Ubiquitous Comovement*

William Grieser† Jung Hoon Lee‡ Morad Zekhnini‡

December 15, 2019

Abstract

Rational and behavioral asset pricing theories offer conflicting interpretations of the covariance structure of asset returns. Return comovement beyond what prespecified 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 indistinguishable from the comovement of economically motivated groupings advanced in the literature. Our finding is consistent with the presence of a latent factor that could be derived from multiple sources of systematic variation, including rational sources. We propose new statistical tests that account for latent factors when detecting excess comovement.

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

JEL Codes: G11, G12, G14, G40

*We thank Kerry Back, John Campbell, Scott Cederburg, Zhi Da, Paul Irvine, Ryan Israelsen, Zoran Ivkovich, Kris Jacobs, Travis Johnson, Ron Kaniel, Nishad Kapadia, Mark Schroder, Jay Shanken, Richard Sias, Andrei Simonov, Malcolm Wardlaw, James Weston, and Toni Whited for thoughtful comments. We also thank additional seminar participants at Emory University, Michigan State University, Rice University, Southern Methodist University, Texas Christian University, Tulane University, the University of Arizona, the University of Houston, the University of New Orleans, and the University of Texas at Dallas. All errors are our own. †Department of Finance, Neeley School of Business Texas Christian University, phone: (817) 253-3535, email: w.grieser@tcu.edu. ‡Department of Finance, A.B. Freeman School of Business Tulane University, email: 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 market imperfections. To illustrate this point, we show that randomly grouped assets generally exhibit substantial within-group residual return comovement, which is indistinguishable from that of economically motivated groups advanced in the literature. We further illustrate that these findings are consistent with a latent factor explanation. Thus, extreme caution should be exercised when interpreting the magnitude of results as evidence of a given explanation.

Traditional asset pricing theory contends that, in a rational framework, return comovement should be driven by commonality in asset fundamentals. Alternatively, market frictions and behavioral biases could lead to deviations from fundamental value. To the extent that these deviations are correlated across assets, they can also cause return comovement (see Barberis et al. (2005)). 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. However, the implicit assumption behind this interpretation is that unmodeled systematic variation has 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 factor on return comovement.[1] Simulations confirm our model's prediction that market-adjusted 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 and in the number of assets included in each group used to estimate comovement. Even when the latent factor accounts for only a small fraction of total asset return volatility, we obtain comovement estimates for random groups that are comparable

[1] Sias et al. (2017) use a similar framework to show that small model misspecifications have a significant impact on estimates of hedge fund contagion.

to those documented in several studies. The novel implication of our model is that even relatively inconsequential latent 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. Indeed, Pollet and Wilson (2010) show that individual stock returns share a common sensitivity to aggregate wealth when the stock market serves as a poor proxy and this sensitivity manifests through pairwise stock return correlations.

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 (MSAs) in our replication of Pirinsky and Wang (2006). We then perform a placebo procedure in which we randomly assign firms to MSAs and estimate the stock return comovement within randomly assigned headquarters locations. The median estimate produced from 1,000 iterations of this procedure (0.660) is even larger than the comovement estimate for actual headquarters locations. Using a similar approach, we compare comovement estimates for randomly grouped assets to groups formed 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 placebo 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 unobserved risk. 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 regardless of the sample period under investigation. For instance, we obtain a median placebo comovement estimate of 0.079 even after controlling for the Fama and French five-factor model augmented with momentum (henceforth, the six-factor model).[2] These findings suggest that a null hypothesis of zero residual return comovement leads to severe overstatements of "excess" comovement.

Next, we show that grouping assets by characteristics, rather than randomly, significantly intensifies comovement estimates. For instance, we find a six-factor residual comovement estimate of 0.259 for stocks grouped by similarity in market equity (i.e., size). We obtain qualitatively similar estimates when we group stocks by similarity in book-to-market, momentum, asset growth, and operating profitability. Commonalities in unobservable criteria will likely generate the same effect. Thus, to conclude that comovement within a particular group of assets is in "excess" of a rational model, and due to a proposed source, requires controlling for commonalities in all other characteristics.

A few studies have acknowledged the potential for latent factors to influence comovement estimates (e.g., Israelsen (2016); Sias et al. (2017)). 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 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 to the distance between firm headquarters locations (Barker and Loughran (2007)).

We use an intensity-based design to show that commonality in characteristics or factor exposure is positively associated with pairwise return correlations. Furthermore, using factor model residuals always attenuates the relationship between asset commonality and pairwise correlation. However, residuals from models that correspond to the characteristics or factors under consideration do not sufficiently account for this relationship. For instance, a significant correlation persists between assets with a similar size after controlling for factor models that include the small-minus-big (SMB) factor. Additionally, similarities in factor exposures and characteristics are not orthogonal. Thus,

[2]Industry adjustments and Characteristics adjustments in the style of Daniel and Titman (1997) do not fully attenuate estimates.

assets that are similar along observable criteria will likely share similar exposures to omitted factors, which makes it difficult to attribute a residual return correlation to any specific source. These findings 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.[3] These tests typically show that within-group comovement becomes stronger after the shock. However, shocks that are either caused by or lead to evolving fundamentals could produce changes in comovement that are challenging to separate from the proposed channel. Through simulations, we show that even mild changes in factor exposure result in a substantive increase in comovement estimates. Furthermore, we provide evidence that factor loadings change significantly surrounding several events that have been explored in the literature, thus violating the exogeneity assumption of these shock-based tests. These findings provide a generalization of those by Chen et al. (2016), who attribute changes in comovement surrounding stock splits and inclusion in the S&P 500 to changes in exposure to momentum.

While our findings support the presence of latent factors, they remain silent on the source of those factors, which can arise for either rational or behavioral reasons. However, the alternative framework that we propose offers testable implications for portfolio volatilities and Sharpe ratios that have not been explored in the literature. In particular, portfolios that exhibit excess comovement are underdiversified and should also exhibit higher volatilities. Moreover, *excess comovement* is commonly defined as covariation between asset returns that is not driven by fundamentals and is not compensated through return premiums. Thus, excess comovement will lead to portfolios with lower Sharpe ratios. Alternatively, if comovement within a portfolio is due to exposure to a priced risk factor, the high volatility will be compensated through higher return premiums.

We implement the variance (Sharpe ratio) test for the five sources of excess comovement we consider. For instance, we build portfolios of stocks from firms headquartered in each MSA, then match each portfolio to a portfolio of firms located outside of the focal MSA.[4] We then compute the ratio of the volatilities between each MSA portfolio and its matched portfolio, after controlling for

[3] For instance, studies have examined comovement surrounding plausibly exogenous shocks from brokerage house mergers (Israelsen (2016); Chung and Kang (2016)), and S&P 500 additions/deletions (Barberis et al. (2005)).

[4] We use nearest-neighbor match based on market capitalization. Details are discussed in Section 4.8.

market exposure. Under the null hypothesis, this ratio follows an F distribution. The alternative hypothesis of a high variance for portfolios that exhibit 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 the tested portfolios. Only the ratio for analyst coverage leads to a rejection at conventional levels of significance. Thus, in these settings, we cannot distinguish the proposed source of comovement from other sources of systematic variation, including rational sources.

To our knowledge, we are the first to demonstrate the ubiquitous nature of residual return comovement and the severity of the latent factor bias. Our findings generalize those of Chen et al. (2016) by revisiting the topic of excess comovement using a linear factor structure and by considering settings outside the shock-based test design. 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, we show that any systematic variation that remains unaccounted for in an empirical model results in substantive residual return comovement. Thus, attributing much of the findings in the literature to specific sources of comovement is premature.

We propose two procedures to mitigate confounding influences of a latent factor. First, rather than testing against a null hypothesis of zero comovement, studies should derive a null from the comovement exhibited within randomly grouped assets. Random groups that are unrelated to the source of comovement being studied will account for common exposure to omitted factors and therefore provide a more appropriate benchmark. One could further enhance this procedure by constructing the null from groups matched on observable characteristics likely to confound the proposed source of excess comovement.[5] Second, we propose comparing the variances and Sharpe ratios of portfolios exposed to a proposed source of comovement to those of matched portfolios that lack this exposure. This test highlights the spirit of studies of excess comovement and exploits the implications for portfolio diversification in the presence of latent factors. In particular, portfolios that exhibit excess comovement will be under-diversified (via their exposure to the proposed source) without a commensurately high return.

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

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 result from rational variation in investors' time preferences or in the prospective cash flows of the underlying assets. For instance, de Bodt et al. (2019) show that correlated operating cash flows result in stronger idiosyncratic return comovement. On the other hand, common 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 these tests evaluated 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, while 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 or 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 nonsynchronous 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 225 index exhibit excess comovement with other stocks in the index, and they comove less with stocks outside the index. Kumar and Lee (2006) find that correlation in retail trading explains the return comovement for stocks that have 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), Barker and Loughran (2007), and Uysal and Hoelscher (2018) find evidence that excess comovement is linked to geography, and Green and Hwang (2009) find excess comovement for stocks within a similar price range. More recently, excess comovement has been documented among stocks that pay dividends (Hameed and Xie, 2019), stocks with high frequency traders (Malceniece et al., Forthcoming), and stocks connected to the same political network (Piotroski et al., 2019). These studies advance some variant of the explanation that excess comovement is caused by correlated sentiment or liquidity needs, thus they 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 that have 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 shared, information to their hedge fund clients, which induces comovement in the returns of clients who trade on this information. Hameed et al. (2015) claim that stocks that have more extensive analyst coverage are priced more accurately, and such "bellwether" stocks lead the price discovery of related firms. In turn, this information spillover causes opaque stocks to comove more strongly with "bellwether" stocks. Box (2018) finds evidence of comovement for stocks with commonality in soft information.

All tests of excess comovement employed in this literature are a joint hypothesis between the asset pricing theory and an empirical model of asset returns. Thus, for the interpretation in these studies to be valid, the empirical model used to adjust returns would have to capture all rational variation in returns. This is a very high bar to cross, and our simulations show that even minor deviations from a perfectly specified empirical model will lead to substantive return comovement. The limitation of this joint hypothesis problem is that investors have more information about the factors that drive returns, the exposure to those factors, and the 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

To illustrate how the presence of an unmodeled factor can affect tests of excess comovement, we consider a standard linear factor representation of asset returns. Under the assumptions of the CAPM of Sharpe (1964) and Lintner (1965), covariance with the market portfolio completely determines the risk of a security and hence its expected return. More generally, the arbitrage pricing theory (APT) of Ross (1976) allows for an arbitrary number of systematic factors. 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. To provide some intuition for this 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 aggregate wealth portfolio (i.e., the value-weighted return of all stocks in CRSP) and Z_t can represent the unobservable component (e.g., human capital). The terms β_i and γ_i are constant for each stock i. We assume that the coefficients β and γ are relatively close to unity, with the average cross-sectional values of each being 1.[6]

In our model, assets are positively exposed to an omitted factor (Z), on average. This assumption offers a slight deviation from typical factor models that assume zero exposure to factors other than the market. These factor models implicitly account for the fact that market factors are estimated as average cross-sectional returns and exposures to other factors are effectively demeaned. However, nothing requires that the original data-generating process have this feature. Indeed, our assumption is empirically motivated to capture differences between value weighting and equal weighting schemes when forming portfolios. To illustrate this point, consider our reduced-form model, in which the omitted factor is a size factor. Given the highly skewed distribution of firm sizes, a random group of stocks is likely to consist of mostly small-cap stocks. As a result, these random groups exhibit positive exposure to the size factor, on average.

