News-Implied Linkages and Local Dependency in the Equity Market*

Shuyi Ge †, Shaoran Li ‡, Oliver Linton §

March 28, 2022

Abstract

This paper studies a heterogeneous coefficient spatial factor model that separately addresses both common factor risks (strong cross-sectional dependence) and local dependency (weak cross-sectional dependence) in equity returns. From the asset pricing perspective, we derive the theoretical implications of no asymptotic arbitrage for the heterogeneous spatial factor model, generalizing the work of Kou et al. (2018). We also provide the associated Wald tests for the APT restrictions in the general case when there are both traded and non-traded factors. On the empirical side, it is challenging to measure granular firm-to-firm connectivity for a high-dimensional panel of equity returns. We use extensive business news to construct firms' links through which local shocks transmit, and we use those news-implied linkages as a proxy for the connectivity among firms. Empirically, we document a considerable degree of local dependency among S&P500 stocks, and the spatial component does a great job in capturing the remaining correlations in the de-factored returns. We find that adding spatial interaction terms to factor models reduces mispricing and boosts model fitting. By comparing the performance of the model estimated using different networks, we show that the news-implied linkages provide a comprehensive and integrated proxy for firm-to-firm connectivity.

Keywords: Spatial asset pricing model; weak and strong cross-sectional dependence; local dependency; networks; big data; large heterogeneous panel

JEL Classification: C33; C58; G10, G12

*The authors would like to thank Hashem Pesaran, Cheng Hsiao, Alexei Onatskiy, Steven Kou, Donald Robertson, Lin Peng, Emmanuel Guerre, Liudas Giraitis, Daniele Bianchi, Robin Lumsdaine, Dick van Dijk, Koen Jochmans, Julius Vainora, Michael Leung, Sampad Mohanty, Weiguang Liu, Chong Shu, Ekaterina Gavrilova, Marco Valerio Geraci, Merrick Li, Seok Young Hong, Chen Wang, Yimeng Xie, Lidan Tan, Grigory Franguridi, Chris Yoo, and Haoxi Yang for their useful comments.

†Correspondence to the School of Finance, Nankai University. Email address: *sg751_shuyige@outlook.com*.

‡The author is at the School of Economics, Peking University. Email address *shaoran0405@outlook.com*.

§The author is at the Faculty of Economics, University of Cambridge. Email address *obl20@cam.ac.uk*.

1 Introduction

Comovement in equity returns is the combined effects of exposures to common risks and local interactions. Classical asset pricing models such as the classical capital asset pricing model (CAPM) developed by Sharpe (1964), and the arbitrage pricing theory (APT) by Ross (1976) focus only on the strong/pervasive component driven by a few common factors. Many studies have found those models focusing on strong dependence only are insufficient to capture all the significant interdependencies in asset returns. Local dependence still plays a non-negligible role (see for example Gabaix (2011), Acemoglu et al. (2012), Israelsen (2016), Barigozzi and Hallin (2017), Kou et al. (2018), Hale and Lopez (2019), Bailey et al. (2019a), and Barigozzi and Brownlees (2019)). To distinguish the two sources of dependencies, imagine a group of people sitting in a room on a chilly winter day. People might catch a cold because the heater is broken (common factors) or someone sitting close to them is ill (local interactions).

In this paper, we study a spatial factor model that separately addresses both common factor risks and the local dependence in equity returns. The factor component and spatial component complement each other, with the former capturing strong cross-sectional dependence in equity returns, and the latter capturing the weak cross-sectional dependence due to local interactions among entities. The network architecture of entities, like the sitting plan in the above example, is the key to studying local interactions. Unlike spatial interactions in geographical systems, where there is a natural network structure, there is no natural network structure for a high-dimensional panel of equity returns. We use extensive business news data to construct firms' linkages through which local shocks transmit. We use those news-implied linkages as a proxy for the connectivity among firms. It has been documented that common news coverage reveals information about linkages among companies, which are related to many economically important relationships like business alliances, partnerships, banking and financing, customer-supplier, and production similarity (Scherbina and Schlusche (2015), Schwenkler and Zheng (2019)). We use news data from RavenPack Equity files Dow Jones Edition from the beginning of 2004 to the end of 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which together produce the most actively monitored streams of news articles in the financial system. We identify linkages among firms by news co-mentioning.

