Ubiquitous Comovement*

William Grieser† Jung Hoon Lee‡ Morad Zekhnini‡

March 26, 2019

Abstract

Rational and behavioral asset pricing theories offer conflicting interpretations of the covariance structure of asset returns. Return comovement beyond what empirical factor models can explain is often interpreted in favor of frictions or behavioral explanations. However, we show that randomly grouped assets exhibit "excess" comovement that is ubiquitous and often indistinguishable from that of economically-motivated sources advanced in the literature. We further show, theoretically and through simulations, that this finding is consistent with an unobservable risk factor in a rational model. We then prescribe a revised procedure and statistical test to identify excess comovement based on distinguishing between rational and behavioral explanations.

Keywords: excess comovement, fundamentals, rational markets, behavioral finance, stock returns, factor models

JEL Codes: G11, G12, G14, G40

*We thank Scott Cederburg, Paul Irvine, Ryan Israelsen, Kris Jacobs, Ron Kaniel, Nishad Kapadia, Mark Schroder, Andrei Simonov, James Weston, and Toni Whited for thoughtful comments. We also thank additional seminar participants at Michigan State University, University of Houston, University of New Orleans, Rice University, Texas Christian University, and Tulane University. All errors are our own. †Department of Finance, Neeley School of Business Texas Christian University, phone: (817) 253-3535, e-mail: w.grieser@tcu.edu. ‡Department of Finance, A.B. Freeman School of Business Tulane University, e-mail: jlee39@tulane.edu, mzekhnin@tulane.edu.

1 Introduction

A large and growing literature has documented asset return comovement beyond what can be explained by common empirical asset pricing models. The central question explored by this literature is whether residual return comovement indicates a violation of the rational market paradigm. However, in this paper we provide empirical evidence that residual return comovement is ubiquitous and does not necessarily characterize imperfect markets. To illustrate this point, we show that randomly grouped assets generally exhibit substantial within-group residual return comovement, which is often indistinguishable from that of economically-motivated sources advanced in the literature. Thus, extreme caution should be exercised when interpreting the magnitude of results as evidence of a given explanation. We prescribe a revised procedure that accounts for potential latent factors and propose a new test that exploits different implications of rational and irrational explanations.

Traditional asset pricing theory contends that, in a rational framework, return comovement should only occur to the extent that assets are exposed to common factors that influence fundamentals. Alternatively, market frictions and behavioral biases can potentially lead to deviations from fundamental value. To the extent that these deviations are correlated across assets, they can cause return comovement.[1] Most tests of excess comovement are attempts to distinguish between these alternative explanations, and are therefore joint tests of comovement and an empirical model of equilibrium asset prices. Residual return correlation in excess of the chosen empirical model is often cited as a contradiction of traditional theory. The implicit assumption behind this interpretation, however, is that unmodeled risk factors have a trivial effect on comovement estimates. On the contrary, we show that this assumption is not as innocuous as it initially seems.

We start by developing a reduced-form model to formally illustrate the impact of an econometrically unobserved risk factor on return comovement. Simulations confirm our models prediction that idiosyncratic returns exhibit substantial comovement in the presence of a latent factor, regardless of the factor's unconditional expected value. Furthermore, residual return comovement is increasing in the variance of the latent factor as well as the number of assets included in each group used to estimate comovement. For example, we obtain a median return comovement estimate of

[1] See Barberis et al. (2005) for an overview of these interpretations.

0.612 for random groups of 80 assets when the volatility of the latent factor equals that of the market portfolio. Even when the latent factor volatility is only 1/8 that of the market, we obtain an average comovement estimate of 0.536. The novel implication of our model is that even relatively inconsequential factors lead to substantive residual return comovement.

Our framework is consistent with the Arbitrage Pricing Theory (APT) of Ross (1976), which allows for an arbitrary number of systematic risk factors. If some factors are unobservable or measured with error, the unobserved component provides a source of common return variation, after adjusting for observable factors. This problem can also be exemplified in a CAPM framework. For example, the Roll (1977) critique posits that the true market portfolio is unobservable. The market factor can then be decomposed into an observable component (e.g., stock market returns) and an unobservable component (e.g., human capital). In empirical tests, idiosyncratic returns with respect to the observed component will continue to exhibit common variation due to common exposure to the unobserved component, leading one to erroneously attribute "excess" comovement to violations of the CAPM.

To illustrate the practical implications of our model, we replicate the primary results for five sources of comovement documented in recent studies. We then show that randomly grouping assets yields within-group return comovement comparable to that of the groupings being replicated. For instance, we find a stock return comovement estimate of 0.636 for firms headquartered in the same Metropolitan Statistical Areas (MSA) in our replication of Pirinsky and Wang (2006). We then perform a bootstrapping procedure in which we randomly assign firms to MSAs and estimate the stock return comovement within randomly assigned headquarter locations. The median estimate produced from 1,000 iterations of this procedure (0.660) is even larger than the comovement estimate for actual headquarter locations. Using a similar approach, we compare comovement estimates for randomly grouped assets to actual groups according to analyst affiliations (Israelsen (2016)), share prices (Green and Hwang (2009)), mutual fund holdings (Anton and Polk (2014)), and prime broker relations (Chung and Kang (2016)). In all cases, the median bootstrapped comovement estimate for randomly grouped assets is comparable to the estimate for the actual groups.

Some studies adjust returns according to a richer empirical model in the hopes of mitigating the potential confounding effects of an omitted risk factor. While controlling for common empirical asset pricing factors attenuates comovement estimates, we find that significant comovement always endures within randomly grouped assets, regardless of the empirical model used to adjust returns and the sample period under investigation. For instance, we obtain a median comovement estimate of 0.079 from 1,000 bootstrap iterations of 160 randomly grouped stocks in the CRSP universe from 1980-2016, even after controlling for the Fama and French five factor model augmented with momentum.[2] This finding suggests that a null hypothesis of zero residual return comovement leads to severe overstatements of "excess" comovement.

Next, we show that grouping assets by characteristics associated with risk significantly intensifies comovement estimates.[3] For instance, we find a comovement estimate of 0.261 for stocks grouped by similarity in market equity (size) after controlling for the Fama-French five factor model augmented with momentum. We obtain qualitatively similar estimates when we group stocks by similarity in book to market, momentum, asset growth, and operating profitability. Even if these characteristics do not proxy for risk, these results suggest that grouping stocks by similarities on observable characteristics leads to substantive comovement estimates. However, similarities on observable characteristics are also likely to lead to similarities in unobservable dimensions. Thus, to draw the conclusion that comovement within a particular group of assets is in "excess" requires the strong assumption that the grouping criteria do not result in the assets having a similar exposure to omitted factors.

While these findings do not rule out behavioral or friction-based explanations of comovement, they do highlight the limitations of commonly used tests and suggest that such interpretations are premature. The alternative framework that we propose offers testable implications for Sharpe ratios that have not been explored in the literature. In particular, excess comovement is defined as covariation between asset returns that is not driven by fundamentals. That is, excess comovement manifests through positive return correlation without an impact on expected return levels. Thus, a portfolio that exhibits excess comovement will have high volatility without a commensurately high expected return. Alternatively, if comovement within a portfolio is due to a priced risk factor, the portfolio will be efficiently diversified and the high volatility will be compensated through high expected returns. Therefore, the Sharpe ratios of comovement portfolios can determine if the risk underlying comovement is priced, and therefore potentially rational, or unpriced.

[2] We also illustrate that industry adjustments, and adjustments in the style of Daniel and Titman (1997), do not fully attenuate comovement estimates.

[3] Thus, our analysis naturally extends to a characteristics based framework (see Daniel and Titman (1997)).

We test whether this proposition holds for our five documented sources of excess comovement. For instance, we build a portfolio of stocks from firms headquartered in each MSA and match each MSA portfolio to a portfolio of firms located outside of the focal MSA.[4] We then compute the ratio of the squared Sharpe ratios between each MSA portfolio and its matched portfolio. Under the null hypothesis of equivalent risk exposure and therefore equivalent expected returns, this ratio follows an F distribution. The alternate hypothesis of a low Sharpe ratio for portfolios with excess comovement can be cast as a rejection of the null. In each of the five settings we consider, we fail to reject the null in more than half of the tested portfolios, and only the ratio for analyst coverage leads to a rejection at conventional levels of significance. Thus, in most settings we fail to reject that the documented comovement from the five sources that we consider is consistent with a risk-based explanation.

A few studies have acknowledged the potential for latent factors to partially influence comovement estimates. Two approaches have been adopted to mitigate this problem: intensity-based sorting (i.e., pairwise return correlations), and shock-based tests. In the intensity based approach, researchers explore whether comovement estimates become stronger as the the grouping mechanism of assets becomes more intense. For example, the strength of comovement has been linked to the degree of common mutual fund ownership (Anton and Polk (2014)) and the distance between firm headquarter locations (Barker and Loughran (2007)). We show that controlling for additional factors always strongly attenuates pairwise return correlations in a variety of settings. These findings are consistent with similarity in characteristics manifesting through common exposure to latent factors, and suggest that an intensity-based design does not circumvent the latent factor bias.

The notion that asset characteristics and returns are jointly determined has likely motivated the shock-based test design, in which researchers identify plausibly exogenous shocks that alter the group to which an asset belongs or the intensity of connections within groups.[5] These tests typically show that within-group correlations become stronger in the destination group after the shock. However, Chen et al. (2016) demonstrate that attributing these results to the proposed mechanism also requires that the comovement with the incumbent group decrease. We find that incumbent

[4]We perform a nearest neighbor match based on market capitalization. Firms in the matched portfolio are not restricted to belong to any particular MSA.

[5]For instance, studies have examined comovement surrounding plausibly exogenous shocks from brokerage house mergers (Israelsen (2016); Chung and Kang (2016)), and S&P500 additions/deletions (Barberis et al. (2005)).

group correlations are usually unaltered by the shock in the settings we consider, suggesting that the mechanisms that lead to excess comovement are not well specified. However, this persistence in incumbent group comovement is consistent with an omitted factor explanation. Simulations of a model with a latent factor show that, if the shocks correspond to an increase in exposure to the latent factor, both incumbent and destination group comovement estimates will increase.

Our paper does not rule out the potential for sources of comovement that cannot be explained by fundamentals, such as informational frictions or market segmentation. However, to our knowledge, we are the first to demonstrate the ubiquitous nature of residual return comovement and the severity of the latent factor bias. We illustrate that controlling for additional empirical factors is unlikely to offer a solution, which suggests a null hypothesis of zero comovement can lead to a severe overstatement of "excess" comovement. Instead, we propose a null of the residual correlation from a benchmark group of stocks, ideally matched on observable criteria to better isolate the proposed source of excess comovement.[6] Furthermore, we are the first to propose a test that exploits the implications of excess comovement for portfolio diversification in the presence of latent factors. Using our approach, we find mixed evidence of excess comovement based on the settings proposed in the literature.

The rest of this paper is organized as follows. In Section 2, we provide an overview of the comovement literature. In Section 3, we provide a theoretical motivation for the relationship between omitted risk factors and comovement, and we develop the distributional properties of our Sharpe ratio test. In section 4, we describe our simulations and replications, test the implications of our model, and discuss the results. Section 5 concludes.

2 Return Comovement

The fundamental question underpinning studies of return comovement is whether observed levels of comovement are consistent with predictions of traditional asset pricing theory. On one hand, common variation in returns across securities could be a result of rational variation in investors time preferences or in the prospective cash flows of the underlying assets. Alternatively, common

[6] e.g., to illustrate excess comovement in geography, one should compare return correlations within each MSA to a portfolio of stocks matched (by size or other characteristics) to those in the focal geography, but that are headquartered elsewhere.

variation could be driven by deviations from fundamental value that are correlated across assets. Early tests of the theory were conducted under the assumption of constant discount rates, and considered whether asset prices were too volatile relative to the volatility of their cash flows or dividends (e.g., Shiller (1983)). Subsequent work challenged the validity of these findings since most asset pricing theories do not require constant discount rates (see Kleidon (1988); Cochrane (1991); Fama (1991)).

Later studies focused on specific assumptions of the traditional theory, including that of well-informed rational investors, perfect competition, and complete financial markets. Barberis et al. (2005) propose three explanations for comovement that rely on frictions or irrational investor behavior: the category view, the habitat view, and the information diffusion view. The category view posits that investors allocate funds across categories of assets, rather than individual assets and the habitat view asserts that transaction costs, trading restrictions, or lack of information cause investors to invest only in a subset of assets. Both category- and habitat-based investment can lead to correlated investor demand, which can induce excessive common variation in the returns of assets within categories/habitats. Chen et al. (2016) refer to the category and habitat views collectively as an asset class effect. Finally, the information diffusion view is based on the non-synchronous incorporation of common information into assets which potentially leads to excess comovement.

Most of the subsequent literature on comovement can largely be classified as interpreting evidence in light of one of the alternative explanations proposed by Barberis et al. (2005). For instance, Greenwood (2008) finds evidence that stocks that are overweighted in the Nikkei 224 index exhibit excess comovement with other stocks in the index, and comove less with stocks outside of the index. Kumar and Lee (2006) find that correlation in retail trading explains return comovement for stocks with a high concentration of retail traders. Anton and Polk (2014) find that excess comovement is related to common mutual fund ownership. Pirinsky and Wang (2006) find that stocks of firms headquartered in the same Metropolitan Statistical Area (MSA) exhibit excess comovement and Green and Hwang (2009) find excess comovement for stocks with a similar price range. These studies advance some variant of the explanation that excess comovement is caused by correlated sentiment or liquidity needs, and thus interpret their findings as evidence of an asset class effect.

A variety of studies also interpret their evidence in support of the information diffusion view. Grullon et al. (2014) find excess comovement in the stock prices of firms with common lead underwriters. The authors claim that investment banks serve as a conduit of information flow between firms and investors, which leads to segmented sets of investors who hold similar stocks and have access to similar information. Similarly, Chung and Kang (2016) claim that prime brokers provide valuable, and common, information to their hedge fund clients, which induces comovement in clients' returns who trade on this information. Hameed et al. (2015) claim that stocks with more extensive analyst coverage are priced more accurately, and that such "bellwether" stocks lead the price discovery of related firms. In turn, this information spillover will cause opaque stocks to comove more strongly with "bellwether" stocks.

