

[2]Notably, commercial banks and Savings & Loans have been prohibited from doing so since 1936 and 1989 respectively.

[3]According to the Vanguard High-Yield Corporate Fund Prospectus (May 28, 2010), the fund invests "at least 80% of its assets in corporate bonds that are rated below Baa3 by Moody's Investors Service, Inc."

[4]CDS spreads are arbitrage bound to bond prices.

value of the underlying bond should a default event occur. We investigate whether the investment-grade boundary has a discernible influence on the correlation structure of the U.S. five-year CDS market and make a comparative analysis to the correlation structure of the equity market, where no comparable friction exists. Our data set consists of the CDS and equity prices of 145 eminent U.S. firms between February 2, 2004 to August 29, 2008.

The correlation structure of the equity market has been documented previously. For example, early work by King (1966) and Meyers (1973) demonstrates that a large group of individual series of stock returns can be separated into smaller sets of correlation clusters related by industry classification. To the best of our knowledge, the CDS market has not been examined from this perspective. However previous authors[5] have suggested that the CDS market is segmented by industry. Through principle components analysis, we examine this issue. We find that the correlation structures of CDS and equity returns are both characterized by respective market factors[6] and also factors which effect specific subsets of firms, such as industries or groups of industries. However differences in their correlation structures are apparent; correlations seem less closely related to industry in the CDS market and we find evidence of common factors related to oil prices influencing equity and not CDS correlations.

We then utilize an agglomerative hierarchical cluster analysis, the Complete Linkage method, to identify correlation clusters in CDS and equity returns. Our motivation for doing so is that if common factors related to rating classification exist in spreads, and make an important contribution relative to market and industry factors, a segmentation of CDS correlations across the boundary will be observed. We use the Global Industry Classification Standard (GICS) as a benchmark for the expected structures of the clusters; industry groups of firms homogenized due to common factors in their security prices. The clusters we find demonstrate two patterns in addition to industry influences. Firstly, correlation clustering of high-yield rated issuers with no industry links and secondly a decoupling of correlations within an industry, across the boundary. By contrast, in the equity market the clustering is entirely explained by industry; there is no segmentation by

[5]For instance Daniels and Jensen (2005) and Longstaff and Rajan (2008).
[6]That are significantly negatively correlated.

rating. These results provide empirical evidence of common factors in CDS spreads related to the boundary; in line with friction and sentiment-based theories of co-movement.

To understand better the relative contributions of market, industry and rating factors on correlations, we then remove from the data set all firms which have crossed the investment-grade boundary during the sample period. As shown by Barberis et al. (2005), the sensitivity of an asset to factors pertaining to its classification alter when the classification changes. Subsequently, we control for the influence of market factors on return correlations. We do so through firm-by-firm time-series regressions of CDS returns on a CDS index and equity returns on the S&P500, in a similar fashion to existing studies in the equity literature (Livingston (1977)). By repeating the cluster analysis on the residuals of these regressions, we show that mean residual CDS correlation coefficients of high-yield clusters with no industry links are of a similar magnitude to that of investment-grade industry affiliates. When clustered high-yield firms share industry affiliation, mean residual correlations approximately double, suggesting that these high-yield factors are as important as, if not more than, industry factors in the spreads of the issuers we analyze. We also find that the market and industry factors play a similar role in CDS and equity returns correlations in this period for well-clustered industries; mean residual correlations are typically half that of mean return correlations for both sets of securities.

In our final analysis we compare industry factors between the CDS and equity returns of the GICS Energy Sector because of the importance of industry factors on the correlation structures of both markets. We focus on this sector for two reasons. Firstly, the results of our PCA provide an observable proxy for a key factor in their equity returns, the return on spot oil prices. Secondly, the cluster analysis identifies the Energy Sector as distinct in the equity correlation matrix; common factors related to industry are particularly important. We find that common dependence on oil prices explains the extent to which their equity returns cluster. This does not hold for their CDS returns, where we conclude the common industry factors are more likely latent in nature. Distinct industry factors across securities relating to the same firms furthers our explanation of why their correlation structures differ.

This study contributes to the literature in a number of ways. Firstly we provide, to the best of our knowledge, the most comprehensive empirical analysis of the correlation structure of the U.S. five-year CDS market. Secondly, we provide an in-depth comparative analysis to the correlation structure of the equity market. Thirdly, we add to the growing empirical literature that demonstrates the importance of the ratings boundary in financial markets. Recent research has shown that the distinction between investment-grade and high-yield affects a firm's cost of capital (Kisgen and Strahan (2009)), debt-issuance and investment (Chernenko and Sunderam (2009)). Finally, we provide the first evidence of friction-based co-movement in credit markets. The rest of this paper is organized as follows. Section I describes our data. Section II conducts a principle components analysis and presents evidence on what drives correlations in the CDS and equity markets. Section III examines the correlation structures of CDS and equity returns using cluster analysis. Section IV concludes.

I. The Data Set and Descriptive Statistics

Our data set consists of end-of-day composite mid-market CDS spreads and common stock closing prices obtained from Thompson Reuters Datastream, Global Industry Classification Standard (GICS) codes obtained from COMPUSTAT and issuer-level estimated senior unsecured historical ratings from the Default Risk Service (DRS) corporate database of Moody's Investors Service.

A. The CDS and Equity Data

Like previous studies in the CDS literature, we focus on five-year quotes which are the benchmark maturity. The CDS data set is provided by CMA DataVision^{TM} and originally sourced from 30 buy side firms including investment banks, hedge funds and asset managers who collectively form the CMA Data Consortium. According to CMA, contributed quotes are validated, filtered and aggregated using proprietary software. For the publication of prices at 5pm New York time, a minimum of three consortium members seeing quotes from three distinct sell side sources must contribute. When there are insufficient observed values to build an entire term structure, CMA fits a proprietary term structure model to the available data to generate the rest of the curve. All premia are expressed as an annualized spread in basis points on U.S. dollar denominated notional amounts and all reference obligations are senior unsecured debt. We retain CMA's derived prices rather than linearly interpolate or hold prices constant between observed quotes[7]. As a further control of data quality, we impose a restriction that limits the maximum number of consecutive trading days of zero spread change to 20.

The existing CDS literature focusses mainly on spread changes, however in using this specification, issuers trading at higher spreads will by construction have more volatile spreads than issuers trading at lower spreads. To circumvent the statistical problems associated with this, we consider log-percentage changes in CDS spreads. We will refer to these as CDS returns, although this definition does not correspond directly to the dollar daily return on a CDS position. Our CDS data are in the form of at-market spreads for newly-issued contracts and the calculation of actual returns requires a pricing model. There is the well-known distinction between those that belong to either the reduced-form or structural-model specifications, yet little academic consensus as to the best method for pricing CDS. We remain agnostic to model specification because our interest lies in relative price dynamics. Nonetheless, as shown by Micu, Remolona and Wooldridge (2006), under reasonable assumptions our specification will provide an approximation to actual returns; albeit an overestimation. The raw equity data set matches that of the CDS in terms of frequency, coverage and time span and returns are calculated in the normal manner, facilitating a direct comparison between the correlation structures of the CDS and equity markets.

B. Industry Classifications

The most widely available industry classification schemes are the Standard Industry Classification (SIC) codes, the North American Industry Classification System (NAICS), the Fama and French (1997) algorithm and the Global Industry Classification Standard (GICS). Bhojraj, Lee and Oler

[7]In our final sample of 145 issuers, the mean (median) percentage of quoted to total spreads is 81% (84%)

(2003) compare the four schemes in a variety of applications common to capital market research, such as their ability to explain cross-sectional variation in stock returns as well as market-based valuation multiples. They reach the conclusion that the GICS is significantly better than the others in almost all applications and argue that it should be more widely adopted by academic researchers. For this reason, the GICS is used in this study[8].

The GICS was developed jointly by Standard & Poors (S&P) and Morgan Stanley Capital International and introduced in 1999 to 'establish a global standard for categorizing companies into sectors and industries'. It is an eight-digit, four-tiered (two digits per tier), hierarchical classification code and currently consists of 10 sectors, 24 industry groups, 67 industries and 147 sub-industries[9]. Due to the hierarchical nature of the GICS, we must stipulate from which level affiliations between firms will be drawn. We follow Daniels and Jensen (2005) in so far as our ex-ante belief is that the CDS market is segmented by industry type. For the remainder of this study, firms are classified by their six-digit industry-level GICS codes.

C. Rating Classifications

Credit rating agencies (CRAs) provide two important functions for market participants. The first is a subjective assessment of a given issuer's ability to meet its financial obligations in the form of a credit rating. In theory, ratings should mitigate agency problems between borrowers and lenders by fulfilling quality revelation and certification functions. The second, through the usage of standardized risk categories, is a means of comparing the relative creditworthiness of issuers. The largest international rating agencies are Moody's and S&P[10] and there is a well-known correspondence between their systems; the key division occurs between Baa3 and Ba1 in Moody's ratings and BBB$-$ and BB$+$ for S&P's, which is known as the investment-grade boundary. We use Moody's

[8] We also implemented our methodology using the NAICS. Whilst our results are not sensitive to the choice of classification scheme, we consistently found a better correspondence between correlation clusters and the GICS in both our CDS and equity data.

[9] According to the GICS guidebook (2006), companies are classified on the basis of their principle business activity. The S&P and MSCI analysts are guided by information from financial statements as well as market perception as revealed through investment research reports.

[10] There are currently a total of 10 rating agencies licensed by the Securities and Exchange Commission.

historical issuer-level senior unsecured ratings[11] to classify investment-grade and high-yield rated firms as well as those that have either been upgraded or downgraded across the boundary over our sample period. For the subsequent analysis, our usage of the term "rating classification" refers only to the distinction made between issuers separated by the boundary, not the finer categories of Moody's twenty-one grade alphanumeric rating scale.