Using the novel text-based network, we estimate the spatial factor model with different sets of common risk factors. We find a considerable degree of local dependence among S&P500 stocks. The spatial interaction term is highly significant after controlling for popular factors, and it continues to be significant even after adding industry-level factors. Different from most of the spatial econometrics literature, where spatial coefficients are assumed to be homogeneous, we adopt a heterogeneous coefficient framework from Bailey et al. (2016) and Aquaro et al. (2020). The model is flexible, allowing us to capture very general interaction patterns among a large number of firms. Using that framework, we are able to not only investigate the average effect among all or some subgroups of firms but also to gauge the individual-level effect. We document that apart from the average spatial effect measured by the mean group (MG) estimate being highly significant, at the individual level, the spatial effect via news-implied linkages is also highly significant after taking into account the multiple testing issue using the methodology of Barras et al. (2010). We find that the percentages of the individual contemporaneous spatial parameters that are statistically significant are over 70% across all specifications we consider (given a false discovery rate of 5%).[1] This high significance ratio implies that the link identification method based on news co-mentioning successfully detects economically important links. The framework also allows us to examine the heterogeneity at subgroup levels. By applying mean group estimation to different industry groups, we document heterogeneity at the industry level. In particular, financial companies have the highest degree of local dependence. We argue that the spatial factor model provides a unified way of addressing both strong and weak/local dependence in equity returns. To investigate how well the spatial factor model captures both types of dependence, and in particular how well the spatial component captures the remaining dependence in the de-factored returns, we examine the changes in the correlation structure before and after adding the spatial component to the traditional factor models. We find that adding the spatial component reduces the number of non-zero pair-wise cross-correlations by a huge margin. These results show that the spatial component constructed with news-implied linkages successfully captures the remaining cross-correlations from the de-factored returns.

We consider the asset pricing implications of the spatial factor model. Specifically, we extend some recent work of Kou et al. (2018) to allow for heterogeneous spatial effects and to allow for both traded and non-traded factors. We derive the empirically testable restrictions under a generalization of their framework. We consider the general case where there is a combination of both tradable and non-tradable factors and also the special case where there are only tradable factors. We provide Wald tests of these restrictions in both cases that exploit the spatial structure. In the case where there are non-tradable factors, the Wald statistic is a little non-standard and we use results of Andrews (1987) to justify the limiting chi-squared null distribution. We test the APT restrictions using a fairly large set of individual stocks. Our method is well suited to the moderately large cross-section case, since the spatial structure implies that the error covariance matrix of the reduced form model has only order n free parameters (where n is the cross-section dimension) rather than the order n^2 found in the unrestricted case and furthermore the inverse covariance matrix has a simple closed form expression. Unlike the conventional Wald test that does not exploit this structure, this test is going to work well for large n as long as T is sufficiently large. We perform numerical studies and they show that these tests perform well as long as T is large. Implementing the spatial factor models empirically, we demonstrate the benefits of addressing spatial/local interaction in terms of reducing mispricing errors and boosting model fittings.

Literature Review. This paper contributes to several strands of literature. The first one is cross-sectional dependence in equity returns. Cross-sectional dependence in a large panel is usually complex, and it reflects different types of inter-dependencies. Chudik et al. (2011), and Bailey et al. (2016) show that strong cross-sectional dependence (CSD) and weak cross-sectional dependence (CWD) have different economic implications and statistical behaviors. Thus they need to be accounted for separately. Kuersteiner and Prucha (2020) consider a short T panel with cross-sectional dependence due to both common factor risks and spatial/local interactions. Kou et al. (2018) consider a spatial asset pricing model with both factors and spatial interactions when the number of assets is small, the spatial effect is homogeneous, and errors are homoskedastic. Our spatial factor model is more general than theirs, as we allow large n, heterogeneous spatial effect, heteroscedastic errors, and spatial-temporal terms can be easily accommodated. Our work is related to the so-called "factor zoo" literature, (Harvey et al. (2016)), where many different traded and non-traded factors have been found, at least provisionally, to associate with the cross-section of stock returns. We allow both types of observed factors in

[1]This is the result when we use a conservative Bonferroni correction. Applying less conservative multiple testing procedures would give a higher significance ratio at the individual level.

addition to our spatial factors (there are essentially n spatial factors in each structural equation).