All tests of excess comovement are joint hypotheses between the asset pricing theory and an empirical model of returns. For their interpretation to be valid, the empirical model used to adjust returns would have to capture all rational variation in expected returns. The limitation of this joint hypothesis problem is that investors have more information about the factors that drive returns, exposure to those factors, and anticipated changes in those factors than are directly observed by the researcher. Our paper contributes to this literature by proposing a test of excess comovement that accounts for the potential presence of latent factors related to unobserved information.

3 A Latent Factor Explanation

Under the assumptions of the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), covariance with the market portfolio completely determines the risk of a security. More generally, the Arbitrage Pricing Theory (APT) of Ross (1976) allows for an arbitrary number of systematic factors to be associated with risk. However, the theory does not provide a method for identifying all relevant factors. To illustrate how the presence of an un-modeled factor in an APT framework can affect tests of comovement, we consider the following Data Generating Process (DGP) for asset returns:

$$r_{it} - r_{ft} = \beta_i F_t + \gamma_i Z_t + \epsilon_{it} \tag{1}$$

where $r_{it} - r_{ft}$ is the excess (over the risk free rate) return of stock i at time t, F_t and Z_t are the realizations of the orthogonal market-wide factors at t, and ϵ_{it} is an idiosyncratic disturbance. The terms β_i and γ_i are constant for each stock i. To avoid degenerate cases, we will assume that the coefficients β and γ are relatively close to unity with the average cross sectional values of each being 1.

We further assume that there are N stocks in the economy and that G denotes a partition of the set $I = \{1, 2, \dots, N\}$ such that the G_g is the g^{th} element of G. The partition G is the mathematical equivalent of creating a subset of stocks. For example, G can represent grouping stocks by industry classifications, geographical locations, market capitalization or any observable criterion. To provide some intuition for the model, we can think of F_t as observable and Z_t as unobservable. For instance, in the spirit of Roll (1977), F_t may represent the observable component of the market portfolio (i.e., value-weighted return of all stocks in CRSP) and Z_t can represent the unobservable component (e.g., human capital).

Common tests of comovement consider the relationship between each stock's return and the average return of all stocks in the group. The method excludes own returns from the average calculation to avoid spurious correlations. In our setting, we can change the subscripts in the DGP to include a group subscript:

$$r_{igt} - r_{ft} = \beta_i F_t + \gamma_i Z_t + \epsilon_{igt} \tag{2}$$

to indicate that stock i belongs to group G_g. We then calculate group averages:

$$r_{-igt} - r_{ft} = \frac{1}{N_g - 1} \sum_{j \in G_g, j \neq i} r_{jgt} - r_{ft}$$

where N_g is the number of stocks in G_g. Then, the level of comovement driven by the partition G can be assessed through the relationship between $r_{igt} - r_{ft}$ and $r_{-igt} - r_{ft}$ after controlling for observed market exposure.

In order to assess how this estimation would work under our assumptions, we define $\beta_{-i} = \frac{1}{N_g-1} \sum_{j \in G_g, j \neq i} \beta_{jgt}$, $\gamma_{-i} = \frac{1}{N_g-1} \sum_{j \in G_g, j \neq i} \gamma_{jgt}$, and $\epsilon_{-igt} = \frac{1}{N_g-1} \sum_{j \in G_g, j \neq i} \epsilon_{jgt}$. It is clear that

$$r_{-igt} - r_{ft} = \beta_{-i} F_t + \gamma_{-i} Z_t + \epsilon_{-igt}.$$

Likewise, we define the average factor loadings $\bar{\beta} = \frac{1}{N} \sum_i \beta_i$, $\bar{\gamma} = \frac{1}{N} \sum_i \gamma_i$, and $\bar{\epsilon}_t = \frac{1}{N} \sum_i \epsilon_{it}$. Then the returns on (the equally-weighted) market portfolio satisfy:

$$r_{mt} - r_{ft} = \bar{\beta} F_t + \bar{\gamma} Z_t + \bar{\epsilon}_t.$$

Estimating the model:

$$r_{igt} - r_{ft} = a + b(r_{mt} - r_{ft}) + c(r_{-igt} - r_{ft}) + e_{igt}$$

is equivalent to estimating:

$$r_{igt} - r_{ft} = a + (b\bar{\beta} + c\beta_{-i})F_t + (b\bar{\gamma} + c\gamma_{-i})Z_t + b\bar{\epsilon}_t + c\epsilon_{-igt} + e_{igt}.$$

Taking expectations, we get:

$$E_t[r_{igt} - r_{ft}] = a + (b + c)F_t + (b + c)Z_t.$$

Our assumption about the cross sectional average of β and γ combined with the standard Gauss-Markov assumptions imply that the true values of b and c satisfy $b + c = 1$. Unbiased estimates of the coefficients b and c will therefore reflect this identity as well. Note that the model does not identify the parameter c, and that for any partition G we will receive an estimate of c that is not necessarily 0. A positive estimate of c therefore does not signify "excess" comovement between the constituents of each group G_g, since in the presence of some unobserved factor Z_t any group will exhibit some comovement.

3.1 Portfolio variance test

It is important to note that the unobserved factor Z_t in Eq. (1) can be an unpriced factor that does not carry a premium. That is, $E[Z_t] = 0$. For the remainder of this section, we will proceed with the case that $E[Z_t] = 0$ to simplify exposition, noting that the assumption is not necessary for the results that we derive. We can express Eq. (1) in vector form:

$$r_t - r_f\mathbf{1} = F_t B + Z_t \Gamma + \epsilon_t \tag{3}$$

where $r_t = [r_{1t}, r_{2t}, \ldots, r_{n_t}]'$, $B = [\beta_1, \beta_2, \ldots, \beta_n]'$, $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n]'$, $\epsilon_t = [\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{n_t}]'$, and $\mathbf{1}$ is a vector of ones.

Thus far, we have motivated Z_t as a latent unpriced risk factor in an APT framework. Thus, comovement due to this factor is consistent with a rational market paradigm. However, nothing in the current model requires the factor Z to represent a rational risk source. If Z were instead a latent factor relating to average sentiment of investors, for instance, then exposure to Z would indeed provide a source of comovement in excess of fundamentals. In other words, the difference between rational and sentiment or friction based explanations of comovement center around the source of Z_t.

In the case that Z is driven by behavioral biases or market frictions, the typical argument of excess comovement can be recast in terms of Eq. (3), by noting that the presence of excess comovement is equivalent to having a particular subset of assets exposed to the factor Z_t. To operationalize this hypothesis, we test if the coefficients Γ that correspond to this group are indeed different from zero and have the same sign.[7] However, the factor Z_t is unobservable and we cannot estimate the coefficients Γ directly. We can however consider the variance of a portfolio with weight vector $w = [w_1, w_2, \ldots, w_n]'$. The return on the portfolio will be $w'r_t$ and its variance will be $\sigma_p^2 = w'BB'w\sigma_F^2 + w'\Gamma\Gamma'w\sigma_Z^2 + \sum_{i=1}^{n} w_i^2\sigma_i^2$, where we make the standard assumption that $E[\epsilon_{it}\epsilon_{jt}] = 0, \forall i \neq j$, and denote $E[\epsilon_{it}^2] = \sigma_i^2$.

Let us consider two groups of stocks: group A consists of all stocks that share a common feature that drives comovement, and group B is an identical group of stocks that do not share this feature. To distinguish the two groups, assume that the portfolio weights for group A (B) are w^A (w^B). Further assume that the portfolio is a long only portfolio, that is $w_i^A \geq 0$ and $w_i^B \geq 0, \forall i$. Under the assumption that the two groups are identical in every aspect except excess comovement, we can formulate the following hypothesis:

$$H_0 : \Gamma' w^A = \Gamma' w^B,$$

$$H_A : \Gamma' w^A > \Gamma' w^B \geq 0.$$

More generally, we can write the hypothesis as:

$$H_0 : (\Gamma' w^A)^2 = (\Gamma' w^B)^2,$$

[7] Our test focuses on the square of a weighted average of the coefficients Γ. Therefore, we can structure our hypothesis as a test of the Γ coefficients being all positive without any loss of generality.

$$H_A : (\Gamma' w^A)^2 > (\Gamma' w^B)^2.$$

Note that the alternate hypothesis implies

$$H_A : \widehat{var(r_t' w^A)} > \widehat{var(r_t' w^B)}.$$

Under the null hypothesis we have:

$$(T-1)\frac{\widehat{var(r_t' w^A)}}{\sigma_p^2} \sim \chi^2(T-1),$$

and

$$(T-1)\frac{\widehat{var(r_t' w^B)}}{\sigma_p^2} \sim \chi^2(T-1),$$

so that:

$$\frac{\widehat{var(r_t' w^A)}}{\widehat{var(r_t' w^B)}} \sim F(T-1, T-1),$$

We can simply test the alternate hypothesis that this ratio is greater than 1.

An extension of this test concerns Sharpe ratios. Given that excess comovement arguments assume that the expected returns of the assets under study do not depend on the level of comovement, it is safe to assume that under the expected returns of the two portfolios A and B are identical ($E[r_t' w^A] = E[r_t' w^B]$). If we assume that the cross section of stocks is large enough, the two portfolio returns will be identical. Under these asymptotic assumptions, the ratio of the Sharpe ratios of the two portfolio will also have the same distributional properties as the ratio of portfolio variances.

4 Empirical Analysis

4.1 Simulations

We start by exploring the properties of traditional tests of comovement in the presence of a latent factor by using simulated data. We simulate the underlying data generating process to follow our analysis in Section 3. In particular, we simulate:

$$r_{it} = F_t\beta_i + Z_t\Gamma_i + \epsilon_{it} \tag{4}$$

where we calibrate the simulation to match the real data as closely as possible.[8] In particular we simulate the ϵ_{it} to be distributed i.i.d $N(0, 0.183)$, where 18.3% is the average market adjusted monthly return volatility in the CRSP universe from 1980-2016. The market factor F_t is simulated as an AR(1) process with a mean of 0.649%, a standard deviation (σ_F) of 4.52%, and an auto-correlation coefficient of 0.0863. The β_i and Γ_i are each distributed with a cross sectional average of 1 and cross-sectional standard deviation of 0.45, which match the distribution of $\hat{\beta}_i$ from CAPM regressions in the CRSP universe from 1980-2016. $Size_{i0}$ (market capitalization at time 0) is simulated via an exponential distribution and grows each year by $(1 + r_{it})$.[9] We simulate Z_t as an AR(1) process with mean 0, and auto-correlation coefficient of 0.0863. We repeat these simulations for different values of σ_Z.[10]

After simulating the data, we assign stocks to random groups of size $N_g = 10, 20, 40, 80$, and 160 and estimate:

$$r_{igt} = \alpha + \beta r_{mt} + \theta r_{-igt} + \epsilon_{igt} \tag{5}$$

where $r_{mt} = \sum_i \frac{size_{it} * r_{it}}{\sum_i sizeit}$ is the estimated market return for time t.

Table 2 reports simulation results of Eq.(5) for 240 months of returns for 2,400 assets. To explore the sensitivity of comovement estimates to sorting on observable characteristics that proxy for risk exposure, we generate a characteristic $X_i = \rho\Gamma + (1 - \rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$ for each asset. We form groups by sorting on values of X_i and analyze within group comovement for different values of ρ. When $\rho = 0$, this procedure amounts to forming groups randomly. Greater values of ρ indicate that the procedure sorts more strongly on exposure (Γ_i) to the latent factor Z. Each column corresponds to a different value of ρ and each panel corresponds to a different value of σ_Z, expressed as a multiple of σ_F. The rows of each panel correspond to simulations produced with different asset group sizes (N_g). The median estimate of θ from 1,000 simulations is reported for each specification.

It is worth noting that even for the case of $\sigma_Z = 0$ and $\rho = 0$ (Panel A, Column 1), the

[8] For simplicity we exclude the risk free rate from these simulations.

[9] For robustness, we also use a lognormal, and a normal distribution to simulate size.

[10] We also explore variations in our presumed data generating process in the Internet Appendix.

median estimate of θ from 1,000 simulations is quite substantive. This result is consistent with the unobservable nature of the market factor (F_t) as described by the famous Roll critique (Roll (1977)). Value-weighted market returns provide a noisy proxy for F_t, and the unobserved components of this proxy contribute to covariances across assets that are not captured by controlling for market returns. Consequently, our assumption of an omitted factor in a multi-factor model is not necessary to generate substantive residual return comovement. Imperfect proxies for the market factor are sufficient, since the unobserved component of the market can serve as a latent factor.

Focusing on the panel corresponding to the factors F_t and Z_t having equal volatilities ($\sigma_Z = \sigma_F$), the median estimate of θ is monotonically increasing in ρ. For large groups, the estimate increases from 0.758 when $\rho = 0$ to 0.842 when $\rho = 1$. This difference in comovement estimates can be interpreted as the effect of sorting on exposure to the omitted factor. The same pattern is observed in all panels. These findings have practical implications. In particular, these finding suggest that characteristic based groups are likely to lead to higher estimates of comovement when the characteristics are even mildly associated with exposure to risk.

Table 1 provides ordered statistics for estimates of θ from simulations with 1000 iterations each for different values of σ_Z with $\rho = 0$ (i.e., purely random sorts). Each column of Panel A corresponds to a different parametrization of the model through a different value for the volatility of Z_t (σ_Z). Each row of Panel A corresponds to simulations produced with different asset group sizes (N_g). The median coefficient estimates of θ range in value from 0.127 to 0.966, and increase monotonically in both N_g and σ_Z. This finding reinforces the intuition of our model in Section 3 that θ increases as the omitted factor constitutes a higher fraction of total return variance. Similarly, as N_g increases, idiosyncratic returns are diversified away and shared exposure to the omitted factor becomes more prominent. This effect also provides intuition for why higher values of N_g generally result in tighter confidence intervals. Overall, estimates are positive and significant in all simulations, with the lowest median estimate of 0.127 corresponding to $\sigma_Z = 0$ and $N_G = 10$. None of the confidence intervals include the value of zero in our simulations for any N_g or σ_Z.