We further assume that there are N stocks in the economy and that G denotes a partition of the

[6] As long as the average cross-sectional exposure is not exactly zero, this assumption is without loss of generality.

set $I = \{1, 2, \ldots, N\}$ such that the G_g is the g^{th} element of G. The partition G is the mathematical equivalent of creating subsets of stocks. For example, G can represent grouping stocks by industry classifications, geographical locations, market capitalization, or by any observable criterion.

Common tests of comovement consider the relationship between each stock's return and the average return of all stocks in its group. This method excludes the focal asset's 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 as

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

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 obtain

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

Our assumption about the cross-sectional average of β and γ, combined with the standard Gauss–Markov assumptions, implies that the true values of b and c satisfy $b + c = 1$. Unbiased estimates of the coefficients b and c therefore also reflect this identity. Note that the model does not identify the parameter c, and note that for any partition G, we obtain an estimate of c that is not necessarily zero. 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 exhibits 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 the exposition, noting that the assumption is not necessary for the results that we derive. We can express Eq. (1) in vector form:

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

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 as a latent unpriced factor in an APT framework. Thus, comovement due to this factor does not preclude a rational market interpretation. However, Kozak et al. (2018) show that linear factors cannot distinguish between alternative models of investor beliefs. Thus, the factor Z need not represent a rational source of risk. If Z were instead a latent factor relating to the 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 explanations and sentiment- or friction-based explanations of comovement center around the source of Z.

In the case that Z is driven by behavioral biases or by 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. To operationalize this hypothesis, we test whether the coefficients Γ that correspond to this group are indeed different from zero and have the same sign.[7] While the factor F is observable and its exposure can be quantified, the factor Z is unobservable and we cannot estimate the coefficients Γ directly. We can, however, consider the variance of portfolio returns adjusted for exposure to factor F (F-adjusted) with weight vector $w = [w_1, w_2, \ldots, w_n]'$. We define the F-adjusted returns by:

$$\tilde{r}_t = r_t - r_{ft}\mathbf{1} - F_t B = Z_t \Gamma + \epsilon_t. \tag{4}$$

The F-adjusted return on the portfolio is $w'\tilde{r}_t$ and its variance is $\sigma_p^2 = w'\Gamma\Gamma' w \sigma_Z^2 + \sum_{i=1}^n w_i^2 \sigma_i^2$, 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 (exposure to Z), and Group B is an otherwise identical group of stocks that do not share this feature. To distinguish these 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 these two groups are identical in every aspect aside from 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 whether the Γ coefficients are all positive without loss of generality.

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

Note that the alternate hypothesis implies

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

If the two portfolios are indeed identical along the observable criteria (i.e., the constituents have the same variances σ_i) and the weights on different constituents are the same, then it is easy to show that $\sigma_A^2 = \sum_{i=1}^n w_{Ai}^2 \sigma_i^2$ and $\sigma_B^2 = \sum_{i=1}^n w_{Bi}^2 \sigma_i^2$ are equal in the absence of exposure to the omitted factor. Therefore, the portfolio variances have the following distributions under the null hypothesis:

$$(T-1)\frac{\widehat{var(\tilde{r}_t' w^A)}}{\sigma_A^2} \sim \chi^2(T-1),$$

and

$$(T-1)\frac{\widehat{var(\tilde{r}_t' w^B)}}{\sigma_B^2} \sim \chi^2(T-1).$$

Given that $\sigma_A^2 = \sigma_B^2$, the ratio of these two statistics is

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

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

An extension of this test concerns Sharpe ratios. Lo (2002) shows that under assumptions similar to ours, a portfolio's Sharpe ratio is asymptotically normally distributed. Since 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 null hypothesis, the expected returns of the two portfolios Group A and Group 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 returns and variances of the two portfolios will be identical. Under these asymptotic assumptions, we can employ a t-test of equality for the two portfolios' Sharpe ratios.

3.2 Simulations

Next, we explore the properties of traditional tests of comovement using simulations of the model. We simulate a panel of asset returns using the underlying data generating process:

$$r_{it} - r_{ft} = \beta_i F_t + \gamma_i Z_t + \epsilon_{it}, \tag{5}$$

where F represents an observable factor (e.g., the market factor) such that factor-adjusted returns are

$$\tilde{r}_{it} = r_{it} - r_{ft} - \beta_i F_t = \gamma_i Z_t + \epsilon_{it}. \tag{6}$$

We simulate the ϵ_{it} to be distributed i.i.d. $N(0, 0.183)$, where 18.3% is the average market-adjusted (i.e., F-adjusted) monthly return volatility in the CRSP universe from 1980–2016. The parameter γ is distributed with a cross-sectional average of 1 and a cross-sectional standard deviation of 0.45. We simulate Z_t to have a mean of 0 (i.e., it is unpriced). We repeat these simulations for different values of σ_Z.[8]

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

$$\tilde{r}_{igt} = \alpha + \theta \tilde{r}_{-igt} + \epsilon_{igt}, \tag{7}$$

where $\tilde{r}_{-igt}$ is the average market-adjusted return for group g, excluding the focal stock i.

Table 1 reports simulation results of Eq.(7) for 240 months of returns for 2,400 assets. To explore the sensitivity of comovement estimates to sorting on observable characteristics that proxy for latent factor exposure, we generate a characteristic $X_i = \rho\gamma_i + (1 - \rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$ 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 the volatility of the average monthly value-weighted market return from 1980–2016 ($\sigma_F = 4.52\%$). The rows of each panel correspond to simulations produced with various other asset group sizes (N_g). The median estimate of θ from 1,000 simulations is reported

[8] We explore additional variations in our presumed data-generating process in the Internet Appendix.

for each specification.

Three distinct patterns emerge from the estimates presented in Table 1. First, $\hat{\theta}$ is increasing in the variance of the latent factor Z. Even with $\sigma_Z = 1/8 \times \sigma_F$ (Panel A), we obtain a comovement estimate of 0.131 for groups containing 160 assets, and $\rho = 0$. This estimate increases to 0.913 when $\sigma_Z = \sigma_F$ (Panel D). Second, $\hat{\theta}$ is increasing in the number of assets (N_g) contained in each group used to estimate comovement. For instance, in Column 1 of Panel D (i.e., $\sigma_Z = \sigma_F$ and $\rho = 0$), $\hat{\theta}$ increases from 0.399 when $N_g = 10$ to 0.913 when $N_g = 160$. Third, $\hat{\theta}$ is monotonically increasing in ρ. The first two patterns reinforce the intuition of our model described in Section 3, in which $\hat{\theta}$ increases as the omitted factor constitutes a higher fraction of total return variance. In these simulations, only Z and the idiosyncratic noise ϵ affect returns. If the variance of Z is large or if the effect of the idiosyncratic term ϵ is diversified away in groups containing many assets (large N_g), then the shared exposure to the omitted factor becomes more prominent, leading to a larger $\hat{\theta}$. Finally, the last pattern shows that grouping based on characteristics that are even mildly associated with factor exposure leads to higher estimates of comovement.

We explore additional parameterizations of these simulations in the Internet Appendix, and we find qualitatively similar results. We also explore the market model, in which the single market factor is not perfectly observable. This set up is analogous to the Roll critique (Roll (1977)). Consequently, our assumption of an omitted factor in a multifactor 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.

Of course, we have not provided an economic motivation for the source of factor Z. We have merely modeled Z as a latent factor in an APT framework. Nothing in our setting rules out the possibility that Z could arise because of behavioral biases or market frictions. The purpose of these simulations is to highlight that even inconsequential latent factors can lead to significant comovement estimates. Investors have more information than econometricians regarding the factors that drive returns, the exposure to those factors, and the 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, leads to substantial and ubiquitous within-group return comovement.

4 Empirical Analysis

4.1 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 the period of 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 a monthly frequency. However, some of our analysis requires daily CRSP data on common shares of stocks.

In Section 4.5, 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 2 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 USD 1 billion and a book-to-market equity ratio of 0.77.

Headquarters locations are determined through addresses filed with the Securities and Exchange Commission (SEC) and are obtained through the SEC's EDGAR service. Compustat also stores addresses, but it does not maintain a history of changes to that field in the database. The SEC's EDGAR service provides all filings from 1994 to 2016, which restricts the sample of firm headquarters locations to that time period.[9] Following Pirinsky and Wang (2006), we aggregate firm headquarters locations to the MSA level. 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 1993 to the end of our sample in 2016. Mutual fund equity holdings are obtained

[9]Some studies have used alternative sources for headquarters locations. However, we do not have access to these sources. Headquarters locations change very infrequently and should not impact our results.

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 they 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 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 2 summarizes asset returns and asset group characteristics. For headquarters location, all companies whose headquarters are 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 stock i's price. 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. Last, 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%.

4.2 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 certainly more than five candidate papers for replication that identify sources of excess comovement, we choose to replicate a set of papers that spans 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*, the *Journal of Financial Economics*, the *Review of Financial Studies*, and the *Journal of Financial and Quantitative Analysis*.[10]

The studies we selected examine various sources of comovement related to investor behavior

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

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 headquarters 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 investor behavior. The violations that these studies emphasize stem from behavioral biases of investors 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 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 we could produce for each of the five studies described above. These five studies use slightly different methodologies to detect excess comovement. For comparability, we start by employing the same parsimonious specification that encapsulates the spirit of these studies and that corresponds to our simulations:[11]

$$r_{igt} - r_{ft} = \alpha + \theta(r_{-igt} - r_{ft}) + \beta(r_{mt} - r_{ft}) + \epsilon_{igt}, \tag{8}$$

where r_{igt} represents the return for asset i in month t; r_{-igt} represents the average return of all assets in the same group g as asset i, excluding asset i from its own group return calculation; and r_{mt} is the market return. 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 headquarters 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 estimates of the coefficient θ are positive and both statistically and

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

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 the entire 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.