D. Descriptive Statistics

The initial data set covers the period February 2, 2004 to December 31, 2008 giving 1283 daily observations on a total of 262 continuously active corporate issuers based in the U.S. We further impose that an industry contains at least seven constituent firms[12], such that our final sample consists of a balanced panel of the daily CDS and equity returns of 145 firms. Appendix A lists them and their distribution across the 14 GICS industries in the data set. These are the Energy Equipment & Services, Oil, Gas & Consumable Fuels, Chemicals, Road & Road, Auto Components, Household Durables, Hotels, Restaurants & Leisure, Speciality Retail, Food & Staples Retail, Food Products, Healthcare Providers, Pharmaceuticals, Insurance and Multi-Utilities industries. As for rating classifications, in total there are 100 firms that have been classified as strictly investment-grade (IG) throughout the sample period, 21 as high-yield (HY), 17 have traversed the investment-grade boundary from above (IG/HY), five from below (HY/IG) and two that were promoted to IG status and subsequently downgraded again (HY/IG/HY).

Table I presents summary statistics for the equity price, CDS spread and corresponding log-return data by year. The average daily CDS spread in the full sample is 124bps with a standard deviation of 282bps, but there is considerable variation across years reflecting changing credit conditions. In 2004 and 2006, we observe relatively low CDS return volatility and negative average CDS returns (Panel C), i.e. tightening spreads as we are considering returns from the perspective

[11]Four firms did not have ratings available in the DRS. We supplemented their ratings data with S&P's domestic long term issuer credit rating obtained from COMPUSTAT.

[12]If we lower or remove this threshold, the conclusions of this paper do not change. Indeed, we would have more high-yield rated firms and clusters in our data. However, the presentation of the clusters in the expanded data set would have been unnecessarily bulky.

of the protection buyer. The volatility of CDS returns is seen to be relatively high in 2005, which is unsurprising given that credit markets were characterized by the downgrades of Ford and GM in May of that year. More recently, credit markets have been distorted by the onset of the subprime crisis in 2007 and the ensuing global financial crisis of late 2008 which led to an economy-wide repricing of credit risk. This is observed in Panels A & C by the dramatic increase in mean spread levels to 291bps, a standard deviation of 548bps and positive average CDS returns, even in the absence of defaulted issuers. The statistics of the equity market reflect a different story. In Panel D, equity return volatility is seen to trend upwards and mean stock prices (Panel B) are only observed to fall in 2008. The equity market seems, at least at an aggregate level, to have been less disrupted than the credit markets in both 2005 and 2007.

For the proceeding analysis, we consider the period February 3, 2004 to August 29, 2008, yielding a total of 173,130 firm-day observations per security class. The sample is truncated to pre-September 2008 due to the extraordinary events of that month[13] and the disfunction of financial markets, particularly the CDS market, during that time.

II. Principle Components Analysis

We begin by examining the correlation structure of the CDS market by use of a principal components analysis (PCA). The idea is to reduce the dimensionality of the data by looking for linear combinations of the firm level CDS returns that can explain the maximum variance. These Principal Components (PCs) are orthogonal by construction. Subsequently, we regress the principal component weights (normalized eigenvectors) on 14 industry dummy variables. If the correlation structure of the CDS market has nothing to do with industry, then the dummies should have no cross-sectional explanatory power for the principal components. For comparative purposes, we do

[13]Lehman Brothers filed for Chapter 11 bankruptcy on September 15, 2008. American International Group, Merrill Lynch, Fannie Mae, and Freddie Mac also failed or were placed under conservatorship by the U.S. government.

the same for the equity returns of these firms[14]. Table II reports summary results from the PC decomposition of the correlation matrices as well as the results from the cross-sectional regressions.

From the PC decomposition we observe that credit risk, like the risks inherent in equities, is not purely idiosyncratic as the first PC explains 18.5% of the variation in CDS returns and 29.2% in equity returns. Yet idiosyncratic risk would appear to be the dominant influence on both, with the first 10 PCs explaining 40% and 60% respectively. This differential suggests the existence of a relatively larger idiosyncratic component to individual firms' CDS spreads, which would help to explain why correlations in CDS returns are lower than equity returns in general; idiosyncratic shocks on spreads and share prices will not induce cross-sectional correlation. The sample mean CDS return correlation in our sample is 20.0% and that of equity returns is 26.5%.

With reference to the cross-sectional regressions, in both markets, all 14 dummy variables are positive and highly significant with respect to the first PC, so to interpret it, we create an index by taking a weighted average of the 145 firms' CDS and equity returns, using the normalized principal component weights. For future reference, we call this the Principle Component Index (PCI). We also introduce a CDS market return index by constructing an equally-weighted average of all 262 firms' CDS spreads in our data set and then taking log-differences to calculate returns. The correlation of the first PCI of CDS returns is 81.7% with CDS index returns and thus resembles a pervasive component to credit risk, consistent with Das, Freed, Geng and Kapadia (2006), Elizalde (2005) and Longstaff and Rajan (2008). In a similar fashion, the correlation of the first PCI of equity returns is 89.5% with S&P500 returns and is therefore related to the market in a traditional sense. The correlation of the first PCIs of CDS and equity returns is -37.7%, suggesting that the markets factors to which these components relate are not distinct across these securities.

Beyond the first PC, there is a significant relationship between all of the remaining PCs and either individual or groups of industries. However, the CDS return PCs are linked to a narrower subset of industries than those of the equity returns, as many load on only one or two industries. The R^2 of the cross-sectional regressions are also typically higher in the equity results suggesting

[14] King (1966) identifies that stock prices co-vary by industry beyond the effects of aggregate market behaviour and Roll (1988) shows that a significant portion of stock return variation can not be attributed to overall market and industry movements, suggesting that firm-specific information is impounded into prices.

that correlation dynamics are more closely related to industry in the equity market, a finding consistent with Longstaff and Rajan (2008). Of particular interest is the second PC of equity returns, which is best explained by the industry dummies across all principle components and has an adjusted R^2 of 81.3%. It explains approximately 10% of the variation in equity returns. In contrast, that of CDS returns explains 4% and the R^2 of the cross-sectional regression is much lower. Closer inspection of Table II shows that the second PC in equity returns loads heavily and positively on the Energy Sector, specifically the Energy Equipment & Services and Oil, Gas & Consumable Fuels industries and to a lesser extent the Chemicals and Multi-Utilities industries. It loads negatively, but to a lesser extent, on the Auto Components, Household Durables, Hotels, Restaurants & Leisure and Insurance industries[15]. We check the correlation of the second PCI of equity returns with the return on West Texas Intermediate[16] (WTI) crude oil, which is often used as a benchmark in oil pricing. They exhibit a sample correlation of 55.5% which is both highly economically and statistically significant. Clearly the risk captured by the second principle component shares a link with the risks posed by fluctuations in oil prices, but is not market-wide in so far as its interaction is asymmetric across industries and exerts no influence on some, consistent with the interpretation of a sectoral component[17]. To test whether a component of this type manifests itself in the comovement of CDS returns, we check the correlation of the first 10 CDS PCIs with WTI returns, but find no significance at conventional levels.

These results indicate that there are empirical similarities in the correlation structures of the two markets. Firstly, they are both characterized by the influence of market factors that are related across these markets. Secondly, we find evidence of factors which effect specific subsets of firms such as industries or groups of industries in both. Yet there is also evidence of important differences; correlation dynamics of CDS returns seem less closely related to industry and oil prices appear to play a role in equity return correlations. We turn to the methodology of cluster analy-

[15] Although they do not explicitly state it, Longstaff and Rajan (2008) also find a strong positive relationship between their second PC of equity returns and the energy and utilities sectors.

[16] Also known as Texas Light Sweet. Obtained from Datastream.

[17] Roll (1988) suggests that higher order APT factors, created via factor analysis, are especially important for oil and utility firms and conclude that these industries have characteristics unlike those of a typical stock in a broad market index. Many of the results in this paper lead us to similar conclusions.

sis to further explore the correlation structures of these markets and understand better how such differences materialize.

III. Correlation Cluster Analysis

The results of the PCA highlight drivers of CDS correlations and provide indicative evidence of common factors related to industry classification. In this section we explore the existence of common factors in CDS returns linked specifically to issuers' classification as either investment-grade or high-yield. If these factors do exist and are prevalent, one might expect a segmentation effect on CDS markets. However, ex-ante it is not clear if one can observe segmentation beyond industry, which has been established as central to both the CDS and equity correlations. It will depend on the relative contribution of these factors to spreads. Cluster analysis identifies homogeneous groups of firms with respect to correlation in their security returns; an ideal tool to examine both the underlying structure of these markets and roots of correlation segmentation.

A. Method

The three key considerations faced when performing a cluster analysis require elaboration. The first is how to define similarity between returns across firms. For this analysis we consider the sample product moment correlation matrix. The second centers on the choice of clustering method. We implement an agglomerative hierarchical clustering procedure, as detailed in Appendix B, which defines a hierarchy of nested clusters, beginning with each firm representing one cluster and ending with a single cluster containing all firms. Specifically we choose the Complete Linkage algorithm[18] from this class of methods. The algorithm identifies from the correlation matrix the two clusters that have the highest correlation, labeled p and q, and merges them. We denote the

[18] Our method shares some similarities with that of King (1966). His study looks at monthly stock returns from firms continuously listed on the NYSE over the period June, 1927 to December 1960 and firms are classified into six two-digit SEC industries as of the 1961 classifications.

resulting cluster r. The correlation (ρ_{rt}) between r and any other cluster t, in Complete Linkage, is defined as the minimum correlation between firms in r and t:

$$\rho_{rt} = \min(\rho_{pt}, \rho_{qr}) \tag{1}$$

In this framework, clusters share the property that all firms within a cluster will exhibit correlations greater in magnitude than that of its two least correlated firms. Complete Linkage produces clusters where the distribution of firms across them is balanced; a property which facilitates studying the structure present in a correlation matrix. It is also less susceptible to outliers than other hierarchical methods[19]. The final consideration, in general, is the number of clusters present in a correlation matrix; nontrivial given that Complete Linkage defines an entire hierarchy of nested clusters and that there is no global optimization of an objective function. Instead, at each level in the hierarchy, it uses a specific criterion, Equation 1, to decide locally which clusters should be merged. However, our interest does not lie in endogenously specifying the number of clusters, but rather how well they can represent segments of the markets that we consider. We therefore introduce two statistics that measure the fit of the clusters against industry structure as defined by the GICS, given our prior results of the prevalent role of industry in the correlation structures of these markets. The first is the F-measure, which determines the extent to which a particular cluster contains only and all firms of a given industry. It ranges from zero (no correspondence) to one (perfect correspondence). Using this, for each level of the hierarchy, a measure of goodness-of-fit of the distribution of firms across clusters to the distribution of firms across the industry-level GICS codes can be calculated[20]. The second statistic, the Phi-measure, is taken as the maximum goodness-of-fit across all levels and provides the final set of clusters we study. See Appendix C for technical details.