The paper also relates to the theory and testing of arbitrage pricing. While the literature focuses on the role of strong dependence (i.e., exposures to common risk factors), local dependence has received much less attention theoretically and empirically. Some recent literature considers testing asset pricing restrictions using individual stocks instead of portfolios. For example, Gagliardini et al. (2016) consider two-pass regressions of individual stock returns with time-varying risk premium, and they propose a Wald test. To consistently estimate the covariance matrix, some regularization techniques like hard thresholding (Bickel and Levina (2008)) are needed.

There is a large literature that uses statistical techniques to implicitly identify linkages from the return data itself. For example, Fan et al. (2011) suppose that the error covariance matrix is sparse (i.e., has lots of zeros, which represent the absence of linkages between firms beyond that contained in the common factors). They identify the location of non-zero entries by applying thresholding methods to the sample error covariance matrix. Our method uses economically relevant information gathered from other sources, specifically news stories, to identify the linkages. We explore a comprehensive news dataset that combines relevant content from multiple sources and identifies linkages among firms by news co-mentioning. With a measure of local connectivity, we can capture correlations that arise from both strong and the remaining weak dependence in a large panel using a single step. Without prior knowledge of local connectivity, Fan et al. (2011) need a two-step procedure (they first estimate a factor model, then use thresholding to estimate the error sample covariance matrix). There is a rising literature on machine learning and finance (see Gu et al. (2020), Giglio et al. (2021), Chen et al. (2020) among others). Especially, a lot of work has been done in quantifying the information embedded in unstructured data like text data (Garcia (2013), Scherbina and Schlusche (2015), Baker et al. (2016), Hoberg and Phillips (2016), Ke et al. (2019), Schwenkler and Zheng (2019), and Schwenkler and Zheng (2020), etc). Alternative data fill the gaps in data availability induced by limited disclosure and slow update, thus complementing traditional economic datasets. Among these aforementioned works, Scherbina and Schlusche (2015) and Schwenkler and Zheng (2019) also infer networks of firms from news. They find the information embedded in those news-implied linkages is highly predictive of stock returns and aggregate outcomes. Instead of using that information as conditional predictors, we build neighbor's contemporaneous returns into the asset pricing model, and we find news-implied interactions matter unconditionally in the cross-section of stock returns.

Our work also contributes to network effect or local risk spillover effect among economically linked firms. Local risks transmit through economic linkages, and firms with links exhibit excess co-movement beyond what is explained by factor models. There has been various proxies for firm to firm networks in the literature, including industry-based peers (Moskowitz and Grinblatt (1999), Fan et al. (2016), and Engelberg et al. (2018)), analyst co-coverage networks (Kaustia and Rantala (2013), Israelsen (2016), and Ali and Hirshleifer (2020)), customer-supplier networks (Cohen and Frazzini (2008)), geographic networks (Pirinsky and Wang (2006), Parsons et al. (2020)), data-driven partial correlation network (Barigozzi and Hallin (2017)), etc. We show that our news-based linkages provide a comprehensive and integrated proxy for firm-level connectivity. The spatial factor model estimated with news-implied network out-performs those aforementioned networks (except the return-based partial correlation network, which very much reflects the true connectivity structure in the sample) in terms of minimizing the mispricing errors. We also show that conditional on all those previously documented links, and our news-implied linkages are still important channels of local risk spillovers.

The rest of the paper is organized as follows. Section 2 describes the difference between the strong and weak cross-sectional dependence and we introduce the spatial factor model. Section 3 develops the asset pricing implications in the presence of local interactions. Section 4 presents the estimation and inference of the heterogeneous coefficient spatial model and provides tests of the APT restrictions. Section 5 presents the results of a simulation study investigating our asset pricing tests. Section 6 presents the empirical study. Section 7 concludes. Proofs, technical details, and supplementary figures and tables are in the Appendices.

Notations: If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequence of real numbers, then $f_n = \mathcal{O}(g_n)$ (the exact order of magnitude of f_n and g_n are the same) if there exist $n_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{n \geq n_0}(f_n/g_n) \geq C_0$, and $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$. For a $n \times n$ real matrix $A = (a_{ij})$, define its maximum column sum norm by $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} \mid a_{ij} \mid$, its maximum row sum norm by $\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} \mid a_{ij} \mid$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of matrix A, define the spectral radius of A by $\rho(A) = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$, and its spectral norm by $\|A\| = \sqrt{\rho(A^\intercal A)}$. $int(x)$ gives the integer part of x. We let $\mathbf{1}$ denote an $n \times 1$ unit vector.