Of course, we have not provided an economic motivation for the source of factor Z_t. We have merely modeled Z_t as a latent risk factor in an APT framework. Nothing in our setting rules out that Z_t could arise because of behavioral biases or market frictions. The purpose of these simulations is to highlight 1) an imperfect proxy for the market factor is sufficient to generate substantial residual return correlation and 2) even inconsequential latent factors can amplify this effect. Investors will always have more information than the econometrician regarding the factors that drive returns, exposure to those factors, and anticipated changes to those factors. Thus, the practical implications of these simulations suggest that the econometrician cannot distinguish between alternative explanations for residual return comovement. The existence of a latent factor, regardless of its importance, will lead to substantial and ubiquitous within group return comovement.

4.2 Data and summary statistics

We collect monthly return data from the Center for Research in Security Prices (CRSP). In addition to returns, we also collect share prices, market capitalizations, and historical adjustment factors for each stock in our sample. For comparability across the settings that we consider, we restrict our sample to January, 1970 to December, 2016. In some of the settings that we analyze, the sources of comovement are only available after 1980, and we restrict our sample accordingly. Most of our analysis on stock returns is conducted at the monthly frequency. However, some of our analysis requires daily CRSP data on common shares of stocks.

In Section 4.6, we show that stock characteristics play an important role as determinants of comovement. To construct these characteristics, we use financial statement data from the annual Compustat database. These data are combined with the CRSP return data such that elements reported as of December of year t are matched to the returns for July *t+1* through June *t+2*. All Compustat annual data are obtained for 1968 through 2016 to match our CRSP sample. Panel A of table 1 summarizes the main sample. The average excess return for the sample is about 0.7% with a median of about −0.4%. The average firm has a market capitalization slightly above $1 billion, and a book-to-market equity ratio of 0.77.

Headquarter locations are determined through addresses filed with the Securities and Exchange Commission (SEC), and are obtained through Wharton's Data Research Center (WRDS) SEC analytics suite. The SEC analytics suite parses the 10-K annual filings from the SEC's EDGAR service to codify and store the standard fields of each filing. Compustat also stores addresses, but does not maintain a history of changes to that field. The SEC's EDGAR service provides all filings from 1994 to 2016, which restricts the sample of firm headquarter locations to that time period.[11]

[11] Some studies have used alternative sources for headquarter locations. However, we do not have access to these

Following Pirinsky and Wang (2006), we aggregate firm headquarter locations to the Metropolitan Statistical Area (MSA) levels. We use the Census bureau's 2010 ZIP Code Tabulation Area (ZCTA) Relationship files to assign firms in our sample to MSAs.

Analyst coverage comes from the Thomson Reuters IBES database. Each year, we pair analyst i and firm j if analyst i issued at least one report covering firm j in year t. IBES data are available from 1970 to the end of our sample in 2016. Mutual fund equity holdings are obtained from Thomson Reuters Mutual Fund Holdings database. These data allow us to create a mapping between firms and mutual funds following Anton and Polk (2014), and are available from 1979 to 2016. The Mutual Fund Holdings data are collected from the 13-F filings of institutional investors. For each filing, we construct a mapping between stock i and a mutual fund j if stock i appears in the holdings of mutual fund j. Finally, we use the Thomson Reuters Lipper Hedge Fund Database (commonly referred to as TASS) to calculate hedge fund returns as well as to identify the funds that share a common prime broker. The TASS data are available from 1990 to 2016.

Panel B of Table 1 summarizes asset returns and asset group characteristics. For headquarter location, all companies whose headquarter is located within the same MSA are assigned to the same group. On average there are 174 firms in each MSA, and the average return for the sample stocks is 0.89% per month. The sample with analyst coverage has an average return of 0.85% per month. Each stock shares at least one analyst with 68 other stocks on average. For groups formed according to stock price level, we use a sample of monthly CRSP returns from 1926 to 2016. The group for stock i consists of all stocks within 25% of the stock price i . Using this definition, a typical stock is related to 589 stocks in our sample. Two stocks are very likely to be held by at least one common mutual fund, since the average stock is related to about 1,185 stocks in our sample. Lastly, hedge funds share a prime broker with 135 other funds on average, and the average excess returns for hedge funds in our sample is 0.47% over our sample.

4.3 Replications

In this section, we describe our replication of five recently published articles on excess comovement. In particular, we replicate the primary results from Pirinsky and Wang (2006), Green and Hwang (2009), Anton and Polk (2014), Israelsen (2016), and Chung and Kang (2016). While there are

sources. Headquarter locations change very infrequently and should not impact our results.

certainly more than five candidate papers for replication that identify sources of excess comovement, we chose to replicate a set of papers that span a variety of settings and asset classes, and for which we have access to the data. Furthermore, we restrict our replications to recently published articles in the Journal of Finance, Journal of Financial Economics, Review of Financial Studies, and Journal of Financial and Quantitative Analysis.[12]

The studies that we selected examine different sources of comovement that are related to investor behavior and information dissemination. Specifically, Pirinsky and Wang (2006), Green and Hwang (2009), Anton and Polk (2014), and Israelsen (2016) examine comovement in stock returns due to common firm headquarter location, similar share prices, common mutual fund ownership, and common analyst coverage, respectively. Chung and Kang (2016) document comovement in the returns of hedge funds that share the same prime broker. With the exception of Israelsen (2016), these studies attribute their results to violations of rational market behavior. The violations that these studies emphasize stem from behavioral biases of investors and/or information processing channels. For instance, Pirinsky and Wang (2006) and Green and Hwang (2009) claim that stock markets are segmented by geographical proximity and share price similarity for reasons that are not associated with risk. This segmentation, in turn, causes returns to comove beyond what commonality in fundamentals would warrant. The remaining studies contend that analysts, mutual funds, and prime brokers use the same sources of information to price assets. As a result, commonality along these dimensions leads to similar trading behavior and therefore excess comovement.

We report results in Table 3 that correspond to the closest replication that we could produce for each of the five studies described above. The five studies cited use different methodologies to circumvent many of the difficulties associated with detecting excess comovement. For comparability, we start by employing the same parsimonious specification that encapsulates the spirit of these studies:[13]

$$r_{igt} = \alpha + \theta r_{-igt} + \beta r_{mt} + \epsilon_{igt} \tag{6}$$

where r_{igt} represents the excess return for asset i in year t, r_{-igt} represents the average excess return of all assets in the same group g as asset i, excluding asset i from its own group return

[12]These Journals are the four pure finance journals with the highest impact factors.

[13]In unreported results, we closely replicate the exact specification used in each study and obtain qualitatively similar results to those in the original papers.

calculation, and r_{mt} is the excess return on the market. A positive θ estimate is commonly referred to as excess comovement in the existing literature. Each replication amounts to using a different grouping criterion. We group assets by headquarter location in Panel A (Pirinsky and Wang (2006)), common analyst coverage in Panel B (Israelsen (2016)), similar stock price levels in Panel C (Green and Hwang (2009)), common mutual fund ownership in Panel D (Anton and Polk (2014)), and common prime broker in Panel E (Chung and Kang (2016)).

In all five settings, the estimate of the coefficient θ are positive and both statistically and economically significant ranging from 0.09 to 1.03. It is worth noting that in Panels C and D, we make a slight modification to cast all of the analysis in a consistent manner. For the replication of Green and Hwang (2009), we analyze the relationship between stock i's returns and those of a portfolio of all stocks within 25% of stock i's price. For our replication of Anton and Polk (2014), the comparison portfolio for stock i consists of all stocks held by at least one mutual fund that also holds stock i.

4.4 Comovement for randomly grouped assets

As derived in our theoretical motivation in Section 3 and shown in our simulations in Section 4.1, omitted factors can lead to substantive residual return comovement. In light of this implication, a null hypothesis of zero will lead to an overstatement of excess comovement. Instead, a more appropriate null would be the comovement exhibited by a randomly selected group of assets. A random group that is unrelated to the source of comovement being studied will account for common exposure to the omitted factor(s) and therefore provide a more appropriate benchmark. To assess whether the coefficient estimates of our replications in Table 3 provide evidence of excess comovement, we compare the replicated estimates to those obtained from randomly grouped assets, keeping the number of assets per group fixed.

More specifically, for each replication we employ a bootstrap procedure whereby the economically-motivated asset groups are replaced by a randomly selected group of assets. For instance, in the Pirinsky and Wang (2006) replication, we randomly assign Metropolitan Statistical Areas (MSA) to firms and estimate the return comovement of firm i with that of firms randomly assigned to the same MSA. In a similar fashion, we randomly assign analyst affiliations (Israelsen (2016)), share prices (Green and Hwang (2009)), mutual fund holdings (Anton and Polk (2014)), and prime broker relations (Chung and Kang (2016)) to the assets in our sample. We then compare the excess return of asset i with the average excess return of its randomly formed asset group. For each panel, this procedure is repeated 1,000 times to construct a bootstrapped null distribution to compare to the replicated coefficient estimate.

The confidence intervals of the bootstrapped null for each replication are reported in the right half of Table 3. For each panel, the corresponding bootstrapped confidence intervals at the 1%, 5%, and 10% levels are reported, as well as the median value from all sample runs. Consistent with our theoretical model, the confidence intervals do not contain zero for any of the settings we study. These results are highly consistent with the presence of an omitted factor or imperfect market proxy that causes the returns of seemingly unrelated assets to comove. Moreover, common mutual fund ownership is the only source of comovement in our replications that exceeds the 10% upper bound of the corresponding bootstrapped confidence interval. However, even for common mutual fund ownership, the bootstrapped confidence interval suggests that the original coefficient estimate severely overstates "excess" comovement. In other words, a random sampling of assets produces comovement estimates that are not statistically different (at conventional levels) from those of the economically-motivated groupings being replicated.

To explore the time-series of residual return comovement, we repeat our analysis separately for each year from 1980-2016. We plot median comovement estimates from 1000 iterations in Figure 1 of the Internet Appendix. Comovement estimates are obtained by regressing market adjusted returns on groups of randomly selected stocks in the CRSP universe from 1980-2016. We repeat the analysis for groups containing 10, 20, 40, 80, or 160 randomly selected stocks. Median comovement estimates and confidence intervals are obtained from 1000 iterations. From the figure, it is clear that residual return comovement is substantially greater than zero during all time periods and for all group sizes. Furthermore, the figure also illustrates that comovement estimates increase with group sizes for every time period.

The substantive comovement estimates for randomly grouped assets are troubling for several reasons. First, these results suggest that a null hypothesis of zero can lead to severe overstatements of "excess" comovement, and a bootstrapped null should be used instead. Second, the fact that randomly grouped assets exhibit the same level of comovement as economically-motivated sources from the literature suggests that documented estimates need not be driven by the proposed explanations. For instance, our bootstrapped confidence intervals suggest that comovement within stocks grouped by headquarter MSA can be entirely explained by factors unrelated to location. The results suggest similar takeaways for the other sources of comovement that we consider. Thus, the economic motivation for the proposed grouping of assets does not contribute to our understanding of excess comovement without the appropriate counter-factual.

4.5 Adjusting for multi-factor models

The models presented in Table 3 only use excess market returns as a control variable and are therefore analogous to using CAPM adjusted returns. However, some studies have acknowledged the potential for omitted factors to partially drive excess comovement estimates. To mitigate this concern, these studies often control for multi-factor models that perform better than the single factor CAPM in explaining asset returns. The goal of this process is to reduce the confounding effects of an omitted factor and isolate the residual correlation in returns that are due to the economically-motivated grouping criteria. If controlling for additional factors sufficiently accomplishes this task, however, the comovement estimates for randomly grouped stocks should be driven to zero. In this section, we explore whether controlling for more factors alters our conclusions from Section 4.4.

We start our analysis by randomly assigning stocks to groups using monthly data for all CRSP/Compustat firms with common stock from 1970 to 2016. Each stock is randomly assigned to one group for the duration of the sample period. Similar to our analysis in Section 4.4, we regress the risk adjusted excess return of asset i on the average risk adjusted excess return of its randomly formed asset group (excluding asset i). We repeat this bootstrap procedure 1,000 times each for randomly assigned asset groups consisting of 10, 20, 40, 80, and 160 stocks per group.

The average coefficient estimate from 1,000 bootstraps are presented in Panel A of Table 4. Each row of Panel A corresponds to different group sizes (N_g) ranging from 10 to 160, with each subsequent group containing twice the number of stocks as the previous group. In Column 1, we use (raw) excess returns. Each subsequent column reports results for adjusted returns according to the Capital Asset Pricing Model (CAPM), the Fama-French three and five factor models (3 FM and 5 FM), and the Fama-French five factor model augmented with the momentum factor (6 FM).

For each column of Panel A, the comovement estimates exhibit a monotonic relationship that increases with the number of stocks used to form each group. For instance, when raw excess returns are used, the average comovement estimate from 1,000 draws increases monotonically from 0.47, when there are only 10 stocks per group, to 0.93, when there are 160 stocks per group. For each row, the different risk adjustment models also produce significantly different comovement estimates. This finding is consistent with the pattern found in our simulations in which, as the group size becomes larger, idiosyncratic returns will be diversified away and and shared exposure to the omitted factor becomes more prominent. Using the CAPM-adjusted returns yields an average comovement estimate of 0.28, and the estimate attenuates to 0.08 when the six factor model (6 FM) is used. However, for all portfolio sizes and all factor models used to adjust returns, the estimates of comovement remain positive.

In Columns 6-8 of Table 4, we extend our analysis to adjust returns for principal factors from an ex post principal component analysis using the first five (PCA5), ten (PCA10), and twenty (PCA20) factors. Adjusting returns for the first ten principal factors yields a residual comovement estimate of 0.1048 for the groups containing 160 randomly selected stocks. It isn't until we adjust returns for the first 20 principal factors that we obtain inconsequential comovement estimates. This finding suggests that the omitted factor bias is quite pervasive and controlling for a few empirical factors is not sufficient to rule out an omitted factor explanation of comovement.