[19] In results not reported, we have extensively experimented with Average Linkage but the findings of this paper are not significantly altered.

[20] Again on a zero to one basis, calculated from weighted average F-Measures.

B. Correlation Clusters in CDS and Equity Returns

Table III presents the correlation clusters found in the CDS return correlation matrix and Table IV presents that of equity returns. The clusters contained in these tables have been ordered by number of constituents, there is no other significance to this ordering. Table V reports the F-measures and Phi-measures used to define these sets of clusters.

Focusing on the overall picture presented in Tables III & IV, we first note that there is close correspondence between industries and clusters found in the correlation matrices of returns, for both markets. By comparing Phi-measures, 67% for CDS 75% for the equity clusters as seen in Table V, the correspondence to industry classifications is notably higher in the equity market[21]. This finding is supportive of the PCA; correlation dynamics are more closely related to industry in equity returns.

By focusing on the relationships between individual clusters and industries, we gain further insight into the empirical similarities and differences in the correlation structures of these markets. Notable similarities, with reference to Table V, are that many industries form well-defined clusters in both markets as observed from their high F-measures. In particular the Chemicals (GICS 151010), Road & Rail (GICS 203040), Auto Components (252010), Household Durables (253010), Insurance (403010) and Multi-Utilities (551030) industries. Also, the clustering of firms in much of the Consumer Discretionary and Staples sectors (two-digit GICS prefixes 25 & 30) are typically misaligned with the GICS industry definitions in both markets. In particular, the Hotels, Restaurants & Leisure, Speciality Retail and Food & Staples Retailing industries. Cluster 1 in in Table III resembles a broad consumer sector consisting of firms from various related industries and many of the smaller clusters in Table IV share some consumer sector affiliations[22].

[21] In results not reported, we also considered rolling correlation matrices of 3, 6 and 12 months and by plotting the time-evolving Phi-measure, we observe that this result always holds in our data.

[22] We further investigated the breakdown of industry links between constituents of clusters 3 and 6 in Table III. Cluster 3 is largely Healthcare Equipment & Service with two exceptions: Polyone Corp. (Chemicals) and Rite Aid Corp (Food & Staples Retail). However both companies have very strong business links with the health care industry. Equally, cluster 6 is largely Chemicals and to which the Sherwin-Williams Company again shares strong links.

Conversely, there are striking differences between the two markets. Firstly, the Energy Sector, forms the most prevalent cluster in the equity correlation matrix[23] (cluster 1, Table IV). By comparison, the related firms are less clustered in the CDS correlation matrix. This finding is consistent with the PCA, which identified a strong sectoral component related to the Energy Sector in the equity market but no comparable component in the CDS market. Secondly, we observe segmentation of the Energy Sector across the investment-grade boundary as demonstrated by cluster 13 in Table III. This result helps explain why these firms are less clustered in the CDS market, but more importantly, it demonstrates that common factors related to rating classification exist and make a significant contribution to spreads relative to, and in the presence of, industry factors. This segmentation across the boundary is common to many industries in the CDS correlation clusters in Table III. For example, the HY firm in the Road & Rail industry separates from its IG members (clusters 19 and 9 respectively), the HY firm in the Healthcare industry separates from the IG members (clusters 19 and 4) and the same applies to the Utility industry (clusters 7 and 10).

With reference to cluster 7 in Table III, we also observe the clustering of high-yield firms with no industry affiliation in CDS returns. What is particularly striking is that its six constituent firms belong to four different GICS industries, each from a different GICS sector, i.e. Multi-utilities & Unregulated power, Oil, Gas & Consumable fuels, Hotels, Restaurants & Leisure and Food Products[24]. Clearly no industry argument can explain why these firms form a correlation cluster in the CDS market. Surprising too are the relatively high minimum and average correlation between them. This cluster in particular provides strong empirical support for the predictions of the sentiment and friction-based co-movement theories of Barberis and Shleifer (2003) and Barberis et al. (2005); i.e. common factors related to the high-yield labeling of these firms. As well as further helping to explain why correlation dynamics are in general more closely related to industry in equity returns, this particular result highlights that in some instances the common factors in CDS

[23]Although table V suggests that the constituent industries, Energy Equipment & Services and Oil & Gas Producers are not highly clustered with F-measures of 48% and 77% respectively, here the clustering of returns occurs at the sector level in the GICS classifications, not at the industry level which is what these statistics represent.

[24]The firms are CMS Energy Corp, El Paso Corp, MGM Mirage, Royal Caribbean Cruises Ltd, Smithfields Foods Inc and Williams Companies Inc.

returns relating to rating classification make a larger contribution to spreads than those pertaining to industry.

It is worth pointing out that our methodology maximizes the industry structure of the data and naturally biases clusters in the CDS market away from rating affiliation. The sample is also dominated by investment-grade issuers. Even so, our results confirm that there is correlation segmentation in the CDS market across the ratings boundary with two results in particular highlighting this. Firstly, we observe that issuers in a given industry divided by rating classification do not move as homogenous groups. Secondly, issuers from unrelated industries that share the high-yield label cluster together. The existence of the investment-grade boundary, as predicted by the friction-based co-movement theories, generates common factors in CDS returns. Our results indicate that such factors play a particularly important role for high-yield rated issuers. In contrast, in the equity market we have found no direct evidence of common factors relating to rating classification. Industry effects dominate the segmentation of this market.

C. The Effect of Market, Industry and Rating Factors on Correlation

The PCA demonstrated the importance of the market factors on the correlation structures of both CDS and equity markets. Our cluster analysis demonstrated the impact of the rating boundary on correlations in five-year CDS spreads in addition to segmentation by industry. In this section, in order to understand better the relative contributions of market, industry and rating factors, we attempt to isolate their influences along two dimensions. Firstly, we remove from the data set all firms that have traversed the investment-grade boundary[25] as they confound the influence of rating factors. As shown by Barberis et al. (2005), the sensitivity of an asset to factors pertaining to its classification alter with a change in classification. Of the 105 firms in this sub-sample, 88 are classified as investment-grade, 17 high-yield and the Speciality Retail, Food & Staples Retail and Auto Components industries are excluded via the minimum industry size restriction[26]. Secondly,

[25] We retain that an industry must contain at least seven firms.

[26] With the exception of Auto Components, these consumer sector industries cluster relatively poorly in our framework as mentioned previously.

we control for the market factors and so need viable proxies for them in both CDS and equity returns. In the equity literature, two methods have been employed. In line with these studies, the first possibility is to use the 'market' index generated by the PCA on our sample data. The second is to create a market index by averaging a large population of security returns. Livingston (1977) compares these alternatives applied to equity data and concludes that whilst both methods give a market factor and residuals, using a broad market index is preferable for a number of reasons. PCA is sample sensitive and hence the results are not easily generalizable. Also, tests reported on simulated data show that factor analysis does not in general uncover the 'true' market and residual industry effects that generate his data. Following this advice, we construct an investment-grade credit market return index (IGRI) from the 180 firms that have been classified as strictly IG in our full sample[27]. For the equity market factor we utilize S&P500 returns in the traditional manner. We then regress, for each firm separately, their CDS (equity) returns on the IGRI (S&P500 returns), plus a constant[28]. Finally, we repeat the cluster analysis on the correlation matrix of the residuals from these regressions. The resulting clusters can been seen in Tables VI and VII.

Having controlled for the market factors, beginning with the clustering of equity return residuals in Table VII, we observe they exhibit a near one-to-one mapping onto the GICS, with a Phi-measure of 85.6%. Again the Energy Sector[29] (cluster 1) forms a prevalent cluster and the disagreement between equity return correlations and the Consumer industry definitions remains (cluster 4). There are again no clusters with rating associations. In the equity data, for every industry with firms of both ratings classifications, high-yield and investment-grade firms cluster together and industry explanations account for the correlation structure of these residuals. Mean correlations of the residuals within the clusters are high; the Energy Sector (cluster 1) has a mean residual correlation of 51.6% across all 23 firms with a minimum linkage of 15.4%, those of the

[27]We tried various other specifications of the credit market index, including: varying the scope of firms to include high-yield issuers, considering the median instead of the mean spread level and various PCA constructed market indexes. However, the one we consider here results in the most significant decrease in correlations across and within clusters suggesting it captures more commonality across issuers.

[28]Across the 105 firms, the mean loading on S&P500 returns is 0.986 and that of the IGRI is 0.897. The mean t-stats are 20.4 and 17.1 and the mean R^2 is 25.7% and 20.1% respectively.

[29]It is noteworthy to point out, that King (1966) finds very similar results in his sample over the period 1927-1960, noting the prevalence of the clustering of the 'petroleum', rail and to a lesser extent utilities industries. Taken in conjunction with his, our results suggest that the correlation structure of the equity market has not significantly altered since the second quarter of the 20th century.

seven utility firms (cluster 6) have a mean correlation of 43.6% and are all linked by a minimum of 33.3%. Similar stories are seen for all of the other clusters.