4.3 Comovement for randomly grouped assets

As derived in our theoretical motivation in Section 3 and shown in our simulations in Section 3.2, omitted factors can lead to substantive residual return comovement. In light of this implication, a null hypothesis of zero leads to an overstatement of excess comovement and a tendency to over-reject the null. 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 would 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 placebo procedure whereby the economically motivated asset groups are replaced by a randomly selected group of assets. For instance, in the Pirinsky and Wang (2006) setting, we estimate the return comovement between firm i and 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 an empirical null distribution to compare to the replicated coefficient estimate.

The confidence intervals of the placebo procedure for each replication are reported in the right half of Table 3. For each panel, the corresponding placebo confidence intervals at the 1%, 5%, and 10% levels are reported, as well as the median value from all sample runs. 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 an imperfect market proxy that causes the returns of seemingly unrelated assets to comove strongly. Moreover, common mutual fund ownership is the only source of comovement in our replications that exceeds the 10% upper bound of the corresponding confidence interval. However, even for common mutual fund ownership, the placebo confidence interval suggests that the original coefficient estimate severely overstates "excess" comovement.

The positive comovement estimates for randomly grouped assets are consistent with our theoretical model. In this model, all assets are positively exposed to an omitted factor, which leads to a positive correlation between any two asset returns, on average. Adjusting returns for market exposure exffectively demeans exposure to non-market factors. However, if portfolios use equal-weighted returns, they exhibit non-zero exposure to (value-)demeaned factors. For instance, given the highly skewed distribution of firm sizes, a group of randomly selected stocks contains mostly small stocks. As a result, these random groups exhibit positive exposure to the SMB factor, on average. To investigate this implication, we change the random selection criteria so they are proportional to firm size, such that the probability of sorting a stock into a group equals the fraction of that stock's market capitalization relative to overall market capitalization. Therefore, the expected return of each group of randomly selected assets equals the value-weighted market return. In the Internet Appendix, we present results that illustrate comovement estimates using these size-adjusted selection criteria are significantly smaller than the estimates in Table 3.[12]

These 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 comovement from a placebo group should be used as the null 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 confidence intervals suggest that comovement within stocks grouped by headquarters MSA can be entirely explained by factors unrelated to location. Our results suggest similar takeaways for the other sources of comovement we consider. In summary, attributing comovement to the proposed economic motivations is challenging without an appropriate counterfactual.

[12]We thank John Campbell for suggesting this exercise.

4.4 Adjusting for multifactor models

The models presented in Table 3 use only excess market returns as a control variable. However, some studies have acknowledged the potential for omitted factors to influence excess comovement estimates. To mitigate this concern, these studies often control for multifactor models that perform better than the single-factor market model 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. However, if controlling for additional factors sufficiently accomplishes this task, 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 4.3.

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 4.3, 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 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 placebos are presented in Panel A of Table 4. Each row of Panel A corresponds to various 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 market model, 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 simulations increases monotonically from 0.48, 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 are diversified away and shared exposure to the omitted factor becomes more prominent. Using the market-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 for groups of 160 assets. However, the estimates of comovement remain positive for all portfolio sizes and for all factor models used to adjust returns.

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 0.33 for the groups containing 160 randomly selected stocks. This finding suggests that the omitted factor bias is quite pervasive and that 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.

4.5 Characteristic sorts

An alternative approach to using empirical factor models is to adjust returns according to asset characteristics (see Daniel et al. (1997)). To the extent that factor models fail to capture the predictability in cross-sectional returns, commonality among assets along 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 4.4.

For this analysis, we consider five of the most common characteristics: *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 10 smallest stocks (according to market cap) in group 1, the next 10 smallest stocks in group 2, and so on. We then regress residuals from a factor model on the average residuals for each 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 are 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 according to 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 that we consider exhibit positive comovement estimates despite controlling for an empirical factor that corresponds to the focal characteristic. For instance, the comovement for groups sorted on size exhibit substantive comovement after controlling for exposure to the SMB factor. This finding highlights the inadequacy of adjusting returns for commonly used empirical factor models.

In Columns 6-8 of Table 4, we adjust returns for the first five (PCA5), 10 (PCA10), and 20 (PCA20) principal factors. The comovement estimates under these specifications continue to exhibit a strong positive relationship, and they are generally of the same magnitude as the estimates derived from the six-factor model. Operating profitability produces the lowest comovement estimate, 0.17, even after adjusting for the first 20 principal factors (for $N_g = 160$).

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 do not fully capture the effect of characteristics on realized returns. 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 in observable characteristics leads to substantive comovement estimates. However, similarities in observable characteristics also likely 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 one or more omitted factors or similar characteristics. The results in Table 4 also reaffirm the inadequacy of the zero null hypothesis.

4.6 Intensity-based tests

As we illustrate in Sections 4.3 - 4.5, 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; instead, they 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 headquarters locations (Barker and Loughran (2007)) and the degree of common analyst coverage (Israelsen (2016)). These tests implicitly assume that the degree of similarity in the chosen criterion is uncorrelated with exposure to omitted factors, thus circumventing the latent factor bias.

While we cannot directly test for similarities in latent factor exposure, we can consider similarities in various observable determinants of risk. Specifically, we estimate how the intensity of comovement is related to the distance between each of the six factor loadings and five characteristic variables that we consider in Section 4.5. In Panel A of Table 5, we estimate the following specification:

$$\rho_{i,j,t} = \lambda \frac{-|x_{i,t} - x_{j,t}|}{\sigma(X_t)} + \varepsilon_{i,j,t}, \tag{9}$$

where $\rho_{i,j,t}$ is the pairwise stock return correlation[13] between stock i and stock j in year t, and $\sigma(X_t)$ is the cross-sectional standard deviation of characteristic x.

Table 5 presents results for similarities in factor loadings. Each entry of Panel A corresponds to an estimate of λ from a univariate regression. In Column 1, we use (raw) excess returns. Each subsequent column adjusts returns according to the market model, the Fama–French three- and

[13] Stock return correlations are estimated using daily returns.

five-factor models (3 FM and 5 FM), and the Fama–French five-factor model augmented with the momentum factor (6 FM). In Columns 7-9, we adjust returns for principal factors from an ex post principal component analysis using the first five (PCA5), 10 (PCA10), and 20 (PCA20) factors. In all cases, λ 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 six factor loadings on pairwise stock return correlations are estimated simultaneously.

Two patterns that arise from these results are worth noting. First, using additional factors to calculate residuals always attenuates the pairwise correlations attributed to factor similarity. However, even adjusting returns for the first 20 ex post principal components does not fully attenuate estimates. Second, the coefficient estimates in the multivariate regressions remain highly statistically significant, but they are smaller in virtually all cases compared to the univariate estimates. This reduction in point estimates highlights the fact that factor exposures are correlated. In other words, similar exposure to one factor likely indicates similar exposure to other factors. This finding is consistent with that of de Bodt et al. (2019), who show that two firms with a high pairwise return correlation have greater similarity in product market portfolios, which would lead to similar cash flows and similar discount rates. Thus, it is unlikely that assets that are similar along observable criteria do not share similar exposure to omitted factors. In conjunction, these findings demonstrate that an intensity-based test design is likely to suffer from a latent factor bias, which reaffirms the intuition from our characteristic-based sorts described in Section 4.5.[14]

Daniel et al. (1997) claim that "characteristics rather than the covariance structure of returns appear to explain the cross-sectional variation in stock returns." In light of their results, we repeat the analysis of Model (9) using the five asset characteristics considered in Section 4.5 and report the results in Table 6. Similar to the factor loading results, all characteristic similarities are associated with higher pairwise correlations, and the relationship attenuates when richer factor models are used to adjust returns. Our chosen characteristics are a small subset of those documented in the literature as having an association with returns. Consequently, intensity-based tests have the high hurdle of proving that a proposed explanation is not due to an omitted characteristic or factor.

[14] In Table IA10, we report results from simulations of Equation 9 according to the DGP in Equation 6 of Section 3.2. These simulations reaffirm our findings in this section.

4.7 Shock-based tests

While some studies have opted for the intensity-based design, others have recognized the limitations of standard tests of comovement by implementing a shock-based test design. For example, changes in comovement have been examined surrounding stock splits (Green and Hwang (2009)), headquarters relocations (Pirinsky and Wang (2006)), and S&P 500 index inclusions (Barberis et al. (2005)). These studies investigate shocks that ostensibly alter the nature of return comovement without affecting asset fundamentals (i.e., risk or cash flows). If the shock in question is truly exogenous to underlying fundamentals, significant changes in comovement would indicate a violation of rational asset pricing theory. However, shocks that are either caused by or lead to evolving fundamentals result in changes in comovement that may be challenging to separate from the proposed channel.

To illustrate this point, we simulate shocks that cause a change in latent factor exposure and we explore the impact on comovement estimates within our framework. We start by repeating our simulation in Section 3.2, except we increase the exposure to the latent factor Z halfway through the sample period (i.e., at month 240) for one asset in each randomly sorted group. Specifically, for a select number of random assets, we increase the factor loading γ_i by σ_γ. This procedure mimics a setting in which an asset experiences a shock that moves it from one group to another. Yet, instead of the shock having no effect on fundamentals, we impose that the shock alters the asset's latent factor exposure.

Table 7 reports the median within-group comovement estimates for each asset with an altered factor loading, both before and after the change. We repeat the exercise described above for various parameter choices regarding group size (N_g = 10, 20, 40, 80, and 160) and the volatility of Z, expressed as a multiple of the volatility of F (1/4, 1/2, 1, 2). In every specification, the magnitudes of comovement estimates increase after the parameter change. 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. These findings illustrate that changes in factor loadings are sufficient to generate substantive changes in comovement, even though only one asset in each group received the shock. Therefore, the validity of shock-based tests of comovement rely heavily on factor loadings remaining unchanged surrounding the shock.

Next, we turn to real data to explore changes in comovement estimates and factor loadings in a variety of shock-based settings proposed in the literature. We start by exploring index additions following the work of Barberis et al. (2005). Barberis et al. (2005) show that comovement with respect to the S&P 500 index increases for stocks after their addition to the index, and the authors attribute this increase in comovement to the trading behavior of "style" investors. We replicate their findings by calculating the changes in comovement ($\beta_{S\&P}$) over symmetric windows surrounding each addition to the S&P 500 index. However, Denis et al. (2003) show that S&P 500 inclusions are associated with substantive changes in earnings. In light of these findings, we also consider changes in comovement surrounding additions to the Russell 1000 and 2000 indices, which have less subjective inclusion criteria.