The results presented in this section lead to a few important takeaways. First, adjusting returns for additional factors always attenuates comovement estimates for randomly selected groups of stocks. This finding is consistent with an omitted factor explanation of excess comovement. Second, regardless of the factor model used to adjust returns, randomly grouped assets always appear to exhibit positive comovement estimates. These positive estimates demonstrate that existing empirical factors fail to capture all significant, common cross-sectional variation in stock returns. Given that these factors perform well in identifying cross-sectional risk premia, these results suggest that a small residual component in returns can lead to significant positive comovement, on average, for any subset of assets. Thus, comovement is ubiquitous, and tests of excess comovement appear to suffer from a severe form of the joint hypothesis problem discussed in Fama (1991). This finding also reiterates the importance of conducting comovement tests with a nonzero null.

4.6 Characteristic sorts

An alternative approach to using empirical factor models is to adjust returns according to asset characteristics, which have been associated with returns (see Daniel et al. (1997)). To the extent that factor models fail to capture the predictability in cross-sectional returns due to empirical factors, commonality among assets along these characteristics will be closely related to comovement. In this section, we quantify the extent to which these characteristics lead to excess comovement relative to the factor models that we present in Section 4.5.

For this analysis we consider five of the most common characteristics associated with risk: Size, Book-to-Market (B/M), Momentum, Asset Growth, and Operating Profitability. With the exception of momentum, these characteristics correspond to the sources of risk explored in Fama and French (2015). For each characteristic, we form groups of 10, 20, 40, 80, and 160 stocks based on sorts of the focal characteristic. For example, for the specification involving the size characteristic and a group of 10 stocks, we place the ten smallest market cap stocks in group 1, the next smallest ten stocks in group 2, and so on. For each stock, we then regress its residual from a factor model on the average residuals of all other members of its group, excluding the focal stock.

We present the results from this exercise in Panels B-F of Table 4. For all characteristics, the comovement estimates remain substantially higher than those of the corresponding return adjustment for randomly grouped stocks in Panel A. Grouping stocks by the momentum characteristic yields the highest comovement estimates across all specifications. For groups of 160 stocks, the momentum characteristic exhibits a comovement estimate of 0.28 for returns adjusted for the six factor model (6 FM). It is worth noting that the six-factor model includes the momentum factor up-minus-down (UMD) to adjust returns. Similarly, all five characteristics we consider exhibit positive comovement estimates, despite controlling for an empirical factor that corresponds to the focal characteristic (i.e., adjusting returns according to the six-factor model).

In Columns 6-8 of Table 4, we adjust returns for the first five (PCA5), ten (PCA10), and twenty (PCA20) principal factors. The comovement estimates under these specifications continue to exhibit a strong positive relationship and are generally of the same magnitude as the estimates derived from the six factor model. Operating Profitability produces the lowest comovement estimates of 0.08, even after adjusting for the first 20 principal factors.

One interpretation of the findings in Table 4 is that comovement within groups based on similarity in characteristics reflects similar exposure to unobserved factors. Another interpretation is that linear factor models contain more measurement error than characteristic sorts. Both interpretations highlight the importance of characteristics as determinants of return comovement. Even if these characteristics do not proxy for risk, these results suggest that grouping stocks by similarities on observable characteristics leads to substantive comovement estimates. However, similarities on observable characteristics are also likely to lead to similarities in unobservable dimensions. Thus, to draw the conclusion that comovement within a particular group of assets is in "excess" and due to a proposed source (e.g., correlated sentiment) requires the strong assumption that the grouping criteria do not result in the assets having a similar exposure to omitted factor(s) or similar characteristics. The results in Table 4 also reaffirm the inadequacy of the zero null hypothesis. A more prudent test of excess comovement would use a benchmark from a portfolio of stocks matched on observable characteristics, which will help account for common exposure to omitted factor(s) or sources of comovement that differ from the proposed source.

4.7 Sharpe ratio tests

In this section we reformulate tests of excess comovement to focus on implications beyond residual return correlations, which can have several causes. In particular, we exploit different implications between an omitted factor explanation of comovement and comovement that is driven by behavioral biases or informational frictions. In particular, excess comovement is defined as covariation between asset returns that is not driven by fundamentals. That is, excess comovement manifests through a non-zero correlation between returns without an impact on expected return levels. Under friction-based explanations, within-group excess comovement is an indication that investors are not efficiently diversified (i.e., a portfolio with only within-group assets will exhibit a low expected Sharpe ratio).

In contrast, a rational investor facing fewer informational frictions would be able to diversify more effectively. Thus, if an omitted risk factor is the reason for comovement, then all diversified portfolios will have the same expected Sharpe ratio. For example, suppose investor A elects to overweight stocks of firms headquartered in his/her MSA, while investor B holds a more geographically dispersed portfolio. Then investor A is restricting his/her diversification benefits compared to investor B.

To formalize this intuition, we propose a squared Sharpe ratio test in the spirit of Gibbons et al. (1989). Continuing with the example of investors A and B, we will refer to the squared Sharpe ratio of investor A's portfolio as s_A^2 and that of investor B's portfolio as s_B^2. Assuming that the two portfolios contain the same number of relatively similar assets, then under the null hypothesis of no excess comovement, the two expected squared Sharpe ratios will be equivalent. Thus, under the null, $s_A^2 = s_B^2$, or $s_A^2/s_B^2 = 1$. However, under the alternative hypothesis that excess comovement exists for the assets in investor A's portfolio: $s_A^2 < s_B^2$, or $s_A^2/s_B^2 < 1$. We showed in Section 3.1 that a variant of this statistic exhibits an F distribution under the null hypothesis of equivalent expected sharp ratios (a ratio of 1).

We test whether this proposition holds for each of the five documented sources of excess comovement that we consider in Table 3. For instance, we build a portfolio of stocks of firms headquartered in each MSA and match each MSA portfolio to a portfolio of firms located outside of the focal MSA. For each stock in each MSA, we find the nearest neighbor match based on market capitalization. The potential matching pool consists of all firms headquartered outside of the focal stock's MSA, but are not otherwise restricted to belong to any particular location. Note that only matching on firm size is a fairly lenient restriction, and that firms clustered on observable dimensions are likely to be similar across many different characteristics. One could easily extend our analysis to impose additional restrictions, such as belonging to the same industry or a particular bin in a multi characterstic sort.

Table 5 presents results from this analysis. For each of the settings that we consider, we construct equally- and value-weighted portfolios. We form portfolios of assets according to common headquarter location (Panel A), common analyst coverage (Panel B), similar share price (Panel C), common mutual fund ownership (Panel D), and common prime broker (Panel E). For each grouping criterion we report the mean (median) squared Sharpe ratio, the t-stat, and the number of portfolios for which we reject the null of unity. The last column of Table 5 presents the number of test portfolios (N).

In many of the settings that we consider, both the average and median ratio of squared Sharpe ratios are close to unity. Analyst coverage and prime broker connections provide the largest deviations of the ratio from one with an average ratio of 1.59 and 1.56 (1.64 and 1.54) for equally-weighted (value-weighted) portfolios, respectively. In these settings, we reject the null $36-37\%$ of the time for common analyst portfolios and $33-44\%$ of the time for prime broker portfolios.

Table 6 presents additional tests of the variances and Sharpe ratios of the portfolios in Table 5 and their respective matched portfolios. The table presents $t-$tests of the difference in variances and Sharpe ratios between the portfolios and their matched counterparts. For equally-weighted portfolios, only those formed on the basis of common analyst coverage and mutual fund holdings have statistically higher volatilities and lower Sharpe ratios than their respective matches on average. For value-weighted portfolios, the same pattern is only present among portfolios formed on the basis of common analyst coverage and share price.

4.8 Intensity-based tests

As we illustrate in Sections 4.4 - 4.6, residual return correlation is ubiquitous and consistent with a simple omitted factor explanation. Some studies have implicitly recognized the potential bias imposed by latent factors and instead explore whether the degree of return comovement is a function of similarity between assets based on observable criteria. For instance, (Anton and Polk (2014)) find that the pairwise correlation between risk adjusted stock returns is positively related to the intensity of common mutual fund holdings. Similarly, the strength of comovement has been linked to the distance between firm headquarter locations (Barker and Loughran (2007)) and the degree of common analyst coverage (Israelsen (2016)). The hope of these studies is that the grouping criterion is arguably uncorrelated with the loading on potentially omitted factors, thus circumventing the latent factor bias.

However, grouping assets based on similarity in observable dimensions is likely to result in similarity on unobservable dimensions. Thus, a positive relationship between asset similarity and comovement strength is also consistent with a latent factor explanation in which exposure to the latent factor(s) is related to similarity in the grouping characteristics. To explore the potential severity of this effect, we estimate how the intensity of comovement is related to the distance between the five characteristic variables that we consider in Section 4.6. In panel A of Table 7, we estimate the following specification:

$$\rho_{i,j,t} = \gamma \frac{-|x_{i,t} - x_{j,t}|}{\sigma(X_t)} + \varepsilon_{i,j,t} \tag{7}$$

where $\rho_{i,j,t}$ is the pairwise stock return correlations between stock i and stock j in year t and $\sigma(X_t)$ is the cross sectional standard deviation of characteristic x. Each entry of Panel A corresponds to an estimate of γ from a univariate regression. In Column 1, we use (raw) excess returns. Each subsequent column adjusts returns according to the Capital Asset Pricing Model (CAPM), the Fama-French three and five factor models (3 FM and 5 FM), and the Fama-French five factor model augmented with the momentum factor (6 FM). In all cases, γ is positive and statistically significant at conventional levels.

In Panel B, we present results for multivariate regressions in which the effect of similarities in all five characteristics on pairwise stock return correlations are estimated simultaneously. The coefficient estimates are smaller in virtually all cases, suggesting that similarity in one characteristic is related to similarity in others. In both Panels A and B, the comovement estimates are attenuated as we control for additional factors.

These findings demonstrate that an intensity-based test design is likely to suffer from a severe latent factor bias, reaffirming the intuition from our characteristic based sorts. First, controlling for additional factors always attenuates these comovement estimates. Even adjusting returns for the first 10 ex post principal components does not fully attenuate coefficient estimates. Second, since all of the characteristics that we consider have been linked to risk, these findings are consistent with excess comovement being a manifestation of common exposure to omitted factors.

4.9 Shock-based tests

Many studies in the comovement literature, including some of the settings we replicate, recognize the limitations of direct tests of correlation. These studies investigate shocks to the proposed structure that drives comovement. These shocks ostensibly alter the nature of comovement without affecting the fundamentals of the firm or asset under investigation. For example, Green and Hwang (2009) use stock splits as a shock to a stock's nominal share price. In these shock-based tests, an asset moves from one group to another group as a result of this shock. A simple difference-in-differences setting can therefore reveal whether the shock altered the nature of comovement.

Chen et al. (2016) argue that for such tests to favor the presence of excess comovement, an additional requirement needs to be met. Namely, excess comovement dictates that the movement from group 1 to group 2 be associated with simultaneously an increase in comovement with group 2 and a decrease in comovement with group 1. In the settings that we consider, these shocks do not exhibit this pattern. The assumption that the shock did not affect the fundamentals of the asset is therefore likely violated. In other words, if the asset comoves more with the destination group but continues to comove highly with the source group then it is likely that the move made the returns more volatile or different in some other fundamental manner.

In untabulated results, we replicate the main analysis in Green and Hwang (2009) and confirm that the comovement estimates from the source group are unchanged after a stock split. This pattern is consistent with omitted-factor(s). If a stock split is associated with a change in exposure to certain risk factors, then the comovement estimates will change but the stock will continue to comove with the source group since common exposure to the factor is not completely eliminated.

For shock-based tests, if the shock coincides with a change in the asset's exposure to a latent factor, then the comovement estimates will reflect this change in exposure. Table 9 presents simulation results that illustrate how shock-based tests can result in high comovement estimates with the destination group. We simulate the model in section 3 but change the exposure of a few assets to the latent factor Z_t as a proxy for a shock that alters factor exposure. Specifically, at the halfway mark of our simulation, we increase the value of γ_i by one standard deviation for the first asset in each random group of assets, and randomly rotate the other member of the group. Table 9 reports the median coefficient estimates from regressing the market-adjusted returns for the first asset in each group on the average returns of the remaining assets in the group both before and after the parameter change.

For comparability to prior results, we repeat the exercise described above for various parameter choices regarding group size (10, 20, 40, 80, and 160) and volatility of Z_t as a fraction of the volatility of F_t (1/4, 1/2, 1, 2). In every specification, the estimate of the comovement of asset 1 with the remaining assets in the group increases in the latter half of the sample. For example, with groups of 10 assets and $\sigma_Z = 1/2 \times \sigma_F$, the comovement estimate is 0.146 before the shock, but rises to 0.327 after the shock. Therefore, a change in factor exposure can lead to elevated estimates of comovement. Chen et al. (2016) find that in the settings they investigate, inclusion in the S&P

500 index and stock splits, past winners are over represented. That is to say, firms that changed groups based on these shocks experienced an increased exposure to the momentum factor. The higher comovement estimates for the prospective (and incumbent) group are therefore consistent with the latent factor model.

5 Conclusion

Tests of excess comovement are typically joint hypothesis test of an asset pricing theory and an empirical empirical analogue. For these tests to contradict that comovement in returns is driven by fundamentals, the empirical model used to adjust returns must capture all rational variation in expected returns. The limitation of this joint hypothesis problem is that investors have more information about the factors that drive returns, exposure to those factors, and anticipated changes in those factors than are directly observable to the researcher. We propose a test of excess comovement that accounts for the potential presence of latent factors.

We first show that random portfolios exhibit comovement estimates that are indistinguishable from that of several economically-motivated sources advanced in the literature. In addition, stocks grouped by similarity in characteristics that have been associated with risk exhibit substantial comovement estimates. Adjusting returns for additional empirical factors strongly attenuates comovement estimates, but does not eliminate the spurious comovement that we document. Furthermore, we propose a test that exploits the implications of excess comovement for portfolio diversification in the presence of latent factors. While our results do not rule out the potential for non fundamental explanations of comovement, they are strongly consistent with a latent factor explanation.