2 Modelling Cross-Sectional Dependence by Spatial Factor Model

2.1 Strong Dependence: The Factor Model

Consider the factor model for returns at time t,

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t,$$

where: $\boldsymbol{r}_t$ is the $n \times 1$ vector of equity returns at t, $\mathbf{f}_t$ is the $K \times 1$ vector of common risk factors, and $\mathbf{B}$ is the $n \times K$ factor loadings, where β_{ik} is the loading of asset i on factor k. Let

$$\sum_{i=1}^{n} \mid \beta_{ik} \mid = \mathcal{O}(n^{\alpha_{\beta_k}}), \quad \text{for } k = 1, \ldots, K,$$

$$\|\mathbf{B}\|_1 = \max_{1 \leq k \leq K} \sum_{i=1}^{n} \mid \beta_{ik} \mid = \mathcal{O}(n^{\alpha_{\mathbf{B}}}),$$

where α_{β_k} measures how pervasive the k^{th} factor is, and $\alpha_{\mathbf{B}} = \max_k(\alpha_{\beta_k})$ measures how pervasive the factor component $\mathbf{B}\mathbf{f}_t$ is. For standard factor models, it is assumed that $\alpha_{\beta_k} > 0$ for $k = 1, \ldots, K$, and $\alpha_{\mathbf{B}} = 1$. Under these assumptions, each factor has a non-diminishing effect on the system, and exposures to common risk factors give rise to strong cross-sectional dependence, which is systematic and non-diversifiable.

2.2 Weak/Local Dependence: The Spatial Model

Consider the canonical spatial autoregressive model with a homogeneous spatial coefficient

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \psi W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t,$$

where W is the $n \times n$ adjacency matrix that specifies the channels from which shocks transmit, with a typical entry w_{ij} being the influence of the returns of j on that of i. The strength of spatial risk spillovers is represented by the scalar parameter ψ.

Spatial dependence usually characterizes weak cross-sectional dependence where interactions are local. To show this, we re-write the spatial autoregressive model as $\boldsymbol{r}_t = G(\psi)\boldsymbol{\nu}_t$, where $G(\psi) = (I_n - \psi W)^{-1}$, and $\boldsymbol{\nu}_t = \boldsymbol{\alpha} + \boldsymbol{\epsilon}_t$. The spatial autoregressive model can be seen as a factor model with n factors, where $G(\psi)$ is the

$n \times n$ matrix of factor loadings, where g_{ij} is the loading of i on factor j. All factors are weak and only have local effects if the following absolute summability condition is true: for c some positive finite constant

$$\sum_{i=1}^{n} \mid g_{ij} \mid \leq c,$$

for all $j = 1, \ldots, n$. The absolute summability condition is equivalent to a bounded column sum matrix norm condition on the Leontief inverse $G(\psi) = (I_n - \psi W)^{-1}$. As in LeSage (2008), $G(\psi) = (I_n - \psi W)^{-1} = I + \psi W + \psi^2 W^2 + ... = I + \sum_{j=1}^{\infty} \psi^j W^j$. The Leontief inverse takes accounts of direct interaction effect and higher-order indirect effects. The assumption that the column sum of $G(\psi) = (I_n - \psi W)^{-1}$ is uniformly bounded is usually assumed in spatial econometrics (see Kelejian and Prucha (1998), Kelejian and Prucha (1999), Lee (2004)) to limit the cross-sectional correlation to a manageable degree. Although some recent developments show that we may relax this assumption (Aquaro et al. (2020), Pesaran and Yang (2021)[2]), we take that assumption as a modelling assumption to distinguish between strong and weak/local dependence. In particular, for weak dependence, no cross-sectional unit exerts pervasive effects on the system, and the interactions are local. There will be more discussions in Section 2.3.

2.3 Strong and Weak Dependence: The Spatial Factor Model

We study a heterogeneous coefficient spatial factor model in which the factor component and the spatial component complement each other, with the former addressing strong dependence and the latter addressing spillovers that are non-pervasive/local in nature (i.e.,cross-sectional weak dependence (CWD) defined in Chudik et al. (2011)). Specifically, we suppose that

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\boldsymbol{\epsilon}_t$ is an error process specified below. Here, $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\psi}) = \text{diag}(\psi_1, \ldots, \psi_n)$ is a diagonal matrix with n individual-specific contemporaneous spatial coefficients on the main diagonal, while $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ is the $n \times K$ matrix of factor loadings, where $\boldsymbol{\beta}_k$ is the vector of loadings on the k^{th} common risk factor.