With regards to the CDS return residual clustering in Table VI, the Phi-measure rises only slightly to 69.0% and whilst there are industry interpretations to the majority of clusters, the results are less clear cut. The most prevalent clusters are the investment-grade Insurance firms (cluster 1), the investment-grade Energy firms (cluster 2) and the investment-grade Healthcare providers (cluster 3) and for all of these industries, we note the decoupling of their high-yield firms from these clusters. There are now at least four high-yield clusters (9, 10,11 & 13) whereby two exhibit industry affiliation, namely the high-yield Energy firms (cluster 13) and the high-yield Household Durables firms (cluster 9), and two which have no clear industry links (clusters 10 & 11).

To gauge the relative contribution of industry factors in these respective markets, we look to industries that cluster well (high F-measure) in both and whose constituents have barely altered before and after controlling for the market factor. For CDS returns, we identify that the mean correlation between constituents in the Road & Rail industry (cluster 9 , Table III) is 45.3% before the market factors' removal and 27.7% afterwards (cluster 5, Table VI). Similarly, for the Insurance industry it is 40.2% before (cluster 2, Table III)) and 19.4% afterwards (cluster 1, Table VI). For equity returns, we observe that the mean correlation between constituents in the Road & Rail industry is 58.9% before (cluster 7, Table IV) and 36.6% afterwards (cluster 8, Table VII). Similarly, for the Insurance industry it is 55.6% before (cluster 2, Table IV) and 24.5% afterwards (cluster 2, Table VII). Notable here is the similarity in the relative importance of the market and industry factors in these CDS and equity return correlations.

To assess the relative contribution of industry and high-yield factors to CDS correlations we look to Table VI. For clusters relating to the Energy Sector, the IG clusters 2, 12, 15 & 21 exhibit mean residual correlations of 17.1%, 12.4%, 13.6% and 13.5% respectively. In comparison the Energy Sector HY cluster 13 exhibits a correlation of 23.5%. For the HY Household Durables (cluster 9) the mean correlation is 24.6%. As for the HY clusters with no industry affiliation, 10 and 11, their mean residual correlations are 14.4% and 13.7% respectively. What is apparent from these numbers, is that the correlation linking HY issuers with no industry affiliation is similar to that linking IG industry affiliates. Also, when firms are industry affiliates in addition to being rated as high-yield, mean residual correlations are markedly higher. We therefore conjecture that high-yield factors have an effect on CDS correlations of equal magnitude, if not greater than, the industry factors pertaining to these firms.

D. Industry Factors Across Markets

From an aggregate perspective, there are major differences in the correlation structures of CDS and equity markets. However, the central role of industry factors in both has been an important finding in this study. In this section we turn to an individual industry in order to better understand whether there are similarities in industry factors across the markets. We choose the Energy Sector for several reasons. Firstly, the cluster analysis showed the prevalence of the Energy Sector in the equity return correlation matrix and hence the important role of industry factors for these firms; even after removing the market factor, the mean residual correlation is 51.6%. Secondly, given the results of the PCA analysis, oil prices are an important factor in the equity returns of these firms and given their primary business operations, provide an observable common factor[30]. Thirdly, the relatively poor clustering of the CDS returns of these firms by industry would suggest that no equally dominant industry factor is contained within them. In the following analysis we investigate separately the impact of this observable factor on the CDS and equity returns of this sector. We find evidence that the nature of common industry factors in security returns across these markets is different. This result strengthens our understanding of why the correlation structures of these two markets are strikingly different.

D.1. Equity Returns

We first perform a PCA on the equity return residuals from the 23 constituents of the Energy Sector (cluster 1) in Table VII. The PCs are now orthogonal to the market factor. Using the index

[30]We also experimented with other industries. Whilst economically sensible proxies for industry factors can be thought of, in this framework we require them to be available at a daily frequency and therefore market-traded assets are the limiting choice.

construction procedure, we test the correlation of their first PCI with the returns on WTI crude oil; they exhibit a sample correlation of 63.3%. Furthermore, the first PC explains 57.1% of the variation in this sector's equity return residuals. Although not reported, inspection of the normalized eigenvectors also shows that the first component in these residuals acts as a parallel shift factor, with all the weights being strictly positive and remarkably uniform. Their mean and standard deviation are 0.0435 and 0.012 respectively. To quantify the relative importance of S&P500 returns and WTI returns, two key factors in the equity returns of this sector, we make use of pooled OLS regressions. The results can be seen in Regressions 1 & 2 in Panel A of Table VIII. Both variables are positive and highly statistically significant, with panel robust t-stats of 19 and 17.1 respectively. The economic impact of WTI returns on the equity returns of this sample is essentially of the same order of magnitude as that of S&P500 returns, with coefficients of 1.015 and 0.952 respectively. By comparison to Regression 1, which excludes oil returns, we also note that the R^2 of the regression doubles with its inclusion from 16% to 34%. These results confirm that this industry factor plays a central role in the clustering of the Energy Sector in the equity return correlation matrix.

D.2. CDS Returns

We repeat the same analysis on the CDS returns of the most highly-clustered constituents of this sector, the nine investment grade firms in cluster 2 of Table VI. This subset of Energy Sector firms allow us to draw stronger conclusions about common industry factors than the smaller clusters. Inspection of the PC weights for the first PCI in these residuals again shows a remarkably uniform weighting on each of the nine firms, resembling a parallel shift factor, with a mean of 0.111 and a standard deviation of 0.017. The first PC explains 26.16% of their variation. Yet the correlation of the first PCI with WTI returns is a mere -5.89%; essentially statistically indifferent from zero. The industry factors driving correlation in the CDS return residuals of this subset are therefore unlikely to be related to oil prices. At best we can conclude that this PCI resembles latent factors related to their industry affiliation (and conditional on their investment-grade status) . For completeness, we also look to see whether WTI returns exert any meaningful statistical impact on the CDS returns of these firms. The pooled regressions in Panel B of Table VIII using the full 23 firms show that oil prices do exert a statistically significant influence on their CDS returns. The negative coefficient, -0.098, on WTI returns shows that increasing oil prices correspond to a tightening of spreads, however the economic significance is an order of magnitude smaller than that of the market factor. The coefficient on the return of the investment-grade credit market index is 0.788. Including this variable does not increase the R^2 of the regression.

IV. Conclusion

In this paper we have analyzed the correlation structures of the U.S. CDS and equity markets. We have shown that equity and CDS correlations are driven by market factors common to all firms as well as factors common to subsets of firms, such as industries or groups of industries. These factors appear to play similar roles in the correlation structures of these markets although correlations in CDS returns are typically lower given that they contain a relatively larger idiosyncratic component. In addition, CDS correlations provide direct evidence of factors related to rating classification above and below the investment-grade boundary whereas equity correlations do not. The existence of this boundary drives a wedge, intra-industry, between CDS correlations of firms on opposite sides of the boundary, and also leads to substantial correlation in the spreads of high-yield rated issuers from completely unrelated industries. The investment-grade boundary therefore has a discernable influence on the correlation structure of the CDS market.

Our results have direct implications for pricing and provide evidence of classification induced correlation, in line with the friction and sentiment-based co-movement theories. The exact mechanisms of how rating classification factors are generated and a better understanding of their empirical properties present interesting avenues for future research.

Table I: Summary statistics for the equity and CDS data, over the period February 3, 2004 to December 31, 2008 on a year-by-year basis. Panels A and C present statistics for the CDS data and B and D are that of the Equity data.

Panel A: Daily CDS Spread Levels (bp)

Period	No. of Firms	Mean	Std. Dev.	Median	Min.	Max.
2004	145	80.12	101.53	45.00	6.80	955.0
2005	145	75.49	106.61	41.80	4.00	2086.5
2006	145	75.11	113.37	35.50	3.00	1125.0
2007	145	91.64	148.01	38.50	2.90	1958.6
2008	145	290.90	547.62	114.00	13.60	9123.7
Full Sample	145	123.55	282.41	48.50	2.90	9123.7

Panel B: Daily Equity Prices ($)

Period	No. of Firms	Mean	Std. Dev.	Median	Min.	Max.
2004	145	32.60	15.84	30.42	3.35	91.31
2005	145	38.73	17.79	36.90	3.28	103.80
2006	145	41.96	19.53	41.29	3.46	119.41
2007	145	46.75	23.07	46.15	2.07	149.05
2008	145	39.47	26.37	36.11	0.28	161.40
Full Sample	145	40.02	21.43	37.84	0.28	161.40

Panel C: Log Daily CDS Returns

Period	No. of Firms	Mean	Std. Dev.	Median	Min.	Max.
2004	145	-0.0011	0.045	0	-0.92	1.27
2005	145	0.0006	0.051	0	-0.75	0.74
2006	145	-0.0013	0.047	0	-1.23	0.76
2007	145	0.0030	0.052	0	-0.51	1.25
2008	145	0.0049	0.055	0.0013	-1.05	0.83
Full Sample	145	0.0013	0.050	0	-1.23	1.27

Panel D: Log Daily Equity Returns

Period	No. of Firms	Mean	Std. Dev.	Median	Min.	Max.
2004	145	0.0007	0.016	0	-0.31	0.19
2005	145	0.0004	0.017	0	-0.18	0.28
2006	145	0.0003	0.017	0	-0.23	0.25
2007	145	-0.0002	0.020	0	-0.66	0.24
2008	145	-0.0025	0.048	-0.0002	-0.94	0.70
Full Sample	145	-0.0003	0.027	0	-0.94	0.70

Table II: Columns 1-3 present results from a principle components decomposition of the sample correlation matrices of CDS and equity returns for the 145 firms over the period February 3, 2004 to August 29, 2008. The first 10 principle components for each are reported. (%) Explain is the percentage of variation explained by the corresponding principle component. Cum (%) Explain is the cumulative percentage of variation explained by each additional component. Columns 4-18 are the results from cross-sectional regressions of the principle component weights on the 14 industry dummy variables. Bold font denotes significance at the 5% level or greater.