References

Anton, M., and C. Polk. 2014. Connected stocks. *Journal of Finance* 69:1099–1127.

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

Barker, D., and T. Loughran. 2007. The Geography of S&P 500 Stock Returns. *The Journal of Behavioral Finance* 8:177–190.

Chen, H., V. Singal, and R. F. Whitelaw. 2016. Comovement revisited. *Journal of Financial Economics* 121:624–644.

Chung, J.-W., and B. U. Kang. 2016. Prime Broker-Level Comovement in Hedge Fund Returns: Information or Contagion? *Review of Financial Studies* 29:3321–3353.

Cochrane, J. H. 1991. Volatility tests and efficient markets: A review essay. *Journal of Monetary Economics* 27:463–485.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *Journal of Finance* LII:1035–1058.

Daniel, K., and S. Titman. 1997. Evidence on the characteristics of cross sectional variation in stock returns. *the Journal of Finance* 52:1–33.

Fama, E. F. 1991. Efficient Capital Markets: II. *Journal of Finance* XLVI:1575–1617.

Fama, E. F., and K. R. French. 2015. Incremental variables and the investment opportunity set. *Journal of Financial Economics* 117:470–488.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society* pp. 1121–1152.

Green, C. T., and B.-H. Hwang. 2009. Price-based return comovement. *Journal of Financial Economics* 93:37–50.

Greenwood, R. 2008. Excess Comovement of Stock Returns: Evidence from Cross-Sectional Variation in Nikkei 225 Weights. *Review of Financial Studies* 21:1153–1186.

Grullon, G., S. Underwood, and J. P. Weston. 2014. Comovement and investment banking networks. *Journal of Financial Economics* 113:73–89.

Hameed, A., R. Morck, S. Jiangfeng, and B. Yeung. 2015. Information, Analysts, and Stock Return Comovement. *Review of Financial Studies* 28:3153–3187.

Israelsen, R. D. 2016. Does Common Analyst Coverage Explain Excess Comovement? *Journal of Financial and Quantitative Analysis* 51:1193–1229.

Kleidon, A. W. 1988. Bubbles, Fads and Stock Price Volatility Tests: A Partial Evaluation. *Journal of Finance* 43:656–659.

Kumar, A., and C. M. Lee. 2006. Retail Investor Sentiment and Return Comovements. *Journal of Finance* LXI:2451–2486.

Pirinsky, C., and Q. Wang. 2006. Does corporate headquarters location matter for stock returns? *Journal of Finance* 61:1991–2015.

Roll, R. 1977. A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics* 4:129–176.

Ross, S. A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13:341–60.

Shiller, R. 1983. Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends? *American Economic Review* LXXI:421–436.

Table 1: Summary Statistics

The table provides descriptive statistics of the main sample of monthly CRSP stock return (Panel A) and the average excess return (in %) for the assets in each replication sample we use (Panel B). Panel A presents the means, medians, standard deviations and 10^{th} and 90^{th} percentiles for excess returns (ex. ret.), book-to-market (B/M), Momentum (Mom), asset growth (AG), operating profitability (OP), and market equity (size) for all stocks in the sample from 1970 to 2016. Panel B presents the average and standard deviation (in brackets) of own asset returns along with the average excess return for the peer group and the market. The last columns report the average number of assets in each peer group and total number of observations in each sample. We use five distinct sample to capture Metropolitan Statistical Areas (MSA) of headquarter locations of firms, common analyst coverage of stocks, individual share price groups, common ownership of stocks by mutual funds, and shared prime brokers for hedge funds.

Panel A. Main sample characteristics					
	Mean	10^{th}%	Median	90^{th}%	Std. dev.
Ex. Ret	0.007	-0.165	-0.004	0.171	0.190
B/M	0.769	0.140	0.602	1.587	0.717
Mom	0.119	-0.488	0.037	0.739	0.572
AG	0.259	-0.152	0.080	0.660	0.771
OP	0.122	-0.289	0.191	0.468	0.619
Size ($ million)	1,103.160	7.314	94.145	2,119.013	3,606.520

Panel B. Own and peer group returns					
Group	r_i	r_{-i}	r_m	# Peers	# Obs
Headquarters	0.89	0.89	0.57	174	824,123
	[18.84]	[7.04]	[4.53]		
Analyst coverage	0.85	0.88	0.66	68	1,048,798
	[16.83]	[8.18]	[4.32]		
Stock price	0.78	0.78	0.56	589	3,405,870
	[17.50]	[7.15]	[4.73]		
Mutual Fund Ownership	0.52	0.52	0.45	1,185	476,640
	[12.46]	[5.29]	[4.53]		
Prime Broker	0.47	0.47	0.51	135	196,822
	[5.37]	[2.57]	[1.96]		

Table 2: Simulations: Latent Factors and Characteristics

This table reports simulation results of Eq.(1) in Section 3 for 240 months of returns for 2,400 assets, where we assume that F_t and Z_t follow an AR(1) process with $\sigma_F = 4.52\%$. Assets are assigned a random size at $t = 0$, which grows by $(1+r_{it})$ each period. We define groups g to contain 10, 20, 40, 80, or 160 assets. For each group size, we report the median estimate of γ in the regression: $r_{igt} = \alpha + \beta r_{mt} + \gamma r_{-igt} + \epsilon_{it}$, where r_{mt} is the value-weighted excess market return at t, and r_{-igt} is the excess return of group g at time t, excluding asset i. We form groups based on sorts of characteristic $X_i = \rho\Gamma + (1-\rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. Each column corresponds to a different value of ρ. Each panel corresponds to a different value of σ_Z, expressed as a multiple of σ_F. The model is simulated 1,000 times for each specification.

	$\rho = 0$	$\rho = .1$	$\rho = .25$	$\rho = .5$	$\rho = .75$	$\rho = .9$	$\rho = 1$
			$\sigma_Z = 0 \times \sigma_F$				
10	0.1256	0.1280	0.1263	0.1243	0.1293	0.1251	0.1254
20	0.2253	0.2214	0.2262	0.2233	0.2238	0.2274	0.2265
40	0.3604	0.3719	0.3611	0.3607	0.3650	0.3679	0.3727
80	0.5334	0.5410	0.5323	0.5316	0.5335	0.5379	0.5330
160	0.6963	0.7006	0.7002	0.7005	0.6938	0.7037	0.6957
			$\sigma_Z = 1/8 \times \sigma_F$				
10	0.1255	0.1272	0.1271	0.1286	0.1306	0.1348	0.1343
20	0.2322	0.2288	0.2307	0.2269	0.2244	0.2267	0.2242
40	0.3670	0.3699	0.3698	0.3697	0.3730	0.3737	0.3701
80	0.5368	0.5368	0.5385	0.5393	0.5451	0.5378	0.5390
160	0.7026	0.6896	0.6992	0.6953	0.6919	0.7026	0.7031
			$\sigma_Z = 1/4 \times \sigma_F$				
10	0.1293	0.1297	0.1291	0.1328	0.1354	0.1387	0.1405
20	0.2294	0.2333	0.2257	0.2397	0.2397	0.2365	0.2392
40	0.3709	0.3779	0.3718	0.3780	0.3837	0.3923	0.3777
80	0.5486	0.5418	0.5456	0.5487	0.5538	0.5569	0.5590
60	0.7025	0.7061	0.7057	0.7012	0.7159	0.7055	0.7069
			$\sigma_Z = 1/2 \times \sigma_F$				
10	0.1373	0.1366	0.1366	0.1487	0.1560	0.1615	0.1614
20	0.2396	0.2421	0.2434	0.2541	0.2794	0.2823	0.2757
40	0.3941	0.3990	0.3918	0.4174	0.4295	0.4362	0.4427
80	0.5576	0.5653	0.5644	0.5795	0.6032	0.6042	0.6031
160	0.7228	0.7198	0.7251	0.7377	0.7507	0.7543	0.7535
			$\sigma_Z = 1 \times \sigma_F$				
10	0.1757	0.1682	0.1808	0.2136	0.2418	0.2497	0.2543
20	0.2778	0.2818	0.2948	0.3469	0.3948	0.4061	0.4039
40	0.4437	0.4497	0.4565	0.5173	0.5620	0.5702	0.5686
80	0.6100	0.6135	0.6317	0.6812	0.7198	0.7281	0.7282
160	0.7579	0.7686	0.7750	0.8133	0.8360	0.8422	0.8420
			$\sigma_Z = 2 \times \sigma_F$				
10	0.2798	0.2738	0.3046	0.3983	0.4730	0.4950	0.4906
20	0.4261	0.4441	0.4687	0.5703	0.6499	0.6664	0.6652
40	0.6023	0.6225	0.6335	0.7260	0.7830	0.7941	0.7939
80	0.7576	0.7656	0.7731	0.8396	0.8813	0.8848	0.8869
160	0.8651	0.8674	0.8782	0.9129	0.9345	0.9387	0.9388
			$\sigma_Z = 4 \times \sigma_F$				
10	0.6730	0.6731	0.7020	0.7928	0.8562	0.8760	0.8700
20	0.8005	0.8056	0.8232	0.8774	0.9231	0.9293	0.9308
40	0.8918	0.8948	0.9017	0.9395	0.9610	0.9634	0.9654
80	0.9416	0.9422	0.9479	0.9677	0.9795	0.9813	0.9814
160	0.9692	0.9713	0.9729	0.9842	0.9900	0.9908	0.9906

Table 3: Sources of comovement and confidence intervals for bootstrapped null

The table presents estimates of the model $r_{igt} = r_{-igt} + r_{mt} + \epsilon_{igt}$ where r_{igt} represents asset i's excess returns, r_{-igt} represents the average excess return of all other assets in the same group g as i , and r_{mt} is the excess return on the market. The groups considered correspond to the following potential sources of comovement: headquarter location (Panel A), analyst coverage (Panel B), stock price (Panel C), mutual fund ownership (Panel D), and prime broker (Panel E). In each panel, the bootstrapped confidence intervals are calculated by randomly assigning assets to groups and estimating the model on the bootstrapped data. The assets considered are stocks for panels A-D, and hedge funds for panel E.

	Replicated Coefficient Estimates			Confidence Interval for Bootstrapped Null						
	HQ Location	Mkt		1%	5%	10%	50%	90%	95%	99%
Coef	0.636	0.424		0.657	0.658	0.659	0.660	0.662	0.663	0.664
t-stat	150.770	64.710								
	Analyst Coverage	MKt		1%	5%	10%	50%	90%	95%	99%
Coef	0.429	0.675		0.474	0.478	0.480	0.487	0.494	0.496	0.499
t-stat	175.401	145.742								
	Price	Mkt		1%	5%	10%	50%	90%	95%	99%
Coef	0.453	0.399		0.645	0.659	0.663	0.673	0.679	0.681	0.684
t-stat	632.115	503.157								
	Connections	Mkt		1%	5%	10%	50%	90%	95%	99%
Coef	1.027	-0.041		0.924	0.926	0.927	0.932	0.936	0.937	0.939
t-stat	98.866	-3.388								
	Prime Broker	Style	Mkt	1%	5%	10%	50%	90%	95%	99%
Coef	0.087	0.725	0.108	0.033	0.052	0.060	0.097	0.135	0.144	0.167
t-stat	14.050	123.600	11.510							

Table 4: Comovement, characteristic groups, and group size

This table presents comovement estimates of risk-adjusted stock returns on returns of other stocks grouped by different characteristics. Panel A reports the average comovement estimates from a simulation of randomly grouped stocks. Each subsequent panel reports comovement estimates for stocks grouped by Size, B/M, Momentum, Asset Growth, and Operating Profitability, respectively. The rows of each panel correspond to different group sizes, (10 to 160). Column 1 uses (raw) excess returns. Each subsequent column adjusts returns for, the Capital Asset Pricing Model (CAPM), the Fama-French three and five factor model (3 FM and 5 FM), and the Fama-French model augmented with the momentum factor (6 FM). Columns 6-8 adjust returns for principal factors from an ex post principal component analysis using the first five (PCA5), ten (PCA10), and twenty (PCA20) factors. The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

# Stocks	Raw	CAPM	3 FM	5 FM	6 FM	PCA5	PCA10	PCA20
Random Groups								
10	0.4795	0.1424	0.0535	0.0409	0.0407	0.2692	0.1685	0.0504
20	0.6469	0.1922	0.0722	0.0552	0.0549	0.3631	0.2273	0.0680
40	0.7836	0.2328	0.0875	0.0669	0.0665	0.4399	0.2754	0.0824
80	0.8762	0.2603	0.0978	0.0748	0.0743	0.4918	0.3079	0.0921
160	0.9313	0.2766	0.1040	0.0795	0.0790	0.5228	0.3273	0.0979
Market Equity								
10	0.5481	0.2469	0.1675	0.1660	0.1560	0.3602	0.2702	0.1647
20	0.7081	0.3188	0.2162	0.2142	0.2014	0.4652	0.3489	0.2126
40	0.8283	0.3710	0.2503	0.2480	0.2329	0.5428	0.4063	0.2461
80	0.9065	0.4038	0.2713	0.2688	0.2522	0.5928	0.4428	0.2667
160	0.9528	0.4195	0.2791	0.2764	0.2589	0.6202	0.4610	0.2742
Book to Market								
10	0.5109	0.1898	0.1051	0.1035	0.0929	0.3105	0.2146	0.1021
20	0.6741	0.2491	0.1371	0.1349	0.1209	0.4089	0.2820	0.1331
40	0.8052	0.2975	0.1637	0.1611	0.1444	0.4885	0.3368	0.1590
80	0.8902	0.3272	0.1788	0.1760	0.1573	0.5389	0.3707	0.1736
160	0.9414	0.3438	0.1864	0.1834	0.1637	0.5687	0.3902	0.1809
Momentum								
10	0.5571	0.2613	0.1832	0.1817	0.1719	0.3725	0.2841	0.1805
20	0.8325	0.3885	0.2714	0.2691	0.2544	0.5554	0.4228	0.2673
40	0.8325	0.3885	0.2714	0.2691	0.2544	0.5554	0.4228	0.2673
80	0.9104	0.4226	0.2940	0.2915	0.2754	0.6061	0.4604	0.2896
160	0.9547	0.4394	0.3037	0.3011	0.2841	0.6334	0.4794	0.2990
Asset Growth								
10	0.5110	0.1910	0.1065	0.1049	0.0943	0.3113	0.2157	0.1036
20	0.6763	0.2523	0.1405	0.1383	0.1243	0.4117	0.2851	0.1366
40	0.8047	0.2986	0.1651	0.1625	0.1458	0.4889	0.3377	0.1603
80	0.8899	0.3300	0.1822	0.1793	0.1608	0.5404	0.3731	0.1770
160	0.9394	0.3480	0.1919	0.1890	0.1694	0.5702	0.3937	0.1865
Operating Profitability								
10	0.5035	0.1797	0.0943	0.0926	0.0819	0.3014	0.2047	0.0913
20	0.6703	0.2403	0.1269	0.1247	0.1104	0.4020	0.2735	0.1229
40	0.7999	0.2854	0.1497	0.1471	0.1300	0.4788	0.3251	0.1449
80	0.8868	0.3165	0.1660	0.1631	0.1443	0.5309	0.3605	0.1608
160	0.9378	0.3347	0.1757	0.1726	0.1526	0.5614	0.3811	0.1699

Table 5: Squared Sharpe ratio test

This table presents estimates for the ratio of squared Sharpe ratios between stocks grouped by common headquarter location (Panel A), common analyst coverage (Panel B), similar share price (Panel C), common mutual fund ownership (Panel D), and hedge fund returns grouped by common prime broker (Panel E). In each setting, an asset is matched to the nearest match based on size, where size refers to market equity for stocks and assets under management for hedge funds. For each group of assets, an equally-weighted (value-weighted) portfolio is formed as well as corresponding portfolio of the matched assets. The table reports the mean and median of the ratio of Sharpe ratios of the sample portfolio and its match as well as the number of instance where the null of equal ratios is rejected by one-sided F-tail test. The first four samples use monthly returns for all CRSP stocks with available data, and the last sample uses all hedge funds in the TASS database.