We present our results in Panel A of Table 8. Changes in comovement estimates are positive and statistically significant at conventional levels for the S&P 500 for all event windows. The Russell 1000 exhibits statistically insignificant changes (at the 5% level) for all windows. Estimates for the Russell 2000 are positive and statistically significant for all event windows except the 4- and 5-quarter windows. The significant changes for the Russell 2000, but not for the Russell 1000, are likely driven by the fact that additions to the Russell 2000 typically become some of the largest stocks in the index (i.e., stocks that drop from the 1000 into the 2000). Consequently, the value-weighted index is highly correlated with these additions, resulting in higher beta estimates for the stock relative to the index. On the contrary, additions to the Russell 1000 typically become the smallest stocks in the index (i.e., they are typically added from the Russell 2000).

Next, we explore changes in factor loadings surrounding inclusion events for each of the three indices. Factor loadings are estimated for each stock one year before and one year after inclusion in its respective index. Table 8 presents loadings for the Fama–French five factors as well as the UMD momentum factor. For S&P 500 index additions, only the change in UMD exposure is statistically significant (i.e., it decreases by 0.04). This finding is consistent with that of Chen et al. (2016), who show that firms typically experience high returns leading up to their inclusion in the S&P 500 index. Index additions to the Russell 1000 and 2000 indices experience large changes in several factor exposures. For instance, Russell 1000 additions experience a 0.17 decrease in exposure to the SMB factor. This change is consistent with relatively small firms being added to the Russell 1000 as a result of rapid growth in market capitalization. Thus, inclusion in the index coincides with decreased exposure to the SMB factor. We also replicate the shock-based analysis for stock splits outlined in Green and Hwang (2009) and for headquarter relocations outlined in Pirinsky and Wang (2006). We find significant changes in factor loadings surrounding these events. The results are presented in the Appendix (Table IA11).

Our simulations emphasize that factor loadings must be stable in order for shock-based settings to be valid. However, our findings also suggest that significant changes in factor loadings are common surrounding plausibly exogenous events. Thus, testing for changes in factor loadings should constitute an important validation of any shock-based empirical design to study comovement. These findings complement the work of Chen et al. (2016), who argue that, in order for shock-based tests to provide evidence of excess comovement, the reassignment of an asset from an old group to a new group must be associated with an increase in comovement with the new group and a simultaneous decrease in comovement with the old group. An increase in the comovement with both groups likely violates the exogeneity assumption. Evolving factor exposure, however, could explain why the comovement of an asset would increase with both the source and destination groups following a shock.

4.8 Variance and 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 nonzero correlation between returns with no impact on expected return levels.

Under friction-based explanations, within-group excess comovement is an indication that investors are not efficiently diversified (i.e., portfolios with excess comovement will be more volatile and therefore exhibit low Sharpe ratios, on average). In contrast, an investor facing fewer frictions or subject to less behavioral bias could diversify more effectively. Thus, if an omitted systematic factor is the reason for comovement, then there are no diversification benefits to be realized by considering assets outside the group. For example, suppose Investor A elects to overweight the stocks of firms headquartered in her MSA, while Investor B holds a more geographically dispersed portfolio. If geography is a source of comovement, then Investor A is restricting her diversification benefits compared to Investor B.

To formalize this intuition, we propose a variance ratio test in the spirit of Gibbons et al. (1989). Continuing with the example of Investors A and B, we will refer to the variance of Investor A's portfolio as σ_A^2 and that of Investor B's portfolio as σ_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 variances would be equivalent. Thus, under the null, $\sigma_A^2 = \sigma_B^2$, or $\sigma_A^2/\sigma_B^2 = 1$. However, under the alternative hypothesis that excess comovement exists for the assets in Investor A's portfolio: $\sigma_A^2 > \sigma_B^2$, or $\sigma_A^2/\sigma_B^2 > 1$. We showed in Section 3.1 that a variant of this statistic follows an F distribution under the null hypothesis of equivalent volatilities (i.e., 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 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 the focal stock's MSA, but this pool is not restricted to firms that belong to a particular location. Note that only matching on firm size is a fairly lenient restriction, and 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 9 presents results from this analysis. For each of the settings we consider, we construct equally-weighted and value-weighted portfolios and calculate the residual variance with respect to the market portfolio. We form portfolios of assets according to common headquarters 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) variance ratio, the t-stat, and the number of portfolios for which we reject the null of unity. The last column of Table 9 presents the number of test portfolios (N).

In many of the settings we consider, both the average and median ratios of variances are close to unity. Shared headquarters location provides the largest deviations of the ratio from 1 with a mean ratio of 1.48 (1.43) for equally-weighted (value-weighted) portfolios. In this setting, we reject the null 36−37% of the time.

Table 10 presents additional tests of the variances and Sharpe ratios of the portfolios in Table 9 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 shared 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 present only among portfolios formed on the basis of common analyst coverage and similar share price.

5 Conclusion

We revisit the question of excess comovement using a linear factor structure in returns. Using this framework, we show that residual comovement is a ubiquitous feature of asset returns in the presence of a latent systematic factor. We confirm this ubiquity using both simulations and real data. Adjusting returns for additional empirical factors strongly attenuates, but does not eliminate, comovement within groups of randomly sorted assets. Thus, a null hypothesis of zero comovement can lead to a severe overstatement of "excess" comovement. Furthermore, grouping assets by characteristics, which offers an alternative to linear factor adjustments, generally magnifies comovement estimates. Therefore, a more appropriate null should also account for similarity in characteristics.

Our study remains silent on the source of latent systematic factors, which can arise for either rational or behavioral reasons. However, our study highlights the fact that all tests of excess comovement are a joint hypothesis between an asset pricing theory and an empirical model of returns. Thus, in order to attribute comovement to informational frictions, behavioral preferences, or market segmentation, the empirical model used to adjust returns must capture all rational variation in returns. Furthermore, attributing comovement to a particular source requires the formidable task of controlling for all alternative sources of systematic variation. While these findings do not rule out behavioral or friction-based explanations of comovement, they do highlight the limitations of commonly used tests, and they suggest that several proposed explanations need to be revisited.

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.

Box, T. 2018. Qualitative similarity and stock price comovement. *Journal of Banking and Finance* 91:49–69.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of finance* 52:57–82.

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* 52:1035–1058.

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

de Bodt, E. J., E. B. Eckbo, and R. W. Roll. 2019. Corporate rivalry and return comovement. *Working Paper* .

Denis, D. K., J. J. McConnell, A. V. Ovtchinnikov, and Y. Yu. 2003. S&P 500 index additions and earnings expectations. *The Journal of Finance* 58:1821–1840.

Fama, E. F. 1991. Efficient capital markets: II. *Journal of Finance* 46:1575–1617.

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* 57:1121–1152.

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.

Hameed, A., and J. Xie. 2019. Preference for dividends and return comovement. *Journal of Financial Economics* 132:103–125.

Hoberg, G., and G. Phillips. 2016. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124:1423–1465.

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.

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73:1183–1223.

Kumar, A., and C. M. Lee. 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61:2451–2486.

Lo, A. W. 2002. The statistics of Sharpe ratios. *Financial analysts journal* 58:36–52.

Malceniece, L., K. Malcenieks, and T. J. Putnins. Forthcoming. High frequency trading and co-movement in financial markets. *Journal of Financial Economics* .

Piotroski, J. D., T. Wong, and T. Zhang. 2019. Political Networks and Stock Price Comovement: Evidence from network-connected firms in China. *Working Paper* .

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

Pollet, J. M., and M. Wilson. 2010. Average correlation and stock market returns. *Journal of Financial Economics* 96:364–380.

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* 71:421–436.

Sias, R., H. J. Turtle, and B. Zykaj. 2017. Hedge fund return dependence: Model misspecification or liquidity spirals? *Journal of Financial and Quantitative Analysis* 52:2157–2181.

Uysal, V., and S. Hoelscher. 2018. Local Clientele: geography and comovement in stock returns. *Review of Behavioral Finance* 10:231–251.

Table 1: Simulations: Latent Factors and Characteristics

This table reports simulation results of Eq.(5) in Section 3 for 240 months of returns for 2,400 assets. We define groups g to contain 10, 20, 40, 80, or 160 assets. For each group size, we report the median estimate of θ in the regression: $\tilde{r}_{igt} = \alpha + \theta\tilde{r}_{-igt} + \epsilon_{it}$, where $\tilde{r}_{igt}$ is asset i's market-adjusted return at time t and $\tilde{r}_{-igt}$ is the market-adjusted average return of group g excluding asset i. We form groups based on sorts of characteristic $X_i = \rho\gamma_i + (1-\rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. 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.

	$\rho = 0$	$\rho = .1$	$\rho = .25$	$\rho = .5$	$\rho = .75$	$\rho = .9$	$\rho = 1$
			Panel A: $\sigma_Z = 1/8 \times \sigma_F$				
10	0.0088	0.0104	0.0088	0.0114	0.0103	0.0126	0.0125
20	0.0176	0.0194	0.0185	0.0209	0.0202	0.0232	0.0238
40	0.0374	0.0382	0.0385	0.0426	0.0433	0.0440	0.0444
80	0.0717	0.0730	0.0721	0.0790	0.0823	0.0873	0.0863
160	0.1305	0.1410	0.1432	0.1455	0.1578	0.1606	0.1577
			Panel B: $\sigma_Z = 1/4 \times \sigma_F$				
10	0.0384	0.0382	0.0379	0.0405	0.0434	0.0454	0.0454
20	0.0722	0.0737	0.0746	0.0787	0.0855	0.0870	0.0846
40	0.1346	0.1377	0.1377	0.1477	0.1563	0.1587	0.1552
80	0.2393	0.2347	0.2422	0.2554	0.2723	0.2705	0.2740
160	0.3839	0.3850	0.3924	0.4089	0.4244	0.4230	0.4285
			Panel C: $\sigma_Z = 1/2 \times \sigma_F$				
10	0.1361	0.1357	0.1380	0.1482	0.1610	0.1619	0.1611
20	0.2405	0.2402	0.2455	0.2554	0.2753	0.2815	0.2768
40	0.3908	0.3900	0.3902	0.4124	0.4273	0.4353	0.4351
80	0.5618	0.5594	0.5647	0.5863	0.5935	0.6079	0.6034
160	0.7153	0.7174	0.7204	0.7353	0.7518	0.7538	0.7508
			Panel D: $\sigma_Z = 1 \times \sigma_F$				
10	0.3993	0.3995	0.4030	0.4226	0.4398	0.4463	0.4549
20	0.5674	0.5683	0.5718	0.5958	0.6154	0.6184	0.6266
40	0.7248	0.7233	0.7217	0.7390	0.7587	0.7689	0.7649
80	0.8418	0.8397	0.8452	0.8539	0.8635	0.8649	0.8662
160	0.9126	0.9139	0.9141	0.9200	0.9266	0.9269	0.9281
			Panel E: $\sigma_Z = 2 \times \sigma_F$				
10	0.7579	0.7567	0.7619	0.7860	0.8099	0.8089	0.8106
20	0.8627	0.8622	0.8641	0.8811	0.8928	0.8969	0.8953
40	0.9249	0.9259	0.9271	0.9358	0.9434	0.9457	0.9452
80	0.9616	0.9613	0.9627	0.9664	0.9704	0.9720	0.9712
160	0.9803	0.9801	0.9810	0.9831	0.9848	0.9856	0.9851
			Panel F: $\sigma_Z = 4 \times \sigma_F$				
10	0.9795	0.9778	0.9824	1.0002	1.0144	1.0165	1.0166
20	0.9905	0.9879	0.9910	0.9993	1.0071	1.0076	1.0084
40	0.9948	0.9955	0.9960	0.9999	1.0035	1.0037	1.0041
80	0.9976	0.9973	0.9980	1.0000	1.0016	1.0020	1.0018
160	0.9986	0.9988	0.9990	1.0000	1.0007	1.0010	1.0011