				t-Statistics														
	Prin. Comp.	(%) Explain.	Cum. (%) Explain.	Enrgy. Eq.	Oil&Gas	Chems	Rd&Rail	Auto Comp.	House Dur.	H.R.L	Spec. Ret.	Food Ret.	Food Prod.	Health	Pharm.	Insur.	Utils.	Adj. R^2
	1	18.50	18.50	**11.5**	**19.7**	**21.8**	**15.4**	**15.0**	**19.5**	**15.2**	**15.1**	**14.9**	**15.3**	**16.6**	**13.6**	**25.9**	**15.2**	0.258
	2	4.04	22.54	−0.21	−1.05	−0.27	**−4.62**	−1.16	−1.10	−0.66	−1.10	0.96	−0.58	**11.3**	−0.21	−0.59	−0.71	0.544
	3	3.39	25.93	−0.30	−0.54	−0.17	−0.007	−0.90	−1.49	−0.66	−0.47	−0.66	−0.71	−.53	**5.76**	0.78	0.06	0.229
	4	2.79	28.72	0.55	0.12	−0.53	−0.59	−0.52	−0.88	−0.35	−0.18	0.32	0.11	0.20	1.00	−1.83	**5.38**	0.213
CDS Ret.	5	2.39	31.11	0.09	−0.17	−0.83	0.07	**−2.11**	**−2.75**	−1.16	−0.97	−1.47	−1.32	1.35	**5.50**	**3.11**	−0.09	0.317
	6	2.11	33.22	−0.60	−1.26	−0.88	−0.17	**4.03**	**4.58**	1.88	1.19	1.14	1.61	**2.08**	**3.02**	**−8.75**	−1.12	0.516
	7	1.92	35.14	−0.36	−1.01	**−3.41**	−1.18	0.26	−0.93	−1.11	−1.25	−0.05	0.23	3.28	0.02	**4.51**	0.84	0.275
	8	1.82	36.96	1.35	1.54	**3.83**	**2.07**	**−2.86**	**−4.14**	1.22	0.51	0.36	1.90	1.47	1.73	−7.09	2.02	0.464
	9	1.64	38.60	1.44	**2.50**	−1.80	**−3.27**	**2.64**	**1.84**	−0.71	−0.10	−0.41	**−2.03**	**−3.73**	**3.30**	0.05	1.90	0.334
	10	1.57	40.17	**3.09**	**4.41**	**5.07**	**−2.88**	−0.54	0.42	−1.61	**−2.67**	**−2.53**	**−3.51**	−0.53	**−3.66**	0.73	1.94	0.456
	1	29.20	29.20	**6.45**	**9.69**	**10.2**	**8.94**	**12.8**	**21.4**	**9.56**	**8.04**	**5.77**	**4.74**	**5.66**	**4.43**	**14.2**	**4.80**	0.635
	2	9.65	38.85	**13.7**	**18.2**	**3.65**	0.21	**−2.09**	**−4.80**	**−2.48**	−1.83	−0.86	0.45	0.36	−0.14	**−5.50**	**2.43**	0.813
	3	5.70	44.55	**−2.34**	**−2.99**	1.00	1.02	**2.76**	**2.44**	1.51	1.64	1.21	0.83	**2.20**	1.15	**−3.45**	0.55	0.290
	4	4.89	49.44	−1.20	−1.43	**5.55**	**3.45**	**7.58**	**−14.9**	**3.33**	**2.16**	**2.94**	**3.53**	**3.78**	**3.68**	**6.02**	**2.41**	0.749
Stock Ret.	5	2.46	51.90	−1.20	−0.30	1.33	−0.33	**−12.7**	0.70	1.21	1.54	0.43	**2.80**	**7.54**	**3.85**	**5.39**	**3.72**	0.674
	6	1.87	53.77	0.65	**−2.11**	**2.64**	0.70	−0.13	−0.19	**−3.38**	−1.04	−0.47	0.10	**2.81**	1.86	0.033	0.93	0.220
	7	1.64	55.41	−0.143	0.59	−1.81	−0.54	−1.57	−1.51	1.25	−0.43	**3.44**	0.11	**4.53**	−0.42	−1.08	−1.00	0.256
	8	1.62	57.03	−0.061	1.57	**−5.81**	**−4.38**	2.16	0.019	−1.85	**−2.01**	**−2.78**	0.044	**6.48**	0.49	**3.43**	0.44	0.497
	9	1.38	58.41	−0.93	−0.19	1.12	1.34	**−2.45**	0.30	−1.92	0.88	1.88	0.28	−1.58	0.79	1.23	0.23	0.148
	10	1.22	59.63	**2.57**	−1.63	−0.33	0.36	0.25	0.46	0.094	**−2.81**	−0.21	0.32	0.87	0.69	−0.041	0.22	0.127

Table III: Clusters in the CDS return correlation matrix over the period February 3, 2004 to August 29, 2008. 22 unclustered firms have been omitted. Each firm is represented by: Six-digit GICS code - Rating Classification - Ticker symbol. Min. corr is the complete linkage correlation between cluster constituents. Mean. Corr is their mean pairwise correlation. The GICS code are: 101010 - Energy Equip. & Srvcs., 101020 - Oil, Gas & Cons. Fuels, 151010 - Chemicals, 203040 - Road & Rail, 251010 - Auto Components, 252010 - Household Durables, 253010 - Hotels, Rests. & Leisure, 255040 - Speciality Retail, 301010 - Food & Staples Retail, 302020 - Food Products, 351020 - Health Care Provs. & Srvcs., 352020 - Pharmaceuticals, 403010 - Insurance, 551030 - Multi-Utils.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
	255040-IG-AZO 253010-IG-CCL 301010-IG-KR 255040-IG-LOW 255040-IG/HY-LTD 253010-IG-MAR 252010-IG-NWL 255040-IG/HY-RSH 301010-IG-SWY 301010-IG/HY-SVU 302020-IG/HY-TSN 252010-IG-WHR	403010-IG-ACE 403010-IG-ALL 403010-IG-AIG 403010-IG-CB 403010-IG-CNA 403010-IG-HIG 403010-IG-LNC 403010-IG-L 403010-IG-MET 403010-IG-PRU 403010-IG-TRV 403010-IG-XL	351020-IG-AET 351020-IG-CAH 351020-IG-CI 351020-IG-MCK 351020-HY/IG-MHS 151010-HY-POL 301010-HY-RAD 351020-IG-UNH 351020-IG-UHS 351020-IG-WLP	252010-HY-BZH 252010-IG/HY-CTX 252010-HY/IG-DHI 252010-HY-KBH 252010-IG/HY-LEN 252010-IG-MDC 252010-IG/HY-PHM 252010-HY/IG/HY-RYL 252010-HY-SPF 252010-IG/HY-TOL	101020-IG-APC 101020-IG-APA 101020-IG-COP 101020-IG-DVN 101010-IG-HAL 101020-IG-HSE 101020-IG-MRO 101020-IG-OXY 101020-IG-VLO 101020-IG-XTO	151010-IG-CYT 151010-IG-DOW 151010-IG-DD 151010-IG-EMN 151010-IG/HY-OLN 151010-IG-PPG 151010-IG-ROH 151010-IG-RPM 255040-IG-SHW	551030-HY-CMS 101020-HY-EP 253010-HY-MGM 253010-HY-RCL 302020-HY-SFD 101020-HY/IG-WMB	251010-IG/HY-AXL 251010-HY-ARM 251010-IG/HY-CTB 251010-HY-GT 251010-IG/HY-LEA 251010-HY-VC	203040-IG-BNI 203040-IG-CNW 203040-IG-CSX 203040-IG-NSC 203040-IG-R 203040-IG-UNP
Min. Corr.	0.258	0.258	0.245	0.313	0.297	0.274	0.305	0.363	0.266
Mean. Corr.	0.374	0.402	0.463	0.554	0.375	0.355	0.377	0.482	0.453

	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18
	551030-IG-ED 551030-IG-D 551030-IG-DTE 551030-IG-SRE 551030-HY/IG-TE	302020-IG-CPB 302020-IG-CAG 302020-IG-KFT 253010-IG-MCD 302020-IG-SLE	352020-IG-ABT 352020-IG-BMY 352020-IG-WYE	101020-HY-CHK 101010-HY-PDE 101020-HY-TSO	101010-IG-DO 101010-IG-RIG 101010-IG-WFT	252010-IG-BDK 301010-IG-CVS 302020-IG-HN	101020-IG-KMP 101020-IG-OKS	253010-IG-DRI 253010-IG-YUM	403010-IG-AOC 403010-IG-MMC
Min. Corr.	0.317	0.246	0.254	0.334	0.232	0.234	0.306	0.267	0.514
Mean. Corr.	0.390	0.283	0.293	0.375	0.248	0.252	0.306	0.267	0.514

	Cluster 19	Cluster 20	Cluster 21	Cluster 22	Cluster 23	Cluster 24	Cluster 25
	203040-HY-CAR 351020-HY-THC	403010-IG/HY-ABK 403010-IG/HY-MBI	151010-IG-APD 151010-IG-PX	151010-IG-MON 253010-HY/IG-HOT	251010-IG-BWA 251010-IG-JCI	551030-IG-WEC 551030-IG-XEL	351020-HY/IG-ABC 351020-IG/HY-HMA
Min. Corr.	0.276	0.555	0.311	0.238	0.346	0.278	0.373
Mean. Corr.	0.276	0.555	0.311	0.238	0.346	0.278	0.373