The spatial component has several main features. Firstly, the spillover coefficients are heterogeneous. One might reasonably suspect that the sensitivities to neighbors' risks are different from firm to firm. While the restrictive assumption that all entities share the same spatial coefficient is necessary for small T, it can be relaxed when T is big. Some recent work in the spatial literature (see LeSage and Chih (2018), Aquaro et al. (2020), and Chen et al. (2021)) consider heterogeneity in spatial parameters explicitly. We follow the framework from Aquaro et al. (2020). They show that a heterogeneous spatial autoregressive model like Equation 1 can be consistently estimated with large T. We utilize this nice feature to explore the heterogeneity in the strength of local dependency. Moreover, we could examine the heterogeneity pattern at subgroup levels (such as industry levels) using mean-group estimation, which is a popular tool in heterogeneous panel literature. Secondly, it is possible to add weakly exogenous spatial-temporal terms $\sum_{l=1}^{L} \boldsymbol{\Psi}_l W \boldsymbol{r}_{t-l}$ to Equation 1. $\boldsymbol{\Psi}_l W \boldsymbol{r}_{t-l}$ corresponds to the spatial-temporal term at the lth lag for $l = 1, \ldots, L$, where $\boldsymbol{\Psi}_l = diag(\boldsymbol{\psi}_l) = diag(\psi_{l,1}, \ldots, \psi_{l,n})$ is a diagonal matrix of spatial-temporal parameters at the lth lag. These dynamic terms may account for potential market microstructure effects, which are important for financial data with high frequency.

We impose the following assumptions on the heterogeneous coefficient spatial factor model (Equation 1).

[2]They explicitly consider the case where there are dominant units that generate pervasive effects.

Assumption 1. $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$, $E(\mathbf{f}_t) = \mathbf{0}$, $E(f_{kt}\boldsymbol{\epsilon}_t) = \mathbf{0}$ for $k = 1, \ldots, K$, $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\intercal) = \Omega = D_{\boldsymbol{\sigma}} = \text{diag}(\boldsymbol{\sigma}^2) = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, where $0 < \underline{\sigma}^2 \leq \sigma_i^2 \leq \bar{\sigma}^2 < \infty$.

Assumption 2. (a) *Let* $\sum_{i=1}^n \mid \beta_{ik} \mid = \mathcal{O}(n^{\alpha_{\beta_k}})$, for $k = 1, \ldots, K$, *and* $\|\mathbf{B}\|_1 = \max_{1\leq k\leq n} \sum_{i=1}^n \mid \beta_{ik} \mid = \mathcal{O}(n^{\alpha_{\mathbf{B}}})$. $\alpha_{\beta_k} > 0$ *for* $k = 1, \ldots, K$, *and* $\alpha_{\mathbf{B}} = 1$. (b) $\lambda_{min}(\mathbf{B}^\intercal\mathbf{B}) \to \infty$ *as* $n \to \infty$.

Assumption 3. (a) *The adjacency matrix* W *is known and time-invariant with its diagonal elements being zeros. (b)* W has bounded maximum row sum norm (i.e., $\|W\|_\infty < M < \infty$), and *for all* $i = 1, \ldots, n$

$$\mid \psi_i \mid < \frac{1}{\|W\|_\infty}.$$

Assumption 4. *Define* $G(\boldsymbol{\Psi}) = (I_n - \boldsymbol{\Psi}W)^{-1}$*, where* g_{ij} *is a typical entry of* G*. Then suppose that* $\|G(\boldsymbol{\Psi})\|_1 < c < \infty$ *for a positive constant* c*.*

Remark 1. Assumption A2 guarantees that each factor has a non-diminishing effect on the system, and the exposures to common risk factors give rise to strong cross-sectional dependence. We only assume that at least one factor is strong (i.e., with $\alpha = 1$), and all other factors are not weak. This is because, in practice, we may want to add many factors, which have different degrees of pervasiveness. For example, Bailey et al. (2020) find that for the factors proposed in the finance literature for asset pricing, only the market factor is strong over all the windows they consider. The credibility of this assumption depends on how many and what type of factors are being considered. There is a growing literature on weak factor models, see for example Bryzgalova (2015), where these conditions are weakened for a subset of the factors and the consequences traced through for estimation and testing.