	Equally-weighted portfolios			value-weighted portfolios			
	Mean	Median	# Reject.	Mean	Median	# Reject.	N
Panel A. Headquarter MSA							
Coef	1.03	0.89	23	1.06	0.91	23	110
t-stat	14.59			15.00			
Panel B. Analyst Coverage							
Coef	1.59	1.26	2368	1.64	1.27	2425	6576
t-stat	91.14			89.39			
Panel C. Share Price							
Coef	1.02	0.95	15	1.30	1.22	32	50
t-stat	26.25			26.21			
Panel D. Mutual Fund Holdings							
Coef	1.16	1.01	135	1.10	1.01	54	5175
t-stat	37.86			134.93			
Panel E. Prime Broker							
Coef	1.56	1.33	19	1.54	1.03	14	43
t-stat	9.66			6.78			

Table 6: Portfolio volatilities and Sharpe ratios

This table presents tests volatility and Sharpe ratio estimates for stocks grouped by common headquarter location (Panel A), common analyst coverage (Panel B), similar share price (Panel C), common mutual fund ownership (Panel D), and hedge fund returns grouped by common prime broker (Panel E). In each setting, an asset is matched to the nearest match based on size, where size refers to market equity for stocks and assets under management for hedge funds. For each group of assets, an equally-weighted (value-weighted) portfolio is formed as well as corresponding portfolio of the matched assets. The table reports the average volatility and Sharpe ratio of the sample portfolios and their matches. The table also reports the difference between these estimates and t-statistic of the significance of the difference. The first four samples use monthly returns for all CRSP stocks with available data, and the last sample uses all hedge funds in the TASS database.

	Equally-weighted portfolios						Value-weighted portfolios					
	Volatility			Sharpe ratio			Volatility			Sharpe ratio		
	Sample	Match	Diff.	Sample	Match	Diff.	Sample	Match	Diff.	Sample	Match	Diff.
Panel A. Headquarter MSA												
Coef	0.297	0.321	-0.024	0.407	0.337	0.070	0.300	0.320	-0.020	0.365	0.311	0.055
t-stat	25.113	23.204	-1.75	12.029	9.538	1.69	22.669	20.350	-1.38	12.291	9.282	1.29
Panel B. Analyst Coverage												
Coef	0.264	0.225	0.039	0.386	0.430	-0.044	0.225	0.192	0.032	0.325	0.376	-0.051
t-stat	189.526	245.401	30.89	48.985	59.620	-6.9	192.096	229.318	30.76	43.274	51.114	-7.36
Panel C. Share Price												
Coef	0.278	0.272	0.006	0.418	0.415	0.003	0.285	0.247	0.037	0.358	0.389	-0.031
t-stat	23.604	41.740	0.96	60.373	52.012	0.3	23.699	42.560	5.79	36.165	68.677	-2.64
Panel D. Mutual Fund Holdings												
Coef	0.175	0.169	0.007	0.243	0.271	-0.028	0.152	0.148	0.005	0.248	0.258	-0.010
t-stat	232.882	227.898	25.26	15.481	15.693	-3.79	241.621	246.809	18.54	12.850	14.075	-1.31
Panel E. Prime Broker												
Coef	0.115	0.098	0.017	0.587	0.692	-0.105	0.117	0.103	0.014	0.596	0.689	-0.093
t-stat	13.269	23.410	2.2	9.183	14.205	-1.74	10.164	23.820	1.4	8.512	12.465	-1.26

Table 7: Pairwise return correlations and characteristics

This table presents regression estimates for pairwise stock return correlations on a measure of similarity in size, book to market (B/M), momentum (mom), asset growth (AG), and operating profitability (OP). Similarities in characteristics are calculated as $-|x_{i,t} - x_{j,t}|/\sigma(x_t)$, where $x_{i,t}$ is the characteristic for stock i, $x_{j,t}$ is the characteristic for stock j, and $\sigma(x_t)$ is the cross sectional standard deviation of characteristic x at time t. In Panel A, we estimate univariate regressions of each characteristic and in Panel B, we estimate a multivariate regression for all characteristics jointly. In each panel, we consider pairwise correlations of residuals from a specific asset pricing model. These models correspond to the columns of the table: CAPM, Fama-French three- and five-factor models (3FM and 5FM), the three- and five-factor models augmented with momentum (4FM and 6FM), as well as three models based on principal factors (PCA5, PCA10, and PCA20). The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

	CAPM	3FM	5 FM	4FM	6FM	PCA5	PCA10	PCA20
Panel A. Univariate regressions								
Size	0.0034	0.0010	0.0010	0.0009	0.0009	0.0022	0.0014	0.0002
	(14.01)	(15.76)	(14.64)	(14.47)	(14.39)	(5.36)	(3.89)	(4.90)
B/M	0.0104	0.0052	0.0048	0.0050	0.0047	0.0166	0.0164	0.0066
	(5.33)	(4.66)	(4.25)	(4.36)	(3.99)	(5.16)	(4.95)	(4.30)
Mom	0.0059	0.0054	0.0041	0.0024	0.0021	0.0058	0.0053	0.0030
	(6.39)	(8.02)	(8.75)	(14.79)	(13.77)	(3.34)	(3.59)	(10.25)
AG	0.0225	0.0220	0.0186	0.0217	0.0192	0.0309	0.0170	0.0115
	(1.91)	(2.78)	(2.73)	(2.95)	(2.89)	(1.89)	(1.45)	(3.76)
OP	0.0145	0.0134	0.0112	0.0129	0.0098	0.0007	0.0076	0.0054
	(1.98)	(2.21)	(2.18)	(2.07)	(2.08)	(0.07)	(0.94)	(2.72)
Panel B. Multivariate regressions								
Size	0.0035	0.0011	0.0011	0.0010	0.0010	0.0015	0.0009	0.0002
	(16.36)	(11.41)	(11.63)	(11.14)	(10.52)	(3.43)	(2.20)	(5.15)
B/M	0.0078	0.0034	0.0036	0.0039	0.0035	0.0146	0.0121	0.0057
	(3.82)	(3.72)	(3.34)	(3.94)	(3.48)	(4.15)	(4.02)	(3.73)
Mom	0.0050	0.0052	0.0041	0.0021	0.0020	0.0042	0.0035	0.0029
	(4.78)	(6.57)	(6.81)	(9.17)	(9.15)	(2.75)	(2.86)	(12.35)
AG	0.0011	0.0155	0.0162	0.0196	0.0195	0.0280	0.0140	0.0053
	(0.12)	(1.77)	(1.93)	(2.30)	(2.33)	(1.27)	(0.80)	(2.36)
OP	0.0125	0.0112	0.0094	0.0121	0.0089	0.0057	0.0042	0.0035
	(1.77)	(1.98)	(1.88)	(1.96)	(1.93)	(0.55)	(0.60)	(1.89)

Table 8: Pairwise return correlations, characteristics and alternate sources

This table presents regression estimates for pairwise stock return correlations on a measure of similarity in size, book to market (B/M), momentum (mom), asset growth (AG), and operating profitability (OP), Fcap (see Anton and Polk (2014)), Analyst overlap (see Israelsen (2016)), geographic distance (see Barker and Loughran (2007)), and stock price level (see Green and Hwang (2009)). Similarities in characteristics are calculated as $-|x_{i,t} - x_{j,t}|/\sigma(x_t)$, where $x_{i,t}$ is the characteristic for stock i, $x_{j,t}$ is the characteristic for stock j, and $\sigma(x_t)$ is the cross sectional standard deviation of characteristic x at time t. In Panel A, we estimate univariate regressions of each characteristic and in Panel B, we estimate a multivariate regression for all characteristics jointly. In each panel, we consider pairwise correlations of residuals from a specific asset pricing model. These models correspond to the columns of the table: CAPM, Fama-French three- and five-factor models (3FM and 5FM), the three- and five-factor models augmented with momentum (4FM and 6FM), as well as three models based on principal factors (PCA5, PCA10, and PCA20). The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

	CAPM	3FM	5 FM	4FM	6FM	PCA5	PCA10	PCA20
Panel A. Univariate regressions								
Fcap	0.0128	0.0074	0.0059	0.0059	0.0049	0.0183	0.0139	0.0082
	(12.89)	(13.76)	(13.89)	(14.61)	(14.65)	(12.35)	(13.52)	(14.27)
Analyst overlap	0.1079	0.0939	0.0831	0.0858	0.0778	0.0674	0.0626	0.0541
	(4.62)	(4.51)	(4.52)	(4.50)	(4.47)	(4.91)	(4.98)	(5.15)
Distance	0.0020	0.0017	0.0015	0.0015	0.0013	0.0020	0.0018	0.0014
	(6.05)	(6.12)	(6.40)	(6.33)	(6.25)	(6.18)	(6.75)	(7.40)
Price	-0.0147	-0.0099	-0.0079	-0.0067	-0.0060	-0.0160	-0.0115	-0.0064
	(-5.99)	(-6.44)	(-7.36)	(-8.21)	(-8.14)	(-9.16)	(-11.48)	(-15.53)
Panel B. Multivariate regressions								
Size	0.0026	0.0014	0.0013	0.0013	0.0012	0.0013	0.0007	0.0000
	(11.10)	(8.74)	(9.74)	(9.92)	(9.80)	(3.85)	(2.79)	(0.28)
B/M	0.0080	0.0042	0.0053	0.0046	0.0048	0.0144	0.0095	0.0081
	(2.09)	(1.85)	(2.37)	(2.17)	(2.44)	(5.20)	(3.41)	(2.51)
Mom	0.0103	0.0077	0.0030	0.0058	0.0022	0.0064	0.0064	0.0043
	(5.91)	(7.15)	(5.91)	(6.55)	(5.74)	(3.80)	(6.45)	(8.06)
AG	0.0026	0.0014	0.0015	0.0010	0.0012	0.0000	0.0009	0.0003
	(1.13)	(1.25)	(1.43)	(1.22)	(1.43)	(0.09)	(1.80)	(1.37)
OP	0.0072	0.0036	0.0029	0.0011	0.0006	0.0098	0.0150	0.0057
	(0.80)	(0.95)	(0.85)	(0.30)	(0.18)	(1.07)	(2.46)	(1.51)
Fcap	0.0104	0.0063	0.0052	0.0049	0.0041	0.0180	0.0138	0.0078
	(10.20)	(9.48)	(9.06)	(8.75)	(8.03)	(7.99)	(8.59)	(11.01)
Analyst overlap	0.0773	0.0693	0.0620	0.0614	0.0556	0.0517	0.0468	0.0391
	(3.78)	(3.70)	(3.68)	(3.81)	(3.78)	(3.90)	(4.05)	(4.39)
Distance	0.0021	0.0022	0.0017	0.0017	0.0014	0.0020	0.0019	0.0014
	(3.32)	(3.70)	(4.04)	(4.14)	(3.93)	(3.51)	(3.75)	(4.55)
Price	-0.0067	-0.0057	-0.0032	-0.0047	-0.0032	-0.0121	-0.0081	-0.0043
	(-3.74)	(-4.15)	(-3.91)	(-4.55)	(-4.08)	(-10.63)	(-9.59)	(-10.81)

Table 9: Simulations: Shock-based Comovement

This table reports simulation results for a set of assets that experience a change in exposure to a latent factor. The simulation implements Eq.(1) of Section 3 for 480 months of returns for 2,400 assets, where we assume that F_t and Z_t follow an AR(1) process with $\sigma_F = 4.52\%$. Assets are assigned a random size at $t = 0$, which grows by $(1 + r_{it})$ each period. We define groups g to contain 10, 20, 40, 80, or 160 assets. The first asset of each group experiences a one standard deviation increase in exposure to the factor Z_t after $t = 240$. For each group size, we report the median estimate of γ in the regression: $r_{igt} = \alpha + \beta r_{mt} + \gamma r_{-igt} + \epsilon_{it}$, where r_{mt} is the value-weighted excess market return at t, and r_{-igt} is the excess return of group g at time t, excluding asset i. The regression is estimated over two subsamples ($0 < t \leq 240$ and $240 < t \leq 480$) using only observations corresponding to the first asset of each group, with each subsample corresponding to a group of columns in the table. Each column corresponds to a different value of σ_Z, expressed as a multiple of σ_F. The model is simulated 1,000 times for each specification.