Table IV: Clusters in the equity return correlation matrix over the period February 3, 2004 to August 29, 2008. 12 unclustered firms have been omitted. Each firm is represented by: Six-digit GICS code - Rating Classification - Ticker symbol. Min. corr is the complete linkage correlation between cluster constituents. Mean. Corr is their mean pairwise correlation. The GICS codes are: 101010 - Energy Equip. & Srvcs., 101020 - Oil, Gas & Cons. Fuels, 151010 - Chemicals, 203040 - Road & Rail, 251010 - Auto Components, 252010 - Household Durables, 253010 - Hotels, Rests. & Leisure, 255040 - Speciality Retail, 301010 - Food & Staples Retail, 302020 - Food Products, 351020 - Health Care Provs. & Srvcs., 352020 - Pharmaceuticals, 403010 - Insurance, 551030 - Multi-Utils.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
	101020-IG-APC	403010-IG-ACE	151010-IG-APD	252010-HY-BZH	551030-HY-CNP	252010-IG-BDK	203040-IG-BNI	251010-IG/HY-AXL	352020-IG-ABT
	101020-IG-APA	403010-IG-ALL	151010-IG-CYT	252010-IG/HY-CTX	551030-HY-CMS	253010-IG-CCL	203040-IG-CNW	251010-HY-ARM	352020-IG-BMY
	101010-IG-BHI	403010-IG-AIG	151010-IG-DOW	252010-HY/IG-DHI	551030-IG-ED	255040-IG-LOW	203040-IG-CSX	251010-IG-BWA	352020-IG-LLY
	101020-HY-CHK	403010-IG-CB	151010-IG-DD	252010-HY-KBH	551030-IG-D	252010-IG-NWL	203040-IG-NSC	251010-IG-JCI	352020-IG-JNJ
	101020-IG-CVX	403010-IG-CNA	151010-IG-EMN	252010-IG/HY-LEN	551030-IG-DTE	253010-HY-RCL	203040-IG-R	251010-IG/HY-LEA	352020-IG-PFE
	101020-IG-COP	403010-IG-HIG	151010-IG-LZ	252010-IG-MDC	551030-IG-SRE	255040-IG-SHW	203040-IG-UNP	251010-HY-VC	352020-IG-WYE
	101020-IG-DVN	403010-IG-LNC	151010-IG/HY-OLN	252010-IG/HY-PHM	551030-HY/IG-TE	255040-IG-SPLS			
	101010-IG-DO	403010-IG-L	151010-HY-POL	252010-HY/IG/HY-RYL	551030-IG-WEC	252010-IG-WHR			
	101020-HY-EP	403010-IG-MET	151010-IG-PPG	252010-HY-SPF	551030-IG-XEL				
	101010-IG-HAL	403010-IG-PRU	151010-IG-PX	252010-IG/HY-TOL					
	101020-IG-HSE	403010-IG-TRV	151010-IG-RPM						
	101020-IG-MRO	403010-IG/HY-UNM							
	101010-IG-NBR	403010-IG-XL							
	101020-IG-OXY								
	101010-HY-PDE								
	101020-IG-SUN								
	101020-HY-TSO								
	101010-IG-RIG								
	101020-IG-VLO								
	101010-IG-WFT								
	101020-HY/IG-WMB								
	101020-IG-XTO								
Min. Corr.	0.374	0.380	0.376	0.607	0.524	0.381	0.433	0.445	0.348
Mean. Corr.	0.642	0.556	0.495	0.771	0.583	0.460	0.589	0.510	0.409

	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18
	255040-IG-AZO	351020-IG-AET	253010-IG-MAR	302020-IG-CPB	203040-HY-CAR	253010-IG-DRI	301010-IG-KR	351020-HY/IG-ABC	302020-IG-CAG
	301010-IG-COST	351020-IG-CI	253010-HY-MGM	302020-IG-HNZ	251010-IG/HY-CTB	253010-IG-MCD	301010-IG-SWY	351020-IG-CAH	302020-IG-SLE
	255040-IG/HY-LTD	351020-IG-UNH	253010-HY/IG-HOT	302020-IG-KFT	251010-HY-GT	253010-IG-YUM	301010-IG/HY-SVU	351020-IG-MCK	
	255040-IG-TJX	351020-IG-WLP	253010-IG/HY-WEN						
	301010-IG-WMT								
Min. Corr.	0.385	0.510	0.335	0.400	0.357	0.378	0.400	0.495	0.360
Mean. Corr.	0.460	0.588	0.484	0.429	0.410	0.437	0.473	0.518	0.360

	Cluster 19	Cluster 20	Cluster 21	Cluster 22	Cluster 23	Cluster 24
	403010-IG-AOC	403010-IG/HY-ABK	352020-IG-MRK	101020-IG-KMP	302020-HY-SFD	302020-IG-ADM
	403010-IG-MMC	403010-IG/HY-MBI	352020-IG-SGP	101020-IG-OKS	302020-IG/HY-TSN	151010-IG-MON
Min. Corr.	0.536	0.782	0.359	0.510	0.374	0.333
Mean. Corr.	0.536	0.782	0.359	0.510	0.374	0.333

Table V: The output of the cluster validation algorithm (Appendix C) which gives the set of clusters presented in Tables III & IV from the entire cluster hierarchy. The 14 GICS industries and number of constituents are presented in columns 1-3. Columns 4 & 6 show the cluster in Table III or IV which is determined by the algorithm as the best representation of that industry, according to their F-measures in Columns 5 & 7. The F-measure is the extent to which the cluster contains only and all firms of that industry. The weighted average F-measure determines the Phi-measure seen in the final row.

			CDS Return Correlation Matrix		Equity Return Correlation Matrix	
GICS Code	Industry	No. Firms	Cluster	F-Measure	Cluster	F-Measure
101010	Energy Equipment & Services	7	14	0.60	1	0.48
101020	Oil & Gas Producers	17	5	0.67	1	0.77
151010	Chemicals	14	6	0.70	3	0.88
203040	Road & Rail	7	9	0.92	7	0.92
251010	Auto Components	8	8	0.86	8	0.86
252010	Household Durables	13	4	0.87	4	0.87
253010	Hotels Restaurants & Leisure	9	17	0.36	12	0.62
255040	Specialty Retail	7	1	0.42	10	0.50
301010	Food & Staples Retailing	7	1	0.32	16	0.60
302020	Food Products	9	11	0.57	13	0.50
351020	Health Care Providers & Services	12	3	0.73	11	0.50
352020	Pharmaceuticals	9	12	0.50	9	0.80
403010	Insurance	17	2	0.83	2	0.87
551030	Multi-Utilities & Unregulated Power	9	10	0.71	5	1.00
			Phi-Measure:	0.67		0.75

Table VI: The clusters found in the CDS return residual correlation matrix over the period February 3, 2004 to August 29, 2008. These residuals result from firm-by-firm time-series regressions of CDS returns on the IGRI and a constant. The GICS codes are: 101010 - Energy Equip. & Srvcs., 101020 - Oil, Gas & Cons. Fuels, 151010 - Chemicals, 203040 - Road & Rail, 252010 - Household Durables, 253010 - Hotels, Rests. & Leisure, 302020 - Food Products, 351020 - Health Care Provs. & Srvcs., 352020 - Pharmaceuticals, 403010 - Insurance, 551030 - Multi-Utils. The clustering Phi-measure is 69%.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
	403010-IG-ACE	101020-IG-APC	351020-IG-AET	352020-IG-ABT	203040-IG-BNI	151010-IG-DOW	302020-IG-ADM	151010-IG-APD	252010-HY-BZH
	403010-IG-ALL	101020-IG-APA	352020-IG-AGN	352020-IG-BMY	203040-IG-CNW	151010-IG-DD	302020-IG-CPB	151010-IG-CYT	252010-HY-KBH
	403010-IG-AIG	101020-IG-COP	351020-IG-CAH	551030-IG-ED	203040-IG-CSX	151010-IG-EMN	302020-IG-CAG	151010-IG-MON	252010-IG-MDC
	403010-IG-CNA	101020-IG-DVN	351020-IG-CI	551030-IG-D	203040-IG-NSC	151010-IG-LZ	302020-IG-HNZ	151010-IG-PX	252010-HY-SPF
	403010-IG-HIG	101010-IG-HAL	351020-IG-MCK	551030-IG-DTE	203040-IG-R	151010-IG-PPG	302020-IG-KFT	151010-IG-RPM	
	403010-IG-LNC	101020-IG-HSE	151010-HY-POL	551030-IG-SRE	203040-IG-UNP	151010-IG-ROH	302020-IG-SLE		
	403010-IG-L	101020-IG-MRO	351020-IG-UNH	352020-IG-WYE					
	403010-IG-MET	101020-IG-VLO	351020-IG-UHS						
	403010-IG-PRU	101020-IG-XTO	351020-IG-WLP						
	403010-IG-TRV								
	403010-IG-XL								
Min. Corr.	0.072	0.083	0.138	0.038	0.055	0.033	0.037	0.053	0.135
Mean. Corr.	0.194	0.171	0.343	0.121	0.277	0.128	0.108	0.118	0.246

	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18
	551030-HY-CMS	203040-HY-CAR	101010-IG-DO	101020-HY-CHK	403010-IG-AOC	101010-IG-BHI	101020-IG-CVX	352020-IG-LLY	253010-IG-CCLI
	302020-HY-DF	253010-HY-MGM	101010-IG-RIG	101010-HY-PDE	252010-IG-BDK	101010-IG-NBR	551030-IG-WEC	352020-IG-MRK	253010-IG-MAR
	101020-HY-EP	253010-HY-RCL	101010-IG-WFT	101020-HY-TSO	403010-IG-MMC	101020-IG-OXY	551030-IG-XEL	352020-IG-PFE	
	302020-HY-SFD	351020-HY-THC							
Min. Corr.	0.067	0.081	0.117	0.187	-0.108	0.121	0.065	0.093	0.191
Mean. Corr.	0.144	0.137	0.124	0.235	0.084	0.136	0.122	0.120	0.191

	Cluster 19	Cluster 20	Cluster 21	Cluster 22	Cluster 23	Cluster 24	Singleton Clusters
	403010-IG-CB	253010-IG-DR	101020-IG-KMP	252010-IG-NWL	253010-IG-MCD	351020-IG-DGX	352020-IG-JNJ
	403010-IG/HY-UNM	101020-IG-SUN	101020-IG-OKS	252010-IG-WHR	253010-IG-YUM	352020-IG-SGP	551030-HY-CNP
Min. Corr.	0.146	-0.087	0.135	0.195	0.110	0.126	
Mean. Corr.	0.146	-0.087	0.135	0.195	0.110	0.126	