Remark 2. Assumption A3 is to ensure that the Leontief inverse $G(\boldsymbol{\Psi}) = (I_n - \boldsymbol{\Psi}W)^{-1}$ exists. The invertibility of $(I_n - \boldsymbol{\Psi}W)$ requires the spectral radius $\rho(\boldsymbol{\Psi}W) < 1$, which holds under Assumption A3 given that $\rho(\boldsymbol{\Psi}W) \leq \|\boldsymbol{\Psi}W\|_\infty \leq \|\boldsymbol{\Psi}\|_\infty \|W\|_\infty = \max_{1\leq i\leq n} \mid \psi_i \mid \frac{1}{\|W\|_\infty} < 1$.

Remark 3. Assumption A4 assumes that the column sums of $G(\boldsymbol{\Psi}) = (I_n - \boldsymbol{\Psi}W)^{-1}$ is uniformly bounded in absolute values as n goes to infinity. This ensures that no cross-sectional unit exerts pervasive effects on the system and the interactions are local. Such assumption has been usually made in the spatial literature (see Kelejian and Prucha (1998), Lee and Yu (2010), Yu et al. (2008), etc). Assumption A4 also guarantees that the cross-sectional dependence in the errors of the reduced form factor model of Equation 1 is weak in the way defined in Chamberlain and Rothschild (1983). When $G(\boldsymbol{\Psi}) = (I_n - \boldsymbol{\Psi}W)^{-1}$ exists, notice that the spatial factor model (Equation 1) can be written as a reduced form factor model

$$\boldsymbol{r}_t = G\boldsymbol{\alpha} + G\mathbf{B}\mathbf{f}_t + G\boldsymbol{\epsilon}_t = \boldsymbol{\alpha}^* + \mathbf{B}^*\mathbf{f}_t + \boldsymbol{\epsilon}_t^*. \tag{2}$$

Under Assumptions A1, A3, and A4, the spatial interactions lead to weak cross-sectional dependence in the reduced form errors $\boldsymbol{\epsilon}_t^*$, namely the sequence of reduced form error covariance matrix $\{\Omega^*\}$ has uniformly bounded eigenvalues. To see this, we have

$$\rho(\Omega^*) = \|G\Omega G^\intercal\| \leq \bar{\sigma}^2 \|G\| \|G^\intercal\| \leq \bar{\sigma}^2 \|G\|_1^2 \leq \bar{\sigma}^2 c^2 < \infty.$$

We point out that Assumption A4 is not required for consistent estimation of the model (see Aquaro et al. (2020)). In principle, we could allow diverging maximum eigenvalues (for example, induced by a block dependence structure) as in the Assumption A3 from Gagliardini et al. (2016). We adopt Assumption A4 mainly for two reasons. Firstly, it is a sufficient condition for the reduced form factor model of Equation 1 to satisfy the complete factor representation condition from Ingersoll Jr (1984), under which the factor premiums can be uniquely identified. Secondly, Assumption A4 is a modelling assumption to distinguish the strong and weak cross-sectional dependence that arises from exposures to common risk factors and spatial/local interactions. From our perspective, pervasive dependence should be addressed by adding sufficient common factors into the model, and the correlation beyond factors should be weak.[3] In a similar vein, Fan et al. (2011) estimate the high-dimensional covariance matrix of asset returns by assuming that after taking out the influence of Fama-French three factors, the remaining cross-sectional dependence is weak in the way defined in Chamberlain and Rothschild (1983).

The spatial factor model provides a unified way of addressing the remaining dependence in the de-factored component. Fan et al. (2011) identify the location of significant correlations by applying thresholding methods to the factor model error sample covariance matrix. To capture both factor-driven strong dependence and remaining weak dependence in a large panel, they need a two-step procedure. Our method provides an alternative, which can be achieved in a single step. Compared with purely statistical methods, our method also has the advantage of being interpretable given that our linkages are constructed using information from business news.

3 Arbitrage Pricing Theory Under the Spatial Factor Model

In this section, we focus on the asset pricing implications of our heterogeneous coefficient spatial factor model. Kou et al. (2018) consider a special case of this model, which they call the Spatial APT model. In particular, they consider the case where n is small, $\psi_i = \psi$ (homogeneous spatial effects), and the errors are homoskedastic. Kou et al. (2018) derives the implications of the absence of arbitrage on the parameters of the model, in particular on the intercept vector $\boldsymbol{\alpha}$. We extend their analysis by deriving the implications of no-arbitrage under our framework.