Table 2: Summary Statistics

The table provides descriptive statistics of the main sample of monthly CRSP stock returns (Panel A) and the average excess return (in %) for the assets in each of our replication samples (Panel B). For the first four replication settings, we use monthly stock returns from CRSP, while for the last setting, we use monthly hedge fund returns from the Thomson Reuters Lipper Hedge Fund Database. Panel A presents the means, medians, standard deviations, and 10th and 90th 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 peer groups and the market. The last columns report the average number of assets in each peer group and total number of observations in each sample. The five grouping criteria relate to the Metropolitan Statistical Areas (MSAs) of firm headquarters locations, common analyst coverage, individual share price level, common mutual fund ownership, 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 3: Sources of Comovement and Confidence Intervals for Placebo Null

The table presents estimates of the model $r_{igt} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \theta(r_{-igt} - r_{ft}) + \epsilon_{it}$, where r_{igt} represents asset i's returns, r_{-igt} represents the average return of all other assets in the same group (g) as i, r_{mt} is the return on the market, and r_{ft} is the risk free rate. The groups considered correspond to the following potential sources of comovement: headquarters 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 confidence intervals are calculated by randomly assigning assets to groups and estimating the model on the placebo data. For panels A–D we use monthly stock returns from CRSP. In Panel E, we use monthly hedge fund returns from the Thomson Reuters Lipper Hedge Fund Database.

	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								
	Common Ownership	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 various 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, book-to-market, momentum, asset growth,* and *operating profitability*, respectively. The rows of each panel correspond to various group sizes, (10 to 160). Column 1 uses (raw) excess returns. Each subsequent column adjusts returns for market returns (Mkt), 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), 10 (PCA10), and 20 (PCA20) factors. The sample uses monthly returns for all CRSP/Compustat stocks with available data from 1970 to 2016.

# Stocks	Raw	Mkt	3 FM	5 FM	6 FM	PCA5	PCA10	PCA20
			Panel A: 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
			Panel B: 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
			Panel C: 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
			Panel D: 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
			Panel E: 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
			Panel F: 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: Pairwise Return Correlations and Factor Exposure

This table presents regression estimates for pairwise stock return correlations on a measure of similarity in factor loadings with respect to the Fama–French 5-factor model augmented with the momentum factor. Similarities in factor loadings are calculated as $-|x_{i,t} - x_{j,t}|/\sigma(x_t)$, where $x_{i,t}$ and $x_{j,t}$ represent the factor loadings for stocks i and j, and $\sigma(x_t)$ is the cross-sectional standard deviation of factor loading x at time t. In Panel A, we estimate univariate regressions of each factor loading similarity. In Panel B, we estimate a multivariate regression for all factor loading similarities. In each panel, we consider pairwise correlations of residuals from a specific asset pricing model. These models correspond to the columns of the table: Mkt, 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). Following Anton and Polk (2014), the sample uses monthly returns for all stocks in the S&P 500 index that have above-median market capitalization and available data from 1970 to 2016.

	Mkt	3FM	5FM	4FM	6FM	PCA5	PCA10	PCA20
Panel A. Univariate regressions								
Mkt-rf	0.0060	0.0060	0.0051	0.0053	0.0050	0.0085	0.0067	0.0037
	(10.49)	(12.04)	(16.67)	(14.76)	(18.18)	(6.26)	(8.71)	(20.76)
SMB	0.0099	0.0058	0.0051	0.0054	0.0050	0.0060	0.0058	0.0046
	(16.65)	(18.20)	(20.86)	(19.35)	(20.53)	(7.54)	(11.91)	(11.40)
HML	0.0079	0.0062	0.0050	0.0053	0.0047	0.0049	0.0052	0.0040
	(8.99)	(12.94)	(21.94)	(15.84)	(25.04)	(4.30)	(6.69)	(12.11)
UMD	0.0079	0.0067	0.0054	0.0054	0.0049	0.0058	0.0050	0.0033
	(7.01)	(12.13)	(16.85)	(17.03)	(18.81)	(4.04)	(6.33)	(14.62)
RMW	0.0056	0.0057	0.0046	0.0049	0.0044	0.0066	0.0052	0.0034
	(7.70)	(11.73)	(19.85)	(16.39)	(21.31)	(5.73)	(7.67)	(16.32)
CMA	0.0070	0.0064	0.0050	0.0053	0.0046	0.0064	0.0051	0.0035
	(9.92)	(16.24)	(22.71)	(21.56)	(27.22)	(4.14)	(5.46)	(13.59)
Panel B. Multivariate regressions								
Mkt-rf	0.0032	0.0039	0.0034	0.0034	0.0034	0.0069	0.0051	0.0025
	(8.00)	(9.95)	(12.44)	(12.27)	(14.11)	(5.86)	(7.33)	(18.39)
SMB	0.0074	0.0032	0.0031	0.0033	0.0031	0.0037	0.0039	0.0033
	(14.39)	(13.01)	(15.46)	(14.50)	(15.60)	(7.45)	(12.41)	(10.26)
HML	0.0047	0.0032	0.0026	0.0028	0.0025	0.0016	0.0026	0.0023
	(6.57)	(9.36)	(11.84)	(10.66)	(13.49)	(2.52)	(5.28)	(8.52)
UMD	0.0046	0.0040	0.0031	0.0030	0.0028	0.0028	0.0026	0.0016
	(4.03)	(7.12)	(8.14)	(8.35)	(8.97)	(2.60)	(4.30)	(10.58)
RMW	0.0016	0.0027	0.0022	0.0024	0.0021	0.0037	0.0026	0.0016
	(3.08)	(6.92)	(12.90)	(12.42)	(17.21)	(5.48)	(6.26)	(10.07)
CMA	0.0025	0.0028	0.0020	0.0022	0.0017	0.0032	0.0019	0.0012
	(4.49)	(7.30)	(9.55)	(8.57)	(9.44)	(2.50)	(2.42)	(7.19)

Table 6: 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 $x_{i,t}$ and $x_{j,t}$ represent the characteristics for stocks i and 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. In Panel B, we estimate a multivariate regression for all characteristics. In each panel, we consider pairwise correlations of residuals from a specific asset pricing model. These models correspond to the columns of the table: Mkt, 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). Following Anton and Polk (2014), the sample uses monthly returns for all stocks in the S&P 500 index that have above-median market capitalization and available data from 1970 to 2016.

	Mkt	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 7: Simulations: Shock-based Comovement

This table reports simulation results of changes in comovement 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 assign groups 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 for all time periods $t > 240$ (i.e., the halfway point in the sample). For each group size, we report the median estimate of θ in the regression: $r_{igt} = \alpha + \beta r_{mt} + \theta r_{-igt} + \epsilon_{it}$, 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 estimates θ only for assets that received a shock both before and after the change in factor loading ($0 < t \leq 240$ and $240 < t \leq 480$). Each column corresponds to a different value of σ_Z, expressed as a multiple of σ_F. For each specification, we report the median θ estimate from 1,000 simulations.

	$0 < t \leq 240$				$240 < t \leq 480$			
	$\sigma_Z =$				$\sigma_Z =$			
# 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

Table 8: Index Additions, Changes in Comovement, and Factor Loadings

This table presents the change in factor loadings before and after index inclusions to the S&P 500, Russell 1000, and Russell 2000 indices. Panel A presents changes in comovement relative to the corresponding index (excluding the added stock), and Panel B presents factor loadings with respect to the Fama-French 5-factor model augmented with the momentum factor. For both panels, the factor coefficients are estimated using daily returns during a symmetric window centered around the index addition date as in Barberis et al. (2005). We require that at least two-thirds of daily return data be available during each estimation window and winsorize the estimates at the 1% and 99% levels. Panel A uses various windows around each index addition date ranging from 2–24 quarters. Factor loadings in Panel B are estimated over a 2-year window surrounding each index addition. Index inclusions in the S&P 500, Russell 1000, and Russell 2000 indices are from 1991 to 2013.

Panel A: Changes in Destination Group Comovement

	S&P 500				Russell 1000				Russell 2000			
	Pre-	Post-	Difference	t-stat	Pre-	Post-	Difference	t-stat	Pre-	Post-	Difference	t-stat
Qtrs: -1 to 1	1.0835	1.1519	0.0684	3.61	0.7441	0.7295	-0.0146	-1.24	0.5742	0.6516	0.0774	6.57
Qtrs: -2 to 2	1.1005	1.1764	0.0759	5.32	0.7224	0.7322	0.0098	1.05	0.6321	0.6702	0.0381	4.49
Qtrs: -3 to 3	1.1219	1.1759	0.0540	3.76	1.0053	1.0046	-0.0007	-0.09	0.7937	0.8146	0.0209	2.88
Qtrs: -4 to 4	1.1101	1.1728	0.0627	4.48	1.2432	1.2317	-0.0115	-1.58	0.9327	0.9296	-0.0031	-0.50
Qtrs: -5 to 5	1.0819	1.1659	0.0840	6.79	1.3330	1.3308	-0.0022	-0.33	1.0123	1.0224	0.0101	1.85
Qtrs: -6 to 6	1.0873	1.1648	0.0775	6.13	1.3670	1.3709	0.0039	0.64	1.0649	1.0781	0.0132	2.85
Qtrs: -7 to 7	1.083	1.1569	0.0739	6.13	1.3683	1.3692	0.0009	0.16	1.1743	1.1888	0.0145	3.60
Qtrs: -8 to 8	1.0773	1.1602	0.0829	7.00	1.3248	1.3266	0.0018	0.31	1.2890	1.2994	0.0104	2.75
Qtrs: -9 to 9	1.0781	1.1529	0.0748	6.68	1.4262	1.4315	0.0053	1.00	1.3960	1.4111	0.0151	4.33
Qtrs: -10 to 10	1.0754	1.1502	0.0748	6.80	1.2596	1.2644	0.0048	0.94	1.2235	1.2374	0.0139	4.17
Qtrs: -11 to 11	1.0698	1.1452	0.0754	7.00	1.3519	1.3614	0.0095	1.84	1.2589	1.2739	0.0150	4.67
Qtrs: -12 to 12	1.0652	1.1442	0.0790	7.24	1.4692	1.4747	0.0055	1.12	1.4369	1.4492	0.0123	3.99