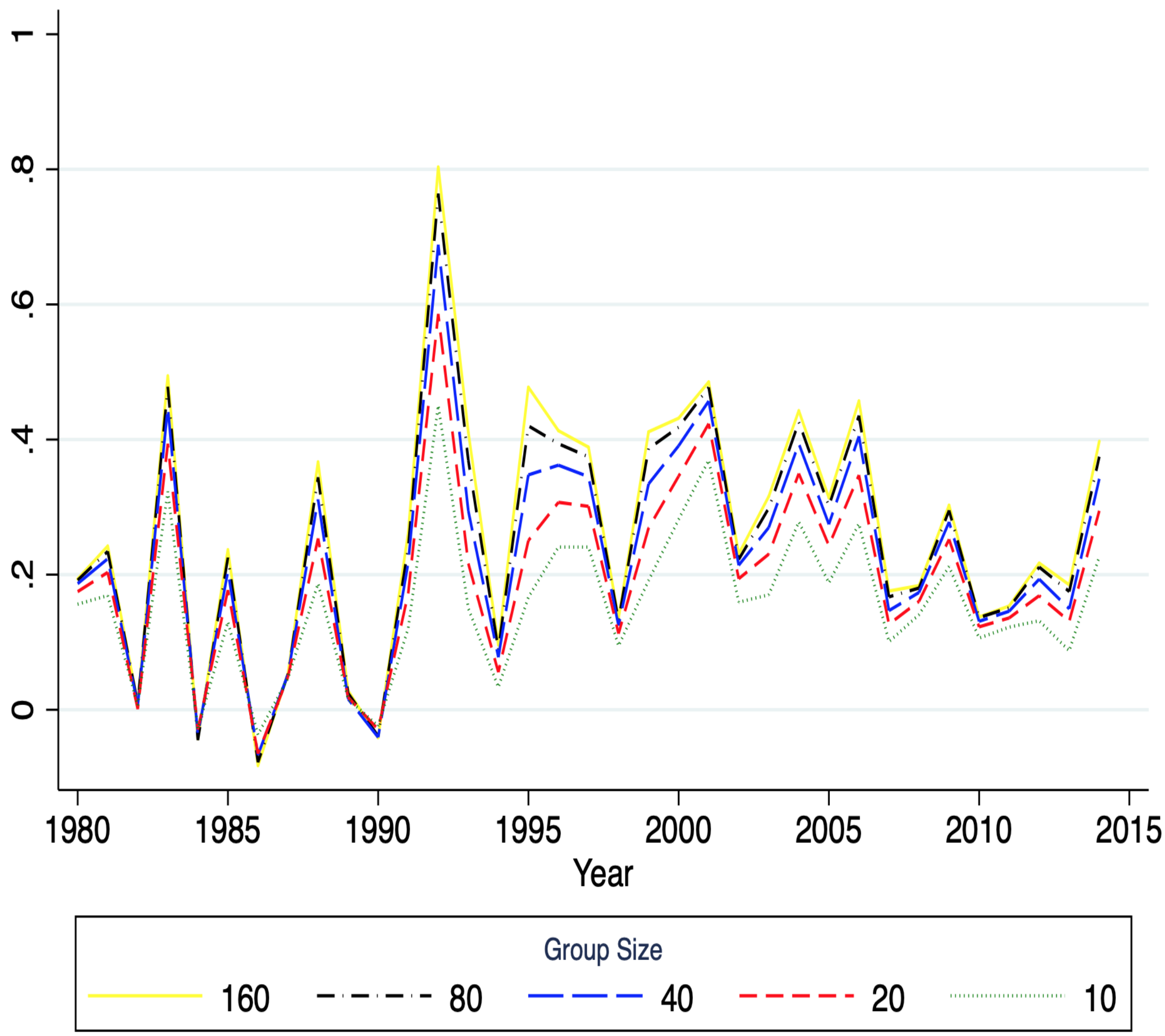
	$0 < t \leq 240$				$240 < t \leq 480$			
	$\sigma_Z =$				$\sigma_Z =$			
# Firms	$1/4 \times \sigma_F$	$1/2 \times \sigma_F$	$1 \times \sigma_F$	$2 \times \sigma_F$	$1/4 \times \sigma_F$	$1/2 \times \sigma_F$	$1 \times \sigma_F$	$2 \times \sigma_F$
10	0.146	0.175	0.176	0.230	0.327	0.333	0.456	0.732
20	0.240	0.252	0.279	0.378	0.458	0.500	0.660	0.977
40	0.398	0.393	0.450	0.524	0.635	0.688	0.864	1.142
80	0.554	0.566	0.613	0.709	0.794	0.858	1.008	1.279
160	0.721	0.664	0.746	0.835	0.872	0.923	1.122	1.339

Ubiquitous Comovement

Internet Appendix

William Grieser, Jung Hoon Lee, and Morad Zekhnini

Figure 1: Excess Comovement: Time Series (alternative)



This figure plots median comovement estimates for each year, obtained by regressing market adjusted returns on groups of randomly selected stocks in the CRSP universe from 1980-2016. For each panel, we repeat the analysis for groups containing 10, 20, 40, 80, or 160 randomly selected stocks. Each stock belongs to the same group throughout the sample period for each iteration. Median comovement estimates are obtained from 1000 iterations.

Table 1: Simulations: Confidence Intervals

The table shows estimate percentiles for simulations of the model in Section 3. The simulation consists of simulating 240 months of returns for 2,400 assets following Eq. (1) where we assume that F_t and Z_t follow an AR1 process. The volatility of F_t (σ_F) is 4.52% and each panel in the table represent different volatilities for Z_t as a multiple of σ_F. Assets are grouped randomly with 10, 20, 40, 80, or 160 assets in each group. For each asset, the peer returns are the average group return excluding the current asset from the group. For each grouping, we report the average coefficient γ in the regression: $r_{it} = \alpha + \beta r_{mt} + \gamma r_{-it} + \epsilon_{it}$ where r_{mt} is the value-weighted market portfolio excess return at t, and r_{-it} is i's peer group excess return at t. Value weighting is achieved by randomly assigning market capitalizations to assets at time $t = 0$, and adjusting sizes based on realized returns. For each specification, the model is simulated 1,000 times to extract the reported confidence intervals.

# Stocks	1%	5%	10%	50%	90%	95%	99%
				$\sigma_Z = 0 \times \sigma_F$			
10	0.0552	0.0701	0.0799	0.1270	0.2145	0.2454	0.3109
20	0.1179	0.1342	0.1472	0.2254	0.3660	0.4104	0.4858
40	0.1963	0.2299	0.2495	0.3674	0.5284	0.5815	0.6396
80	0.3310	0.3793	0.4112	0.5356	0.6944	0.7330	0.7884
160	0.4974	0.5500	0.5797	0.7006	0.8224	0.8430	0.8731
				$\sigma_Z = 1/8 \times \sigma_F$			
10	0.0598	0.0716	0.0796	0.1285	0.2241	0.2553	0.3221
20	0.1072	0.1356	0.1492	0.2262	0.3725	0.4087	0.4873
40	0.2127	0.2386	0.2606	0.3666	0.5332	0.5746	0.6424
80	0.3298	0.3836	0.4172	0.5438	0.7011	0.7310	0.7801
160	0.5092	0.5476	0.5820	0.6992	0.8254	0.8467	0.8731
				$\sigma_Z = 1/4 \times \sigma_F$			
10	0.0596	0.0724	0.0807	0.1259	0.2144	0.2504	0.3183
20	0.1074	0.1344	0.1512	0.2332	0.3628	0.4016	0.4759
40	0.2053	0.2430	0.2614	0.3685	0.5305	0.5819	0.6412
80	0.3344	0.3858	0.4139	0.5473	0.7078	0.7438	0.8042
160	0.4933	0.5464	0.5832	0.7035	0.8257	0.8494	0.8765
				$\sigma_Z = 1/2 \times \sigma_F$			
10	0.0636	0.0782	0.0855	0.1393	0.2534	0.2858	0.3596
20	0.1107	0.1405	0.1572	0.2378	0.3768	0.4200	0.5075
40	0.2227	0.2473	0.2722	0.3921	0.5610	0.6113	0.6809
80	0.3538	0.4057	0.4392	0.5652	0.7226	0.7617	0.8185
160	0.5129	0.5621	0.5971	0.7113	0.8291	0.8500	0.8793
				$\sigma_Z = \sigma_F$			
10	0.0752	0.0919	0.1035	0.1716	0.2949	0.3442	0.4307
20	0.1333	0.1714	0.1863	0.2838	0.4466	0.5036	0.5831
40	0.2439	0.2814	0.3143	0.4478	0.6249	0.6805	0.7659
80	0.3863	0.4534	0.4891	0.6236	0.7795	0.8081	0.8559
160	0.5624	0.6144	0.6445	0.7518	0.8696	0.8944	0.9275
				$\sigma_Z = 2 \times \sigma_F$			
10	0.0964	0.1233	0.1479	0.2815	0.4920	0.5612	0.6833
20	0.1941	0.2383	0.2684	0.4379	0.6619	0.7139	0.7871
40	0.2921	0.3758	0.4121	0.6139	0.8029	0.8452	0.8921
80	0.4645	0.5378	0.5871	0.7560	0.8852	0.9076	0.9402
160	0.6386	0.7046	0.7401	0.8658	0.9399	0.9525	0.9687
				$\sigma_Z = 4 \times \sigma_F$			
10	0.2213	0.3120	0.3616	0.6632	0.8885	0.9288	0.9673
20	0.3450	0.4794	0.5540	0.8046	0.9428	0.9600	0.9832
40	0.5013	0.6421	0.7104	0.8869	0.9701	0.9803	0.9906
80	0.6896	0.7871	0.8360	0.9455	0.9849	0.9902	0.9962
160	0.8236	0.8760	0.9024	0.9664	0.9918	0.9947	0.9981

Table 2: Simulations: Imperfect Market Proxy

This table reports simulation results of Eq.(1) in Section 3 for 240 months of returns for 2,400 assets, where we set $Z_t \equiv 0$, and F_t follows an AR(1) process with $E[F_t] = 0.65\%$. Assets are assigned a random size at $t = 0$, which grows by $(1 + r_{it})$ each period. We define groups g to contain 10, 20, 40, 80, or 160 assets. For each group size, we report the median estimate of γ in the regression: $r_{igt} = \alpha + \beta r_{mt} + \gamma r_{-igt} + \epsilon_{it}$, where r_{mt} is the value-weighted excess market return at t, and r_{-igt} is the excess return of group g at time t, excluding asset i. Each column corresponds to a different value of σ_F, expressed as a multiple of the volatility of the average monthly value-weighted market return from 1980-2016 ($\sigma_F = 4.52\%$). The model is simulated 1,000 times for each specification.

	$\sigma_F = 1/8 \times 4.52\%$	$\sigma_F = 1/4 \times 4.52\%$	$\sigma_F = 1/2 \times 4.52\%$	$\sigma_F = 1 \times 4.52\%$	$\sigma_F = 2 \times 4.52\%$	$\sigma_F = 4 \times 4.52\%$
10	0.0629	0.1697	0.4075	0.6834	0.8892	1.2239
20	0.0637	0.1727	0.4079	0.6829	0.9048	1.2389
40	0.0603	0.1670	0.4170	0.6847	0.9078	1.2156
80	0.0614	0.1717	0.4133	0.6957	0.8971	1.2407
160	0.0615	0.1729	0.4057	0.6934	0.9060	1.2290

Table 3: Simulations: Unpriced Latent Factor and Characteristics

This table reports simulation results of Eq.(1) in Section 3 for 240 months of returns for 2,400 assets, where we set $F_t \equiv 0$, and Z_t follows an AR(1) process with $E[Z_t] = 0$. Assets are assigned a random size at $t = 0$, which grows by $(1 + r_{it})$ each period. We define groups g to contain 10, 20, 40, 80, or 160 assets. For each group size, we report the median estimate of γ in the regression: $r_{igt} = \alpha + \gamma r_{-igt} + \epsilon_{it}$, where r_{-igt} is the excess return of group g at time t, excluding asset i. We form groups based on sorts of characteristic $X_i = \rho\Gamma + (1 - \rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. Each column corresponds to a different value of ρ. Each panel corresponds to a different value of σ_Z, expressed as a multiple of the volatility of the average monthly value-weighted market return from 1980-2016 ($\sigma_F = 4.52\%$). The model is simulated 1,000 times for each specification.