Since the APT operates in a static framework, we drop the time subscript, and consider the one-period spatial factor model:

$$\boldsymbol{r} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f} + \boldsymbol{\Psi} W \boldsymbol{r} + \boldsymbol{\epsilon}. \tag{3}$$

We follow Ingersoll Jr (1984), and consider a fixed infinite economy where a sequence of nested subsets of assets are examined. For the n^{th} economy, a new asset is added to the $(n-1)^{th}$ economy indicated by

$$\boldsymbol{r}^{n\intercal} = (\boldsymbol{r}^{n-1\intercal}, r_n).$$

We denote the size of economy by superscript n. And a portfolio in the n^{th} economy is denoted as $\boldsymbol{c}^n \in \mathbb{R}^n$. $\mathbf{1}^n$ is the vector of n ones. We consider subsequences of assets, where subsequences are indexed by v. There are asymptotic arbitrage opportunities if there is a subsequence of portfolios that satisfy the following conditions:

$$\text{var}(\boldsymbol{c}^{v\intercal}\boldsymbol{r}^v) \to 0 \;\; as \quad v \to \infty,$$

[3] A similar view has been taken in Gagliardini et al. (2020). The authors believe that if we correctly specify the common factors, the errors should be weakly correlated.

$$E(\boldsymbol{c}^{v\intercal}\boldsymbol{r}^{v}) \geq \delta > 0 \text{ for all } v,$$

$$\boldsymbol{c}^{v\intercal}\mathbf{1}^{v} = 0 \text{ for all } v.$$

We say that two $n \times 1$ vectors are asymptotically equivalent, i.e., $\boldsymbol{\alpha}^n \approx \widetilde{\boldsymbol{\alpha}}^n$ whenever there is a positive finite number V such that the sum of squared deviations is uniformly bounded, i.e., for $\boldsymbol{v}^n = \boldsymbol{\alpha}^n - \widetilde{\boldsymbol{\alpha}}^n$ we have

$$(\boldsymbol{v}^n)^\intercal(\boldsymbol{v}^n) \leq V \text{ for all } n. \tag{4}$$

That is to say that the (pricing) error vector $\boldsymbol{v}^n$ does not grow with the number of assets.

Theorem 1. *Assume that the returns on an infinite set of assets are generated by the heterogeneous spatial factor model (3), and that Assumptions A1-A4 hold. If there are no arbitrage opportunities, then there is a sequence of K by 1 vector of factor premiums $\boldsymbol{\lambda}^n$ and a constant λ_0^n such that the following approximation holds*

$$\boldsymbol{\alpha}^n \approx (I_n - \boldsymbol{\Psi}^n W^n)\mathbf{1}^n\lambda_0^n + \mathbf{B}^n\boldsymbol{\lambda}^n. \tag{5}$$

Corollary 1. *Suppose that the factors are traded portfolios and that there exists a risk-free asset with rate r_f. Then, we can re-write the spatial factor model (Equation 3) in terms of excess returns:*

$$\tilde{\boldsymbol{r}} = \breve{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{z}} + \boldsymbol{\Psi}W\tilde{\boldsymbol{r}} + \boldsymbol{\epsilon}, \tag{6}$$

where $\breve{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I_n - \boldsymbol{\Psi}W)\mathbf{1}r_f + \mathbf{B}1r_f - \mathbf{B}E(\tilde{\boldsymbol{f}})$, for $\mathbf{1}$ being the $n \times 1$ unit vector and 1 being a vector of ones that is conformable. $\tilde{\boldsymbol{r}} = \boldsymbol{r} - \mathbf{1}r_f$ *is the vector of asset excess returns, and* $\tilde{\mathbf{z}}$ is the vector of factor excess returns. We have $\tilde{\mathbf{z}} = \tilde{\boldsymbol{f}} - r_f 1$, *where* $\tilde{\boldsymbol{f}}$ *is the factor portfolio real returns, and* $\boldsymbol{f} = \tilde{\boldsymbol{f}} - E(\tilde{\boldsymbol{f}})$. *In such a case, for an infinite economy, no asymptotic arbitrage implies that*

$$\breve{\boldsymbol{\alpha}}^n \approx \mathbf{0}. \tag{7}$$

Corollary 2. *For any $\delta > 0$, there is a constant n_δ such that the number of elements in $\breve{\boldsymbol{\alpha}}^n$ that are bigger than δ in absolute values is uniformly bounded by n_δ . That is,*

$$\lim_{n\to\infty}\sum_{j=1}^{n} I(\mid \breve{\alpha}_j^n \mid > \delta) < n_\delta < \infty. \tag{8}$$