Panel B: Changes in Factor Loadings

	S&P 500				Russell 1000				Russell 2000			
	Pre-	Post-	Difference	t-stat	Pre-	Post-	Difference	t-stat	Pre-	Post-	Difference	t-stat
Mkt - Rf	1.080	1.080	0.000	0.040	1.159	1.082	-0.076	-5.540	1.000	1.092	0.092	9.440
SMB	0.060	0.042	-0.018	-1.310	0.814	0.646	-0.169	-9.990	0.898	1.031	0.133	11.070
HML	0.077	0.090	0.013	0.640	-0.040	-0.094	-0.054	-1.990	-0.011	-0.021	-0.010	-0.590
RMW	-0.121	-0.143	-0.022	-1.020	-0.166	-0.297	-0.131	-4.480	-0.209	-0.187	0.021	1.060
CMA	0.031	0.002	-0.030	-1.290	-0.080	-0.160	-0.081	-2.850	0.090	-0.006	-0.096	-4.970
UMD	-0.080	-0.119	-0.040	-2.570	0.116	0.032	-0.084	-5.000	-0.014	-0.089	-0.075	-6.470

Table 9: Variance Ratio Test

This table presents estimates for the ratio of portfolio volatilities of stocks formed on the basis of common headquarters 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 its 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, we form equally weighted (value-weighted) portfolios of the original assets and their matches. The table reports the mean and median ratio of variances of the sample portfolios and their matched portfolios, as well as the number of instances in which the null of equal ratios is rejected by a one-sided F-test. Portfolio variances are calculated with at least three years of monthly data in each case. The first four samples use monthly returns for all CRSP stocks that have 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. Headquarters location							
Coef	1.48	1.00	37	1.43	1.05	38	104
t-stat	5.48			6.06			
Panel B. Common analyst coverage							
Coef	0.62	0.50	191	0.61	0.50	185	5018
t-stat	55.14			69.94			
Panel C. Same share price grouping							
Coef	1.05	1.11	26	0.72	0.73	2	50
t-stat	27.90			22.86			
Panel D. Connected stocks through mutual fund holdings							
Coef	0.95	0.99	484	1.47	1.24	395	5172
t-stat	237.56			92.77			
Panel E. Common prime broker							
Coef	1.23	0.86	12	1.43	1.02	14	39
t-stat	6.01			6.69			

Table 10: Portfolio Volatilities and Sharpe Ratios

This table presents volatility and Sharpe ratio estimates for stocks grouped by common headquarters 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 a 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 the t-statistic of the significance of the difference. Portfolio variances and Sharpe ratios are calculated with at least three years of monthly data in each case. The first four samples use monthly returns for all CRSP stocks that have 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. Headquarters location												
Coef	0.292	0.314	-0.022	0.388	0.371	0.017	0.294	0.310	-0.016	0.358	0.344	0.014
t-stat	26.241	22.002	-1.65	13.845	13.897	0.74	22.123	19.162	-1.08	16.167	12.266	0.49
Panel B. Common analyst coverage												
Coef	0.244	0.213	0.030	0.466	0.510	-0.043	0.208	0.181	0.026	0.430	0.467	-0.037
t-stat	217.053	233.634	27.71	78.871	108.460	-8.7	224.003	243.494	32.5	72.849	85.796	-7.17
Panel C. Same share price grouping												
Coef	0.276	0.271	0.006	0.412	0.409	0.004	0.283	0.246	0.037	0.354	0.386	-0.032
t-stat	23.641	41.853	0.96	58.470	52.374	0.39	23.716	42.570	5.79	34.350	67.927	-2.73
Panel D. Connected stocks through mutual fund holdings												
Coef	0.170	0.165	0.007	0.361	0.365	-0.028	0.148	0.145	0.005	0.346	0.346	-0.010
t-stat	352.979	363.097	25.13	48.944	49.017	-3.79	368.610	390.725	18.48	35.626	35.572	-1.29
Panel E. Common prime broker												
Coef	0.117	0.099	0.018	0.571	0.675	-0.104	0.119	0.104	0.015	0.618	0.667	-0.050
t-stat	12.691	22.872	2.09	8.582	13.459	-1.6	9.742	22.688	1.39	8.532	11.597	-0.7

Ubiquitous Comovement

Internet Appendix

William Grieser, Jung Hoon Lee, and Morad Zekhnini

A Additional Empirical Analysis

In this Internet Appendix, we outline supplemental analyses that test the robustness of our findings from the main text. We start by illustrating the time series of comovement estimates in Figure 1. We then report confidence intervals (in Table IA1) for the simulation point estimates presented in Table 1 of the main text, and we explore variations in our simulation parameters (presented in Tables IA2-IA4).

In Table IA7, we repeat our analysis of Table 4 in the main text for industry adjusted returns. Specifically, we adjust returns according to the Fama–French 12-, 30- and 48- industry portfolios, as well as the text-based network industry classifications (TNIC) 25 industries defined in Hoberg and Phillips (2016). In Table IA8, we repeat our analysis of Table 4, using characteristic adjusted returns following the procedure outlined in Daniel et al. (1997). In Table IA9 we reconsider our pairwise stock return correlation analysis presented in Table 6 of the main text for similarities in common mutual fund ownership (Anton and Polk (2014)), analyst overlap (Israelsen (2016)), geographic distance (Barker and Loughran (2007)), and stock price level (Green and Hwang (2009)). Finally, in Table IA11, we reconsider shock-based comovement for headquarters relocations and stock splits.

A.1 Exces Comovement: Time Series

To explore the time series of residual return comovement, we repeat our analysis of Panel A in Table 4 of the main text separately for each year from 1980–2016. We plot median comovement estimates from 1,000 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 this analysis for groups containing 10, 20, 40, 80, or 160 randomly selected stocks.

From Figure 1, it is clear that residual return comovement is substantially greater than zero for all group sizes and during all time periods except 1984, 1986, and 1990. These years coincide with a period of significant negative returns to the SMB factor, which suggests that small stocks and large stocks performed very differently during this time period. Since panel regressions place equal weight on each observation, small-cap stocks, which constitute the vast majority of observations, are overweighted in the regression. The fact that small stocks underperformed their larger counterparts might have played a role in reducing the comovement estimates. Furthermore, the figure also illustrates that comovement estimates increase with group sizes for every time period.

A.2 Simulation Extensions

Our simulation results, presented in Table 1 of the main text, are based on the data-generating process in Equation 5 and the estimation of Equation 7. In this section, we explore variation in the parameters that we impose on the simulated data. First, in Table IA1, we report the confidence intervals from the results presented in Table 1 of the main text. These confidence intervals suggest that, despite a wide range of estimates, all magnitudes are substantive and highly statistically significant.

Next, we consider a special case of our DGP in which $Z_t \equiv 0$, and $E[F_t] = 0.65\%$. This setup is akin to the famous Roll critique (Roll (1977)), in which the market factor (F_t) is unobservable and measured with error. In these simulations, assets are assigned a random size at $t = 0$, which grows by $(1 + r_{it})$ each period. Thus, 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.

Table IA2 reports the simulation results of this exercise for 240 months of returns for 2,400 assets. 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 ($\sigma_F = 4.52\%$). Table IA2, Column 4 corresponds to Table1, Panel A, Column 1 in the main text. The model is simulated 1,000 times for each specification.

Comovement estimates are substantive in every specification. We obtain the smallest median comovement estimate of 0.0629 for groups containing 10 assets, when we impose that the variance of the market factor is only 1/8 the volatility of the average monthly value-weighted market return from 1980–2016. These estimates become substantially larger when the variance of the market factor grows. When the simulated variance of the market factor is of the same order of magnitude as the value-weighted CRSP returns (i.e., when $\sigma_F = 4.52\%$), the comovement estimates range from 0.6829–0.6957, suggesting that Roll's critique can have a significant effect on comovement estimates obtained from real data. Overall, this analysis highlights the fact that our assumption of an omitted factor in a multifactor 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.

Next, we consider the impact of an omitted factor in the case in which the market is perfectly observable. Specifically, we consider the special case of our DGP, in which $F_t \equiv 0$, and Z_t follows an AR(1) process. For each group size ($N_g = 10, 20, 40, 80, 160$), 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, where a higher ρ amounts to sorting more strongly on exposure to Z. 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.

We consider this setup first for $E[Z_t] = 0$ (i.e., Z is an unpriced factor), and we present the results in Table IA3. The median comovement estimates are substantive in all cases, and they increase in magnitude as the variance of the omitted factor increases. We repeat this analysis again for $E[Z_t] = 0.65\%$ (i.e., a priced factor), and we present the results in Table IA4. These results are nearly identical to the case in which Z is unpriced, confirming that an omitted factor has a significant effect on comovement estimates whether it is priced or not.

Finally, we simulate and estimate a "kitchen sink" model where we calibrate the simulation to match the real data as closely as possible and we include both an imperfect market proxy and an omitted factor Z. 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 market model regressions in the CRSP universe from 1980–2016. In Panel A, we set $Size_{i0}$ (i.e., the market capitalization at time 0) from the market capitalization of 2,400 randomly selected stocks from the CRSP universe in 1980. In subsequent panels, $Size_{i0}$

is simulated via an exponential distribution, a lognormal distribution, and a normal distribution. In all cases, $size_{it} = size_{i,t-1} \times (1 + r_{it})$. 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.

Table IA5 reports simulation results of Eq.(7) 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 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 multifactor 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. These findings are consistent with those of Pollet and Wilson (2010) who show that unobservable aggregate risk manifests through pairwise stock return correlations.

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 IA6 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.

A.3 Comovement Ubiquity: Robustness

.

In this section, we consider the robustness of our analysis presented in Table 4 of the main text. In Table IA7 we repeat our analysis shown in Table 4 in the main text for industry-adjusted returns. Specifically, we adjust returns according to the Fama–French 12-, 30- and 48- industry portfolios, as well as the TNIC 25 industries defined in Hoberg and Phillips (2016). For ease of comparison, Panel A presents results without industry adjustments. Each subsequent panel uses different industry classifications to adjust returns. Returns are adjusted before calculating the residuals in the respective empirical model used (e.g., CAPM in Column 2). The results indicate that industry adjustments do not fully attenuate comovement estimates, which remain substantive in virtually all specifications.