	$\rho = 0$	$\rho = .1$	$\rho = .25$	$\rho = .5$	$\rho = .75$	$\rho = .9$	$\rho = 1$
			$\sigma_Z = 1/8 \times \sigma_F$				
10	0.0378	0.0382	0.0389	0.0425	0.0448	0.0450	0.0453
20	0.0733	0.0734	0.0747	0.0799	0.0851	0.0873	0.0867
40	0.1355	0.1361	0.1379	0.1469	0.1568	0.1593	0.1600
80	0.2394	0.2393	0.2428	0.2575	0.2736	0.2719	0.2754
160	0.3872	0.3871	0.3914	0.4101	0.4270	0.4307	0.4309
			$\sigma_Z = 1/4 \times \sigma_F$				
10	0.1372	0.1375	0.1401	0.1493	0.1586	0.1617	0.1616
20	0.2411	0.2413	0.2444	0.2602	0.2741	0.2774	0.2763
40	0.3888	0.3890	0.3926	0.4127	0.4308	0.4333	0.4337
80	0.5611	0.5606	0.5664	0.5825	0.5996	0.6054	0.6055
160	0.7171	0.7184	0.7220	0.7367	0.7506	0.7528	0.7533
			$\sigma_Z = 1/2 \times \sigma_F$				
10	0.1501	0.1500	0.1634	0.1984	0.2318	0.2407	0.2387
20	0.2625	0.2688	0.2757	0.3302	0.3750	0.3906	0.3869
40	0.4233	0.4205	0.4355	0.5032	0.5493	0.5586	0.5535
80	0.5997	0.5948	0.6051	0.6610	0.7053	0.7152	0.7113
160	0.7413	0.7459	0.7485	0.7961	0.8274	0.8318	0.8329
			$\sigma_Z = 1 \times \sigma_F$				
10	0.3994	0.3981	0.4023	0.4264	0.4427	0.4503	0.4491
20	0.5698	0.5717	0.5760	0.5953	0.6164	0.6185	0.6203
40	0.7249	0.7255	0.7306	0.7478	0.7618	0.7648	0.7649
80	0.8414	0.8407	0.8434	0.8546	0.8647	0.8668	0.8666
160	0.9132	0.9140	0.9150	0.9215	0.9276	0.9285	0.9286
			$\sigma_Z = 2 \times \sigma_F$				
10	0.7583	0.7581	0.7646	0.7884	0.8083	0.8127	0.8136
20	0.8631	0.8623	0.8665	0.8806	0.8928	0.8954	0.8973
40	0.9265	0.9272	0.9287	0.9367	0.9438	0.9447	0.9449
80	0.9619	0.9615	0.9627	0.9673	0.9707	0.9719	0.9720
160	0.9803	0.9805	0.9810	0.9832	0.9853	0.9854	0.9856
			$\sigma_Z = 4 \times \sigma_F$				
10	0.9811	0.9814	0.9842	1.0008	1.0148	1.0169	1.0176
20	0.9901	0.9907	0.9921	1.0001	1.0072	1.0086	1.0086
40	0.9949	0.9982	0.9961	1.0002	1.0036	1.0043	1.0043
80	0.9975	0.9976	0.9980	1.0000	1.0018	1.0021	1.0021
160	0.9987	0.9989	0.9990	1.0000	1.0008	1.0010	1.0011

Table 4: Simulations: Priced Latent Factor and Characteristics

This table reports simulation results of Eq.(1) in Section 3 for 240 months of returns for 2,400 assets, where we set $F_t = 0$, and Z_t follows an AR(1) process with $E[Z_t] = 0.65\%$. Assets are assigned a random size at $t = 0$, which grows by $(1+r_{it})$ each period. We define groups g to contain 10, 20, 40, 80, or 160 assets. For each group size, we report the median estimate of γ in the regression: $r_{igt} = \alpha + \beta r_{mt} + \gamma r_{-igt} + \epsilon_{it}$, where r_{mt} is the value-weighted excess market return at t, and r_{-igt} is the excess return of group g at time t, excluding asset i. We form groups based on sorts of characteristic $X_i = \rho\Gamma + (1-\rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. Each column corresponds to a different value of ρ. Each panel corresponds to a different value of σ_Z, expressed as a multiple of σ_F. The model is simulated 1,000 times for each specification.

	$\rho = 0$	$\rho = .1$	$\rho = .25$	$\rho = .5$	$\rho = .75$	$\rho = .9$	$\rho = 1$
			$\sigma_Z = 1/8 \times \sigma_F$				
10	0.0382	0.0382	0.0390	0.0431	0.0472	0.0482	0.0482
20	0.0733	0.0733	0.0753	0.0821	0.0897	0.0913	0.0922
40	0.1365	0.1378	0.1396	0.1517	0.1639	0.1677	0.1689
80	0.2396	0.2407	0.2454	0.2648	0.2837	0.2885	0.2867
160	0.3852	0.3850	0.3922	0.4182	0.4398	0.4455	0.4469
			$\sigma_Z = 1/4 \times \sigma_F$				
10	0.1278	0.1308	0.1302	0.1348	0.1329	0.1330	0.1307
20	0.2287	0.2321	0.2300	0.2457	0.2414	0.2385	0.2313
40	0.3775	0.3721	0.3875	0.3752	0.3806	0.3764	0.3841
80	0.5484	0.5381	0.5402	0.5528	0.5527	0.5512	0.5530
160	0.7085	0.7133	0.7108	0.7082	0.7157	0.7168	0.7100
			$\sigma_Z = 1/2 \times \sigma_F$				
10	0.1368	0.1370	0.1388	0.1509	0.1613	0.1624	0.1637
20	0.2420	0.2418	0.2460	0.2623	0.2765	0.2798	0.2815
40	0.3899	0.3883	0.3925	0.4153	0.4345	0.4372	0.4404
80	0.5606	0.5597	0.5659	0.5848	0.6054	0.6090	0.6117
160	0.7167	0.7198	0.7214	0.7385	0.7541	0.7566	0.7577
			$\sigma_Z = 1 \times \sigma_F$				
10	0.3971	0.3975	0.4033	0.4266	0.4446	0.4493	0.4503
20	0.5699	0.5701	0.5749	0.5950	0.6186	0.6203	0.6218
40	0.7252	0.7250	0.7295	0.7473	0.7642	0.7653	0.7667
80	0.8404	0.8408	0.8441	0.8547	0.8647	0.8679	0.8675
160	0.9137	0.9134	0.9153	0.9220	0.9273	0.9289	0.9289
			$\sigma_Z = 2 \times \sigma_F$				
10	0.7576	0.7581	0.7628	0.7886	0.8067	0.8119	0.8131
20	0.8619	0.8638	0.8658	0.8812	0.8933	0.8962	0.8959
40	0.9266	0.9274	0.9286	0.9366	0.9437	0.9450	0.9453
80	0.9617	0.9621	0.9627	0.9674	0.9707	0.9719	0.9720
160	0.9805	0.9805	0.9811	0.9833	0.9853	0.9857	0.9857
			$\sigma_Z = 4 \times \sigma_F$				
10	0.9808	0.9810	0.9848	1.0005	1.0146	1.0169	1.0175
20	0.9901	0.9900	0.9924	1.0000	1.0070	1.0084	1.0085
40	0.9950	0.9983	0.9959	1.0000	1.0035	1.0042	1.0043
80	0.9975	0.9976	0.9981	0.9999	1.0017	1.0021	1.0021
160	0.9987	0.9988	0.9990	1.0000	1.0008	1.0010	1.0010

Table 5: Industry Adjusted Returns and Comovement

This table presents comovement estimates of risk-adjusted stock returns on the returns of portfolios containing randomly selected stocks. In Panel A, we repeat our analysis from Panel A of Table 4 for convenience. Each subsequent panel reports comovement estimates for stocks adjusted for an industry factor model according to the Fama and French 12, 30, and 48 industry portfolios, and the 25 fixed, text-based network industry classifications (TNIC) from Hoberg and Phillips (2015) , respectively. The rows of each panel correspond to different group sizes, (10 to 160). Column 1 uses excess, industry-adjusted returns. Each subsequent column additionally adjusts returns for, the Capital Asset Pricing Model (CAPM), the Fama-French three and five factor model (3 FM and 5 FM), and the Fama-French models augmented with the momentum factor (4 FM and 6 FM). Columns 7-9 adjust returns for principal factors from an ex post principal component analysis using the first five (PCA5), ten (PCA10), and twenty (PCA20) factors. The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

# Firms	Raw	CAPM	3 FM	5 FM	6 FM	PCA5	PCA10	PCA20
				No Industry Adjustment				
10	0.4795	0.1424	0.0535	0.0409	0.0407	0.2692	0.1685	0.0504
20	0.6469	0.1922	0.0722	0.0552	0.0549	0.3631	0.2273	0.0680
40	0.7836	0.2328	0.0875	0.0669	0.0665	0.4399	0.2754	0.0824
80	0.8762	0.2603	0.0978	0.0748	0.0743	0.4918	0.3079	0.0921
160	0.9313	0.2766	0.1040	0.0795	0.0790	0.5228	0.3273	0.0979
				Fama-French 12 Industries				
10	0.1079	0.1006	0.0438	0.0432	0.0353	0.0708	0.0572	0.0428
20	0.1460	0.1361	0.0595	0.0587	0.0481	0.0959	0.0775	0.0582
40	0.1771	0.1651	0.0724	0.0714	0.0585	0.1165	0.0942	0.0708
80	0.1980	0.1846	0.0808	0.0797	0.0653	0.1301	0.1052	0.0790
160	0.2102	0.1959	0.0857	0.0845	0.0692	0.1381	0.1116	0.0838
				Fama-French 30 Industries				
10	0.0761	0.0703	0.0368	0.0361	0.0305	0.0611	0.0529	0.0373
20	0.1031	0.0953	0.0501	0.0491	0.0416	0.0829	0.0718	0.0507
40	0.1251	0.1157	0.0610	0.0598	0.0506	0.1006	0.0872	0.0617
80	0.1398	0.1292	0.0681	0.0667	0.0565	0.1124	0.0975	0.0689
160	0.1484	0.1372	0.0722	0.0707	0.0599	0.1193	0.1034	0.0730
				Fama-French 48 Industries				
10	0.0535	0.0516	0.0328	0.0316	0.0270	0.0463	0.0440	0.0280
20	0.0726	0.0700	0.0446	0.0431	0.0368	0.0629	0.0598	0.0382
40	0.0882	0.0851	0.0543	0.0525	0.0449	0.0764	0.0727	0.0465
80	0.0986	0.0950	0.0606	0.0586	0.0501	0.0854	0.0812	0.0519
160	0.1045	0.1008	0.0642	0.0620	0.0531	0.0905	0.0861	0.0550
				TNIC 25				
10	0.0602	0.0545	0.0322	0.0306	0.0272	0.0480	0.0433	0.0261
20	0.0799	0.0723	0.0430	0.0408	0.0364	0.0638	0.0576	0.0349
40	0.0955	0.0864	0.0514	0.0489	0.0436	0.0763	0.0689	0.0419
80	0.1055	0.0955	0.0568	0.0540	0.0481	0.0843	0.0762	0.0462
160	0.1112	0.1007	0.0598	0.0569	0.0507	0.0888	0.0802	0.0486

Table 6: DGTW Adjusted Returns and Comovement

This table presents comovement estimates of stock returns on returns of other stocks grouped by different characteristics. All stock returns are first characteristic adjusted according to the process outlined in Daniel et al. (1997). Column 1 reports the average DGTW adjusted return comovement estimates from a simulation of randomly grouped stocks. Each subsequent Column reports DGTW adjusted return comovement estimates for stocks grouped by Size, Book to Market, Momentum, Asset Growth, and Operating Profitability, respectively. The rows of each panel correspond to different group sizes, (10 to 160). The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

	Grouping Criteria					
	Random	Market Equity	Book to Market	Momentum	Asset Growth	Operating Profitability
10	0.0133	0.0562	0.0182	0.0793	0.0317	0.0340
20	0.0169	0.0733	0.0229	0.1003	0.0406	0.0451
40	0.0236	0.0841	0.0274	0.1160	0.0483	0.0504
80	0.0310	0.0905	0.0288	0.1231	0.0522	0.0563
160	0.0307	0.0926	0.0288	0.1271	0.0540	0.0591

Table 7: Shock-based Test

This table presents

	N	Mean	Std Dev	Min	Max
Panel A. Headquater Relocation					
Excess Return on the Market (Pre)	6124	0.9145	0.6746	2.1473	6.7285
Excess Return on the Market(Post)	6124	0.8947	0.6488	2.6321	7.0917
Small-Minus-Big Return(Pre)	6124	0.6968	1.1186	16.5935	16.9103
Small-Minus-Big Return(Post)	6124	0.5960	0.9285	3.7981	7.091
High-Minus-Low Return(Pre)	6124	0.2268	1.0902	9.2047	7.1493
High-Minus-Low Return (Post)	6124	0.2402	0.9894	9.2043	6.1811
Momentum Factor (Pre)	6124	-0.1082	0.7028	7.8225	5.3165
Momentum Factor (Post)	6124	-0.1484	0.648	12.0962	3.1874
Panel B. Stock Split					
Excess Return on the Market(Pre)	6242	0.9657	0.6004	1.7155	4.3987
Excess Return on the Market(Post)	6242	1.1011	0.6233	2.6758	4.0881
Small-Minus-Big Return(Pre)	6242	0.6684	0.7294	4.1722	5.2727
Small-Minus-Big Return(Post)	6242	0.6529	0.8076	4.3012	5.1877
High-Minus-Low Return(Pre)	6242	0.0003	0.8385	4.9036	4.3973
High-Minus-Low Return (Post)	6242	-0.1108	1.0345	6.8973	8.1035
Momentum Factor (Pre)	6242	0.1766	0.6162	6.686	5.9106
Momentum Factor (Post)	6242	0.0844	0.7044	10.1072	4.0175

Table 8: Index Additions and Deletions

This table presents

	Difference	t-stat	p-value	Difference	t-stat	p-value	Difference	t-stat	p-value
Panel C. Changes in Beta before and after the addition to Index									
'Qtrs: -1 to 1	0.005	0.64	0.520	-0.289	-1.56	0.120	0.247	2.83	0.005
'Qtrs: -2 to 2	0.005	0.87	0.384	-0.107	-0.82	0.410	0.324	4.87	¡.0001
'Qtrs: -3 to 3	0.020	3.92	¡.0001	-0.153	-1.25	0.210	0.233	4.11	¡.0001
'Qtrs: -4 to 4	0.016	3.26	0.001	-0.125	-1.05	0.292	0.224	4.27	¡.0001
'Qtrs: -5 to 5	0.018	3.73	0.000	-0.131	-1.37	0.170	0.134	3.21	0.001
'Qtrs: -6 to 6	0.015	3.32	0.001	-0.070	-0.90	0.367	0.142	3.88	0.000
'Qtrs: -7 to 7	0.013	2.75	0.006	-0.103	-1.75	0.080	0.114	3.53	0.000
'Qtrs: -8 to 8	0.016	3.50	0.001	-0.061	-1.19	0.233	0.066	2.30	0.021
'Qtrs: -9 to 9	0.016	3.42	0.001	-0.065	-1.56	0.120	0.019	1.00	0.317
'Qtrs: -10 to 10	0.014	2.93	0.003	-0.005	-0.15	0.881	-0.008	-0.57	0.567
'Qtrs: -11 to 11	0.008	1.65	0.100	-0.008	-0.48	0.630	-0.032	-3.67	0.000
'Qtrs: -12 to 12	0.011	2.18	0.029	-0.032	-3.38	0.001	-0.042	-6.63	¡.0001