Table VII: The set of clusters found in the equity return residual correlation matrix over the period February 3, 2004 to August 29, 2008. These residuals result from firm-by-firm time-series regressions of equity returns on S&P500 returns and a constant. The GICS codes are: 101010 - Energy Equip. & Srvcs., 101020 - Oil, Gas & Cons. Fuels, 151010 - Chemicals, 203040 - Road & Rail, 252010 - Household Durables, 253010 - Hotels, Rests. & Leisure, 302020 - Food Products, 351020 - Health Care Provs. & Srvcs., 352020 - Pharmaceuticals, 403010 - Insurance, 551030 - Multi-Utils. The clustering Phi-measure is 86%.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
	101020-IG-APC	403010-IG-ACE	151010-IG-APD	203040-HY-CAR	352020-IG-ABT	551030-HY-CNP	351020-IG-AET	203040-IG-BNI	252010-HY-BZH
	101020-IG-APA	403010-IG-ALL	151010-IG-DOW	253010-IG-CCL	352020-IG-BMY	551030-HY-CMS	351020-IG-CI	203040-IG-CNW	252010-IG-BDK
	101010-IG-BHI	403010-IG-AIG	151010-IG-DD	302020-HY-DF	352020-IG-LLY	551030-IG-ED	351020-IG-DGX	203040-IG-CSX	252010-HY-KBH
	101020-HY-CHK	403010-IG-AOC	151010-IG-EMN	253010-IG-MAR	352020-IG-JNJ	551030-IG-D	351020-HY-THC	203040-IG-NSC	252010-IG-MDC
	101020-IG-CVX	403010-IG-CB	151010-IG-LZ	253010-HY-MGM	352020-IG-MRK	551030-IG-DTE	351020-IG-UNH	203040-IG-R	252010-HY-SPF
	101020-IG-COP	403010-IG-CNA	151010-IG-MON	252010-IG-NWL	352020-IG-PFE	551030-IG-SRE	351020-IG-UHS	203040-IG-UNP	
	101020-IG-DVN	403010-IG-HIG	151010-IG-PPG	253010-HY-RCL	352020-IG-SGP	551030-IG-WEC	351020-IG-WLP		
	101010-IG-DO	403010-IG-LNC	151010-IG-PX	302020-HY-SFD	352020-IG-WYE	551030-IG-XEL			
	101020-HY-EP	403010-IG-L	151010-IG-RPM	252010-IG-WHR					
	101010-IG-HAL	403010-IG-MMC							
	101020-IG-HSE	403010-IG-MET							
	101020-IG-KMP	403010-IG-PRU							
	101020-IG-MRO	403010-IG-TRV							
	101010-IG-NBR	403010-IG/HY-UNM							
	101020-IG-OXY	403010-IG-XL							
	101020-IG-OKS								
	101010-HY-PDE								
	101020-IG-SUN								
	101020-HY-TSO								
	101010-IG-RIG								
	101020-IG-VLO								
	101010-IG-WFT								
	101020-IG-XTO								
Min. Corr.	0.154	0.070	0.111	0.044	0.113	0.331	0.049	0.128	0.131
Mean. Corr.	0.516	0.245	0.234	0.147	0.215	0.436	0.244	0.366	0.424

	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Singleton Clusters
	302020-IG-CPB	151010-IG-CYT	352020-IG-AGN	253010-IG-DRI	302020-IG-ADM
	302020-IG-CAG	151010-HY-POL	351020-IG-CAH	253010-IG-MCD	
	302020-IG-HNZ	151010-IG-ROH	351020-IG-MCK	253010-IG-YUM	
	302020-IG-KFT				
	302020-IG-SLE				
Min. Corr.	0.146	0.081	0.076	0.201	
Mean. Corr.	0.206	0.140	0.188	0.249	

Table VIII: Balanced Panel regressions for the equity and CDS returns of the 23 firms in the GICS Energy Sector on a market variable and WTI crude oil returns. Panel robust standard errors (s.e.) are reported. Panel A presents the results for equity returns and Panel B is that of CDS returns. Regression 1 considers only the market variables, defined as S&P500 returns ($R_{S\&P}$) and the Investment-Grade Return Index (R_{IGI}). Regression 2 adds WTI returns (R_{WTI}) as an explanatory variable.

Panel A: GICS Energy Sector Equity Return Regressions

	Regression					
	1			2		
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat
$R_{S\&P}$	0.948	0.0510	18.6	1.015	0.0535	19.0
R_{WTI}				0.952	0.0557	17.1
Const.	0.0007	0.00006	11.0	0.0003	0.00005	6.3
R^2	0.163			0.340		
No. Observations	27462			27462		

Panel B: GICS Energy Sector CDS Return Regressions

	Regression					
	1			2		
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat
R_{IGI}	0.788	0.0296	26.7	0.788	0.0296	26.7
R_{WTI}				-0.098	0.0220	-4.45
Const.	-0.0001	0.0001	-1.04	-0.00007	0.0001	-0.67
R^2	0.175			0.175		
No. Observations	27462			27462		

References

Anderberg, M.: 1973, *Cluster analysis for applications*, Academic Press.

Barberis, N. and Shleifer, A.: 2003, Style investing, *Journal of Financial Economics* **68**, 161–199.

Barberis, N., Shleifer, A. and Wurgler, J.: 2005, Comovement, *Journal of Financial Economics* **75**, 283–317.

Bhojraj, S., Lee, C. and Oler, D.: 2003, What's my line? a comparison of industry classification schemes for capital market research, *Journal of Accounting Research* **41**, 745–774.

Chernenko, S. and Sunderam, A.: 2009, The real consequences of market segmentation, *Working paper* .

Daniels, K. N. and Jensen, M. S.: 2005, The effect of credit ratings on credit default swap spreads and credit spreads, *Journal of Fixed Income* pp. 16–33.

Das, S., Freed, L., Geng, G. and Kapadia, N.: 2006, Correlated default risk, *Journal of Fixed Income* **14**(1), 7–32.

Elizalde, A.: 2005, Do we need to worry about credit risk correlation?, *Journal of Fixed Income* pp. 42–59.

King, B. F.: 1966, Market and industry factors in stock price behavior, *Journal of Business* **39**, 139–190.

King, B. F.: 1967, Step-wise clustering procedures, *Journal of the American Statistical Association* **62**, 86–101.

Kisgen, D. J. and Strahan, P. E.: 2009, Do regulations based on credit ratings affect a firm's cost of capital?, *Working paper* .

Livingston, M.: 1977, Industry movements of common stocks, *Journal of Finance* **32**, 861–874.

Longstaff, F. A. and Rajan, A.: 2008, An empirical analysis of the pricing of collateralized debt obligations, *Journal of Finance* **63**(2), 529–563.

Meyers, S. L.: 1973, A re-examination of market and industry factors in stock price behaviour, *Journal of Finance* **28**, 695–705.

Micu, M., Remolona, E. and Wooldridge, P.: 2006, The price impact of rating announcements: which announcements matter? BIS working papers.

Roll, R.: 1988, r^2, *Journal of Finance* **43**, 541–566.

A. List of Firms

The 145 firms are grouped by six-digit GICS industries with accompanying equity ticker symbols and rating classifications, defined as investment-grade (IG), high-yield (HY) or some combination of the two depending on rating actions by Moody's during the sample period.

Industry	GICS Code	Ticker	Company Name	Rating Class.
Energy Equipment & Services	101010	BHI	Baker Hughes Inc.	IG
		DO	Diamond Offshore Drilling Inc.	IG
		HAL	Halliburton Company	IG
		NBR	Nabors Industries Ltd.	IG
		PDE	Prode International Inc.	HY
		RIG	Tranocean Ltd.	IG
		WFT	Weatherford International Ltd.	IG
Oil, Gas & Consumable Fuels	101020	APC	Anadarko Petroleum Corp.	IG
		APA	Apache Corp.	IG
		CHK	Chesapeake Energy Corp.	HY
		CVX	Chevron Corp.	IG
		COP	Conocophillips	IG
		DVN	Devon Energy Corp.	IG
		EP	El Paso Corp.	HY
		HSE	Hess Corp.	IG
		KMP	Kinder Morgan Energy LP	IG
		MRO	Marathon Oil Corp	IG
		OXY	Occidental Petroleum Corp.	IG
		OKS	Oneok Partners LP	IG
		SUN	Sunoco Inc.	IG
		TSO	Tesoro Corp.	HY
		VLO	Valero Energy Corp.	IG
		WMB	Williams Cos. Inc.	HY/IG
		XTO	XTO Energy Inc.	IG
Chemicals	151010	APD	Air Products & Chemicals Inc.	IG
		ASH	Ashland Inc.	IG/HY
		CYT	Cytec Industries Inc.	IG
		DOW	Dow Chemical	IG
		DD	E.I. du Pont de Nemours & Company	IG
		EMN	Eastman Chemical Company	IG
		LZ	Lubrizol Corp.	IG
		MON	Monsanto Company	IG
		OLN	Olin Corp.	IG/HY
		POL	Polyone Corp.	HY
		PPG	PPG Industries Inc.	IG
		PX	Praxair Inc.	IG
		ROH	Rohm & Haas Company	IG
		RPM	RPM International Inc.	IG
Road & Rail	203040	CAR	Avis Budget Group Inc.	HY
		BNI	Burlington Northern Santa Fe	IG
		CNW	Con-Way Inc.	IG
		CSX	CSX Corp.	IG
		NSC	Norfolk Southern Corp.	IG
		R	Ryder System Inc.	IG
		UNP	Union Pacific Corp.	IG
Auto Components	251010	AXL	American Axle & Manufacturing Holdings Inc.	IG/HY
		ARM	Arvinmeritor Inc.	HY
		BWA	Borgwarner Inc.	IG
		CTB	Cooper Tire & Rubber Company	IG/HY
		GT	Goodyear Tire & Rubber Company	HY
		JCI	Johnson Controls Inc.	IG
		LEA	Lear Corp.	IG/HY
		VC	Visteon Corp.	HY
Household Durables	252010	BZH	Beazer Homes USA Inc.	HY
		BDK	Black & Decker Corp.	IG
		CTX	Centex Corp.	IG/HY
		DHI	D.R. Horton Inc.	HY/IG/HY
		KBH	KB Home	HY
		LEN	Lennar Corp.	IG/HY
		MDC	M.D.C Holdings Inc.	IG
		NWL	Newell Rubbermaid Inc.	IG
		PHM	Pulte Homes Inc.	IG/HY
		RYL	Ryland Group Inc.	HY/IG/HY
		SPF	Standard Pacific Corp.	HY
		TOL	Toll Brothers Inc.	IG/HY
		WHR	Whirlpool Corp.	IG