In Table IA8 we repeat our analysis shown in Table 4 using characteristic-adjusted returns following the procedure outlined in Daniel et al. (1997). Column 1 presents comovement estimates for market model adjusted returns for randomly grouped assets. Each subsequent column groups assets by market equity, book-to-market, momentum, asset growth, and operating profitability. For instance, in Row 1 (10 assets to each group) the first group would consist of the 10 stocks with the largest market cap, the second group would consist of the next 10 largest, and so on.

Stocks are re-sorted at the beginning of each year. The DGTW adjustments do substantially attenuate the estimates for the randomly sorted groups; however, the estimates remain significantly different than zero. The DTGW adjustments have a smaller attenuation effect for assets grouped by characteristics. These findings suggest that DGTW adjustments do not fully attenuate comovement estimates, even when assets are grouped by the same characteristics used in the adjustment.

A.4 Alternative Tests of Comovement: Robustness

We extend the settings of our pairwise stock return correlation analysis presented in Tables 5 and 6 of the main text. Specifically, we estimates pairwise stock return correlations on a measure of similarity in common mutual fund ownership (Fcap) from Anton and Polk (2014), common analyst coverage (analyst overlap) from Israelsen (2016), geographic distance from Barker and Loughran (2007), and stock price level from Green and Hwang (2009). Similarities in characteristics are calculated as $-|x_{i,t} - x_{j,t}|/\sigma(x_t)$, where $x_{i,t}$ and $x_{j,t}$ represent the characteristics for stocks i and j, and $\sigma(x_t)$ is the cross-sectional standard deviation of characteristic x at time t.

The results of this exercise are presented in Table IA9. In Panel A, we estimate univariate regressions of each characteristic. In Panel B, we estimate a multivariate regression for all similarities 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: Mkt, 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 that have available data from 1970 to 2016.

These findings corroborate those of Tables 5 and 6. In particular, adjusting returns for richer empirical models generally attenuates comovement estimates, but these estimates never reach zero. Further, the coefficient estimates for the multivariate regressions remain statistically significant, but they are smaller in virtually all cases when compared to the univariate regressions. To reiterate our conclusion from the main text, these results suggest that asset characteristics are correlated. Thus, it is unlikely that assets that are similar along observable criteria do not share similar exposure to latent factors. Overall, these findings suggest that the intensity-based test design is likely to suffer from a latent factor explanation.

Next, we extend the settings of our shock-based comovement analysis in Table 8 of the main text. Specifically, Table IA11 presents estimates for Carhart (1997) factor loadings before and after headquarters location changes and stock splits. For headquarters relocations, we follow the methodology of Pirinsky and Wang (2006), and for stock splits, we follow Green and Hwang (2009). For each event, factor loading estimates are obtained from factor regressions using daily return data for the two-year window centered around the event date. For headquarter location changes, the largest change in factor loadings corresponds to the SMB factor. This change is consistent with firms that change location becoming larger around the location change date. For stock splits, the loadings on many of the factors change around the split. On average, market exposure (beta) increases, high-minus-low (HML) exposure decreases, and up-minus-down (UMD) exposure decreases. The UMD exposure results are consistent with the findings of Chen et al. (2016), who contend that stock splits tend to follow episodes of high individual stock returns and hence increased exposure to UMD.

A.5 Firm Size and Tests of Comovement

In this section, we illustrate that the highly skewed distribution of asset sizes can lead to significant comovement estimation problems after controling for value-weighted market returns. In Section 3 we assume that all assets are positively exposed to an omitted factor which leads to positive correlation between any two asset returns, on average. Adjusting returns for market exposure effectively demeans exposure to non-market factors. However, if portfolios use equal-weighted returns they will exhibit non-zero exposure to (value-)demeaned factors. For instance, given the highly skewed distribution of firm sizes, a group of randomly selected stocks will contain mostly small stocks. As a result, these random groups will exhibit positive exposure to the size factor (SMB), on average.

We revisit the question of size and comovement by implementing a modified placebo procedure. Specifically, we change the random selection criteria to be proportional to firm size, such that the probability of picking a stock to belong to a group is equal to the fraction of that stock's market capitalization relative to overall market capitalization. This procedure ensures that large cap stocks are well represented in each sample instead of over weighting small stocks as an equal-weighted sampling scheme would. As a result, a randomly selected group of stocks will, in expectation, have the same return as the (value-weighted) market portfolio.

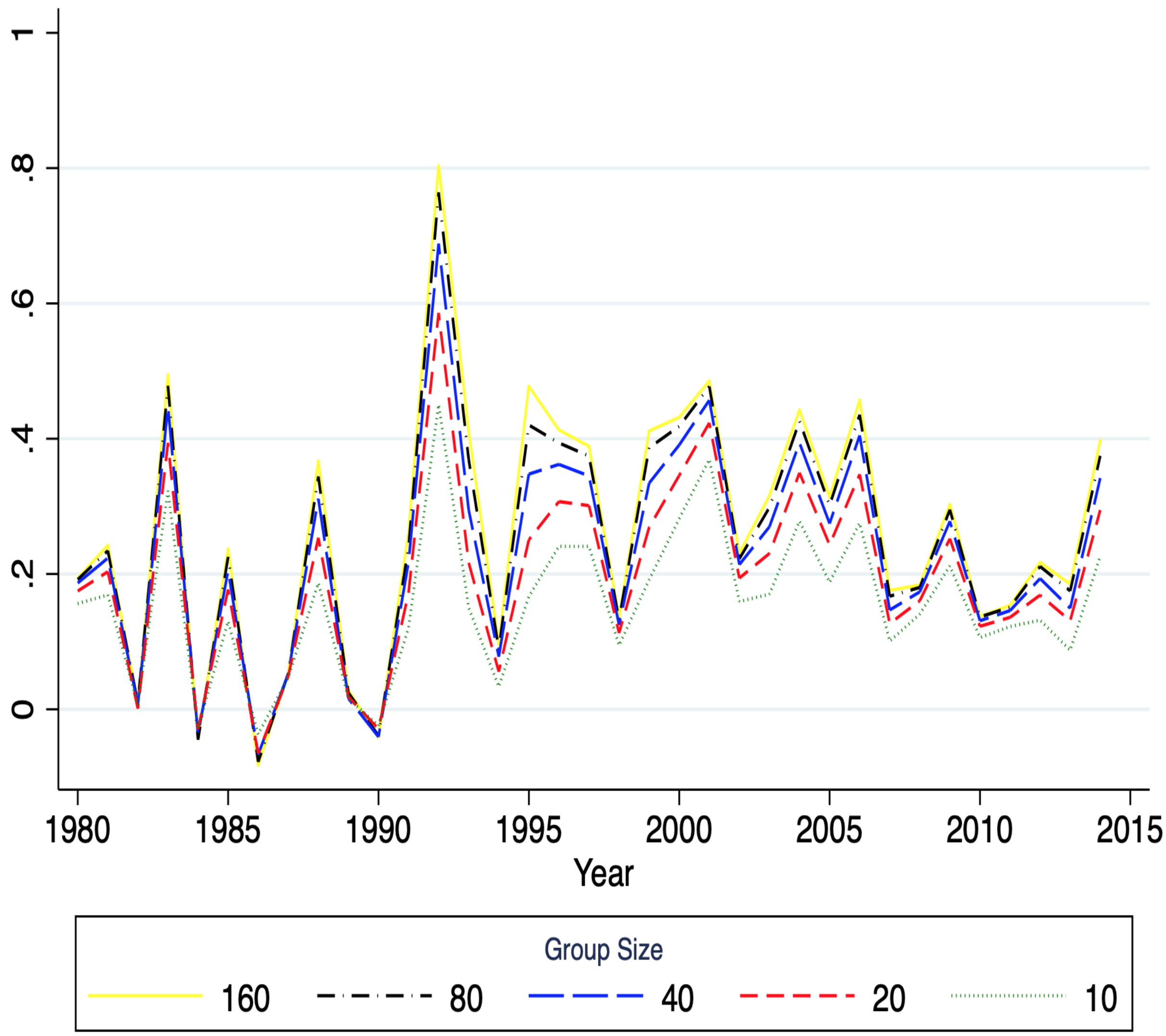
We implement this alternative placebo analysis by constructing a size-adjusted group of stocks from the CRSP universe between 1970 and 2016. At the beginning of each year, we select one group of 10, 20, 40, 80, or 160 stocks randomly in proportion to their beginning of year market equity. For each stock in the group, we calculate the peer return r_{-igt} as the average of the remaining selected stocks, and estimate the model:

$$r_{igt} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \theta(r_{-igt} - r_{ft}) + \epsilon_{it}$$

where r_{igt} and r_{mt} are stock returns and market returns, respectively. Table IA12 reports the confidence intervals from 1,000 repetitions of this procedure. One pattern that emerges from this analysis is the fact that for groupings of 80 or less stocks, the confidence intervals contain the value 0. For groups of 160 stocks, the placebo confidence intervals do not contain 0, but are significantly smaller than the corresponding comovement estimates in the main text.

As stated in the paper, these findings are consistent with the presence of a size factor. Small stocks with a positive exposure to this factor are over-represented in many samples and this common exposure will likely manifest as excess comovement. Generalizing this procedure is difficult in many of the settings considered in the literature since researchers cannot control the constituents in a peer group formed on objective criteria (e.g., firm headquarter location). However, this exercise demonstrates the need for a matching procedure like the one we employ in our volatility and Sharpe ratio tests to eliminate these spurious size effects.

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 IA1: Simulations: Confidence Intervals

The table shows estimate percentiles for simulations of the model described in Section 3. The simulation consists of simulating 240 months of returns for 2,400 assets following Eq. (1), where we impose $E(Z_t) = 0$. The volatility of Z_t (σ_Z) is set to 4.52%, in each panel. We form groups based on sorts of characteristic $X_i = \rho\gamma_i + (1-\rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. Thus, assets are grouped randomly when $\rho = 0$. Each panel corresponds to a different value of ρ, and each row corresponds to groups containing 10, 20, 40, 80, or 160 assets. We report percentiles of the coefficient θ in the regression: $r_{igt} = \theta r_{-igt} + \epsilon_{igt}$, where r_{-igt} is the average excess return of group g, excluding asset i at time t. For each specification, the model is simulated 1,000 times to extract the reported confidence intervals.

# Stocks	1%	5%	10%	50%	90%	95%	99%
				$\rho = 0$			
10	0.3460	0.3658	0.3720	0.3977	0.4257	0.4324	0.4666
20	0.5171	0.5313	0.5397	0.5729	0.6059	0.6117	0.6239
40	0.6613	0.6876	0.6938	0.7206	0.7444	0.7500	0.7599
80	0.8052	0.8172	0.8217	0.8415	0.8576	0.8637	0.8716
160	0.8930	0.8995	0.9016	0.9125	0.9216	0.9244	0.9268
				$\rho = .10$			
10	0.3430	0.3579	0.3718	0.3946	0.4204	0.4287	0.4432
20	0.5181	0.5285	0.5321	0.5650	0.5984	0.6096	0.6296
40	0.6690	0.6892	0.6955	0.7206	0.7424	0.7492	0.7665
80	0.8092	0.8205	0.8252	0.8387	0.8554	0.8596	0.8671
160	0.8929	0.8960	0.9028	0.9139	0.9226	0.9239	0.9256
				$\rho = .25$			
10	0.3297	0.3598	0.3663	0.4020	0.4305	0.4407	0.4611
20	0.4978	0.5278	0.5403	0.5759	0.6017	0.6113	0.6342
40	0.6688	0.6931	0.6987	0.7256	0.7473	0.7504	0.7619
80	0.8119	0.8219	0.8247	0.8453	0.8598	0.8647	0.8720
160	0.8981	0.9019	0.9038	0.9161	0.9238	0.9258	0.9321
				$\rho = .5$			
10	0.3620	0.3898	0.3962	0.4230	0.4532	0.4619	0.4725
20	0.5284	0.5613	0.5658	0.5934	0.6218	0.6275	0.6463
40	0.6851	0.7136	0.7204	0.7451	0.7674	0.7695	0.7774
80	0.8195	0.8311	0.8379	0.8545	0.8678	0.8704	0.8753
160	0.8947	0.9069	0.9107	0.9211	0.9286	0.9300	0.9358
				$\rho = .75$			
10	0.3967	0.4017	0.4102	0.4432	0.4812	0.4905	0.4985
20	0.5761	0.5789	0.5859	0.6146	0.6398	0.6500	0.6648
40	0.7132	0.7330	0.7365	0.7632	0.7821	0.7846	0.7874
80	0.8340	0.8400	0.8452	0.8627	0.8794	0.8832	0.8900
160	0.9117	0.9153	0.9165	0.9257	0.9366	0.9385	0.9427
				$\rho = .9$			
10	0.3709	0.4020	0.4130	0.4455	0.4784	0.4854	0.4997
20	0.5517	0.5772	0.5877	0.6179	0.6463	0.6509	0.6567
40	0.7200	0.7336	0.7404	0.7618	0.7883	0.7938	0.8080
80	0.8355	0.8483	0.8494	0.8635	0.8818	0.8842	0.8918
160	0.9090	0.9143	0.9172	0.9284	0.9376	0.9386	0.9451
				$\rho = 1$			
10	0.3888	0.4043	0.4150	0.4448	0.4789	0.4874	0.4989
20	0.5623	0.5793	0.5860	0.6150	0.6475	0.6527	0.6671
40	0.7237	0.7406	0.7445	0.7608	0.7880	0.7948	0.8086
80	0.8326	0.8391	0.8490	0.8650	0.8815	0.8831	0.8964
160	0.9114	0.9140	0.9165	0.9271	0.9345	0.9377	0.9403

Table IA2: 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. Each Panel corresponds to a different distribution imposed for the random size at $t = 0$. In Panel A, we use the market capitalization of 2,400 randomly select firms from the CRSP universe in 1980. Additionally, we simulate initial size according to an exponential distribution in Panel B, a lognormal distribution in Panel C, and a normal distribution in Panel D.

	σ_F = 1/8 × 4.52%	σ_F = 1/4 × 4.52%	σ_F = 1/2 × 4.52%	σ_F = 1 × 4.52%	σ_F = 2 × 4.52%	σ_F = 4 × 4.52%
Panel A: Empirical Distribution (CRSP Universe - 1980)						
10	0.0257	0.0767	0.2359	0.5134	0.7921	1.1127
20	0.0253	0.0804	0.2412	0.5168	0.7924	1.1178
40	0.0270	0.0789	0.2383	0.5217	0.7913	1.1048
80	0.0268	0.0806	0.2344	0.5165	0.7980	1.1180
160	0.0261	0.0823	0.2396	0.5170	0.7962	1.1147
Panel B: Exponential Distribution						
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
Panel C: Lognormal Distribution						
10	0.0023	0.0067	0.0267	0.0902	0.2678	0.5604
10	0.0023	0.0068	0.0250	0.0911	0.2657	0.5571
40	0.0022	0.0064	0.0251	0.0911	0.2605	0.5599
80	0.0024	0.0064	0.0254	0.0931	0.2673	0.5689
160	0.0023	0.0070	0.0252	0.0905	0.2674	0.5587
Panel D: Normal Distribution						
10	0.0008	0.0046	0.0176	0.0489	0.1621	0.1201
10	0.0008	0.0025	0.0103	0.0429	0.1175	0.0712
40	0.0008	0.0033	0.0104	0.0678	0.2251	0.0759
80	0.0009	0.0035	0.0162	0.0499	0.1539	0.0763
160	0.0005	0.0031	0.0144	0.0747	0.1326	0.0648

Table IA3: 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 IA4: 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 IA5: 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 θ in the regression $r_{igt} = \alpha + \beta r_{mt} + \theta r_{-igt} + \epsilon_{it}$, 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 $X_i = \rho\gamma_i + (1 - \rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. 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 IA6: 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 IA7: 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–French 12-, 30-, and 48- industry portfolios, and the TNIC 25 developed by 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 market model (Mkt), 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), 10 (PCA10), and 20 (PCA20) factors. The sample uses monthly returns for all CRSP/Compustat stocks that have available data from 1970 to 2016.

# Firms	Raw	Mkt	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 IA8: DGTW–Adjusted Returns and Comovement

This table presents comovement estimates of stock returns on returns of other stocks grouped by various 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. The rows of each panel correspond to different group sizes (10 to 160). The sample uses monthly returns for all CRSP/Compustat stocks that have 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 IA9: 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), 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 $x_{i,t}$ and $x_{j,t}$ represent the characteristics for stocks i and 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. 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: Mkt, 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 that have available data from 1970 to 2016.

	Mkt	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 IA10: Pairwise Correlations: Simulations

This table presents simulation results for pairwise stock return correlations for 500 assets for 240 periods. We report the percentile estimates of $\hat{\lambda}$ in the regression $Corr_{i,j} = -\lambda|x_i - x_j|/\sigma(x) + \epsilon_{i,j}$, where x_i and x_j represent characteristics of assets i and j; $\sigma(x)$ is the cross-sectional standard deviation of characteristic x; and $Corr_{i,j}$ is the correlation between the returns of i and j. The data generating process for asset returns follows equation 5 from Section 3.2 : $r_{it} - r_{ft} = \beta_i F_t + \gamma_i Z_t + \epsilon_{it}$. We simulate characteristics according to $X_i = \rho\gamma_i + (1-\rho)u_i$, $u_i \sim N(0, \sigma_\Gamma)$. Thus, higher ρ equates to a stronger relationship between the characteristic X and exposure to the latent factor Z. Each row corresponds to a different value of ρ. We present percentile estimates (1%, 5%, 10%, 50%, 90%, 95%, 99%) of $\hat{\lambda}$ from 1,000 simulations of each specification.

	Placebo Confidence Interval						
	1%	5%	10%	50%	90%	95%	99%
$\rho = 0.00$	-0.0032	-0.0022	-0.0017	0.0000	0.0019	0.0024	0.0033
$\rho = 0.10$	-0.0033	-0.0024	-0.0018	0.0003	0.0023	0.0027	0.0039
$\rho = 0.25$	-0.0032	-0.0020	-0.0013	0.0012	0.0039	0.0046	0.0060
$\rho = 0.50$	0.0003	0.0021	0.0031	0.0065	0.0106	0.0117	0.0144
$\rho = 0.75$	0.0039	0.0059	0.0068	0.0106	0.0146	0.0160	0.0190
$\rho = 0.90$	0.0044	0.0060	0.0069	0.0105	0.0138	0.0149	0.0168
$\rho = 1.00$	0.0039	0.0054	0.0063	0.0094	0.0127	0.0136	0.0156

Table IA11: Shock-based Tests and Factor Loadings

This table presents the factor loading changes before and after the shock (i.e., headquarter location changes and stock splits) as in Pirinsky and Wang (2006) and Green and Hwang (2009). The Fama–French factors augmented with momentum (see Carhart (1997)) are used to estimate factor loadings. Each factor loading is estimated using monthly returns during a symmetric 5-year window around the relocation or split date. Due to uncertainty around the exact relocation date, we ignore do not use the returns that correspond to the fiscal year of the financial report indicating a relocation. The first set of columns report the average loadings before and after headquarters relocations, as well as the change in loadings and a corresponding t-statistic. The second set of columns provide the same statistics for stock splits. We winsorize the factor loading estimates at the 1% and 99% levels to calculate mean values.

	Headquarter Relocations				Stock Price Splits			
	Pre-	Post-	Difference	t-stat	Pre-	Post-	Difference	t-stat
Mkt - Rf	0.902	0.896	-0.006	-0.520	0.985	1.136	0.151	13.920
SMB	0.709	0.706	-0.003	-0.160	0.812	0.566	-0.247	-12.550
HML	0.074	0.068	-0.006	-0.330	-0.266	-0.680	-0.413	-13.790
RMW	-0.127	-0.179	-0.052	-2.240	-0.302	-0.398	-0.096	-3.320
CMA	0.124	-0.008	-0.132	-4.670	0.442	0.362	-0.081	-2.420
UMD	-0.119	-0.183	-0.063	-5.990	-0.139	-0.570	-0.431	-23.980

Table IA12: Size-Adjusted Comovement: Confidence Intervals

This table presents confidence intervals for comovement estimates from randomly selected groups of stocks. At the beginning of each calendar year, we use the beginning of year market equity to select size-adjusted groups of 10, 20, 40, 80, and 160 stocks. The likelihood of being selected is proportional to the market equity of a given stock. For each stock in a group, we calculate the average of the remaining stocks (excluding the focal stock) as the peer return (r_{-igt}), and estimate the model $r_{igt} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \theta(r_{-igt} - r_{ft}) + \epsilon_{it}$. The process is repeated 1,000 times and the median estimate of θ as well as the empirical confidence intervals at the 1, 5, and 10% levels are reported. The sample uses monthly returns for all CRSP/Compustat stocks that have available data from 1970 to 2016.

	Placebo Confidence Interval						
# Stocks	1%	5%	10%	50%	90%	95%	99%
10	-0.233	-0.173	-0.139	-0.042	0.058	0.083	0.135
20	-0.262	-0.197	-0.161	-0.047	0.058	0.083	0.133
40	-0.238	-0.190	-0.155	-0.035	0.067	0.096	0.148
80	-0.171	-0.126	-0.084	0.025	0.117	0.141	0.195
160	0.046	0.095	0.123	0.210	0.290	0.312	0.359