(Continued on next page)

Industry	GICS Code	Ticker	Company Name	Rating Class.
Hotels, Restaurants & Leisure	253010	CCL	Carnival Corp.	IG
		DRI	Darden Restaurants Inc.	IG
		MAR	Marriott International Inc.	IG
		MCD	Mcdonald'S Corp.	IG
		MGM	MGM Mirage	HY
		RCL	Royal Caribbean Cruises Ltd.	HY
		HOT	Starwood Hotels & Resorts Worldwide Inc.	HY/IG
		WEN	Wendy'S Arby'S Group Inc.	IG/HY
		YUM	Yum! Brands Inc.	IG
Speciality Retail	255040	AZO	Autozone Inc.	IG
		LOW	Lowe's Companies Inc.	IG
		LTD	Limited Brands Inc.	IG/HY
		RSH	Radioshack Corp.	IG/HY
		SHW	Sherwin-Williams Co.	IG
		SPLS	Staples Inc.	IG
		TJX	TJX Companies Inc.	IG
Food & Staples Retail	301010	COST	Costco Wholesale Corp.	IG
		CVS	CVS Caremark Copr.	IG
		KR	Kroger Company	IG
		RAD	Rite aid Corp.	HY
		SWY	Safeway Inc.	IG
		SVU	Supervalu Inc.	IG/HY
		WMT	Wal-mart Stores Inc.	IG
Food Products	302020	ADM	Archer-Daniels-Midland	IG
		CPB	Campbell Soup Company	IG
		CAG	Conagra Foods Inc.	IG
		DF	Dean Foods Company	HY
		HNZ	Heinz Company	IG
		KFT	Kraft Foods Inc.	IG
		SLE	Sara Lee Corp.	IG
		SFD	Smithfields Foods Inc.	HY
		TSN	Tyson Foods Inc.	IG/HY
Healthcare Providers	351020	AET	Aetna Inc.	IG
		ABC	Amerisourcebergen Corp.	HY/IG
		CAH	Cardinal Health Inc.	IG
		CI	Cigna Corp.	IG
		HMA	Health Management Assoc.	IG/HY
		MCK	Mckesson Corp.	IG
		MHS	Medco Health Solutions	HY/IG
		DGX	Quest Diagnostics	IG
		THC	Tenet Healthcare Corp.	HY
		UNH	Unitedhealth Group Inc.	IG
		UHS	Universal Health Services	IG
		WLP	Wellpoint Inc.	IG
Pharmaceuticals	352020	ABT	Abbott Laboratories	IG
		AGN	Allergan	IG
		BMY	Bristol-Myers Squibb	IG
		LLY	Lilly (Eli) & Company	IG
		JNJ	Johnson & Johnson	IG
		MRK	Merck & Company	IG
		PFE	Pfizer Inc.	IG
		SGP	Schering-Plough	IG
		WYE	Wyeth	IG
Insurance	403010	ACE	ACE Ltd.	IG
		ALL	Allstate Corp.	IG
		ABK	Ambac Financial Group Inc.	IG/HY
		AIG	American International Group Inc.	IG
		AOC	Aon Corp.	IG
		CB	The Chubb Corp.	IG
		CNA	CNA Financial Corp.	IG
		HIG	Hartford International Services	IG
		LNC	Lincoln National Corp.	IG
		L	Loews Corp.	IG
		MMC	Marsh & Mclennan Companies Inc.	IG
		MBI	MBIA Inc.	IG/HY
		MET	MetLife Inc.	IG
		PRU	Prudential Financial Inc.	IG
		TRV	Travelers Companies Inc.	IG
		UNUM	Unum Group	HY
		XL	XL Group Plc.	IG
Multi-Utilities	551030	CNP	Centerpoint Energy Inc.	HY
		CMS	CMS Energy Corp.	HY
		ED	Consolidated Edison Inc.	IG
		D	Dominion Resources Inc.	IG
		DTE	DTE Energy Company	IG
		SRE	Sempra Energy	IG
		TE	Teco Energy Inc.	HY/IG
		WEC	Wisconsin Energy Corp.	IG
		XEL	Xcel Energy Inc.	IG

B. Hierarchical Clustering

The central procedure underpinning the agglomerative[31] hierarchical[32] algorithms is simple. Whilst the general case holds for any well-defined similarity measure between firms i and j, $s_{i,j}$, we highlight[33] the methodology using the sample product moment correlation coefficient, $\rho_{i,j}$. The lower triangular correlation matrix is first computed and contains all $\frac{1}{2}N(N-1)$ pairwise correlations. The agglomerative clustering procedure is as follows:

1. Begin with a set of N clusters, each cluster containing one firm. Denote the clusters 1 to N.
2. Identify from the correlation matrix which two clusters have the highest correlation, label the clusters p and q.
3. Merge clusters p and q, denoting the resulting cluster r and reduce the number of clusters by 1. Delete the row and column of the correlation matrix relating to cluster p and revise the correlation matrix to reflect the correlation between the unmerged $N-2$ clusters and the newly formed cluster r.
4. Repeat steps 2 and 3 a total of $N-1$ times until all firms are in a single cluster, at each stage recording which clusters have been merged and the value of the correlation coefficient which lead to the merger.

The difference between the agglomerative methods lies in how to update the correlation matrix following the merger of two clusters in step 3 above.

[31] The distinction between the two types of hierarchical cluster method, agglomerative and divisive, lies in the starting point of the clustering procedure. Agglomerative methods begin with a set of N clusters and finish when all firms are in a single cluster. Divisive methods begin with a single cluster containing all firms and work backwards towards a set of N clusters.

[32] The most common distinction among various types of clustering method is that of hierarchical versus partitional. In partitional clustering, firms are grouped into non-overlapping clusters. Hierarchical clustering forms a set of nested clusters that may be organized as a dendrogram, whereby nodes of the tree represent the merger of existing subclusters and the root of the tree is a singleton cluster containing all firms.

[33] This description follows that of Anderberg (1973), Chapter 6, p. 132

C. Cluster Validity

The output of the cluster algorithm is a set of nested clusters, arranged as $l = N$ levels of a dendrogram. The question of which set of clusters gives the best representation of the structure present in a correlation matrix is a nontrivial issue[34]. In looking for industry-based structure to the clusters, we introduce some statistical measures known as the classification orientated measures of cluster validity. By denoting N as the total number of firms in the sample, Y the total number of industries which they represent, n^y as the number of firms in industry y for $y = 1,..,Y$, $n_{x,l}$ as the number of firms in cluster x at level l for $x = 1,..,l$ & $l = N,..,1$ and $n^y_{x,l}$ as the number of firms of industry y in cluster x at level l, we define:

- Precision: The probability that a member of cluster x, for a given level l, belongs to industry y. I.e. the fraction of a cluster that consists of firms of a given industry.

$$P^y_{x,l} = \frac{n^y_{x,l}}{n_{x,l}} \tag{2}$$

- Recall: The probability that a member of industry y is in cluster x, at a given level l. I.e. the fraction of all members of industry y in cluster x.

$$R^y_{x,l} = \frac{n^y_{x,l}}{n^y} \tag{3}$$

- F-Measure: A combination of precision and recall that measures the extent to which a cluster contains only and all firms of a particular industry, calculated as the harmonic mean of precision and recall.

$$F^y_{x,l} = \frac{2 \cdot P^y_{x,l} \cdot R^y_{x,l}}{P^y_{x,l} + R^y_{x,l}} \tag{4}$$

  Which is bounded from above by 1, when there is a perfect correspondence between the constituents of a given cluster and those of a given industry, and by 0 from below when there is no relationship between a given cluster and industry.

[34] As explained by King (1967), methodologies such as ours fall under the heading of 'data analysis'.

By utilizing these statistics, we define the Phi-measure as:

$$\Phi = \max_{N \geq l \geq 1} \left[\sum_{1}^{Y} \frac{n^y}{N} \cdot \max_{1 \leq y \leq Y} F^y_{x,l} \right] \tag{5}$$

Explicitly, for each level l in the cluster hierarchy, the validation algorithm calculates a weighted-average of F-measures, for each industry y, from the cluster which best represents it. The weights are based on industry sizes. Across all levels in the hierarchy, the maximum of these weighted average F-measures determines the Phi-measure. The idea behind this metric is that for the set of clusters provided at each level of the cluster hierarchy, l, the distribution of firms across clusters is mapped onto the distribution of firms across the six-digit industry-level GICS codes and scored on a zero (more concretely it tends to zero with increasing sample size[35]) to one basis. This simultaneously gives the number of clusters, across which the distribution of firms best fits the industry distribution provided by the GICS and a quantitative score of goodness of fit. In the main analysis of this paper, we consider the set of clusters in the two correlation matrices produced from this algorithm.

[35]This is a consequence of using an exclusive and complete agglomerative clustering method, whereby there will always be at least one firm per cluster and all firms belong to a cluster. The term complete in this sense shares no relationship to the name for this agglomerative method, complete linkage, which stems from connections to graph theory. It can be easily shown that in a finite sample, the minimum exists when $l = 1$.