Many different factors have been proposed and used in published articles, with some factors being traded portfolios while some factors are not (Harvey et al. (2016)). The asset pricing implications of traded factors are different from those of non-traded factors. Therefore, we consider the general case with a mixture of both these two types of factors. Suppose that out of the K common risk factors, the first k_1 are tradable factors, and the rest k_2 factors are not tradable. Let $\boldsymbol{f} = (\boldsymbol{f}_1^\intercal, \boldsymbol{f}_2^\intercal)^\intercal$, where $\boldsymbol{f}_1 = (f_1, \ldots, f_{k_1})^\intercal$, and $\boldsymbol{f}_2 = (f_{k_1+1}, \ldots, f_K)^\intercal$. $\tilde{\boldsymbol{f}} = (\tilde{\boldsymbol{f}}_1^\intercal, \tilde{\boldsymbol{f}}_2^\intercal)^\intercal$, where $\tilde{\boldsymbol{f}}_1 = (\tilde{f}_1, \ldots, \tilde{f}_{k_1})^\intercal$, and $\tilde{\boldsymbol{f}}_2 = (\tilde{f}_{k_1+1}, \ldots, \tilde{f}_K)^\intercal$. Here, $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$, where $\mathbf{B}_1$, $\mathbf{B}_2$ correspond to the first k_1 and last k_2 columns of $\mathbf{B}$, respectively. $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\intercal, \boldsymbol{\lambda}_2^\intercal)^\intercal$, where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ correspond to the risk premiums for the k_1 tradable factors and k_2 non-tradable factors, respectively.

Corollary 3. *Suppose there exists a risk-free asset with rate r_f, and there is a combination of both tradable and non-tradable factors as described above. Re-write the spatial factor model (3) as:*

$$\tilde{\boldsymbol{r}} = \underline{\boldsymbol{\alpha}} + \mathbf{B}_1\tilde{z}_1 + \mathbf{B}_2\boldsymbol{f}_2 + \boldsymbol{\Psi}W\tilde{\boldsymbol{r}} + \boldsymbol{\epsilon}, \tag{9}$$

where $\tilde{\boldsymbol{r}} = \boldsymbol{r} - \mathbf{1}r_f$ *is the vector of asset excess returns,* $\tilde{\boldsymbol{z}}_1 = \tilde{\boldsymbol{f}}_1 - r_f 1$ *is the vector of excess factor returns for the* k_1 *tradable factors, and* $\underline{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I_n - \boldsymbol{\Psi}W)\mathbf{1}r_f + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_1)$, for $\mathbf{1}$ being the $n \times 1$ unit vector and 1 being a vector of ones that is conformable. In such a case, no asymptotic arbitrage implies

$$\underline{\boldsymbol{\alpha}}^n \approx \mathbf{B}_2^n \boldsymbol{\lambda}_2^n. \tag{10}$$

Remark 4. Theorem 1 and the corollaries suggest some statistics that we can employ to compare the relative performance of different asset pricing models. In particular, the L_1, L_2 norms of the mispricing errors, and the number of components with big mispricing errors could be useful in measuring how well the approximation is.

Remark 5. Theorem 1 and the corollaries can be easily extended to spatial factor models with more than one spatial spillover channel, for example, the two-W model in Bailey et al. (2016).

Proofs of theorem 1 and the corollaries are in Appendix A.

4 Estimation and Statistical Inference

In this section, we discuss the estimation of the baseline heterogeneous coefficient spatial factor model. We suppose that a sample of returns and factors are observed $\{\boldsymbol{r}_t, \mathbf{f}_t, t = 1, \ldots, T\}$ along with the spatial interaction matrix W obtained from the text datasource. We suppose that the data are generated from

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\Psi}W\boldsymbol{r}_t + \boldsymbol{\epsilon}_t, \tag{11}$$

where $\boldsymbol{\epsilon}_t$ is an error process discussed below.

4.1 Quasi Maximum Likelihood

There are mainly two classes of methods that have been developed in the literature to estimate spatial models: the maximum likelihood method (Lee (2004), Lee and Yu (2010), Shi and Lee (2017), and Aquaro et al. (2020), among others), and the IV/GMM approach (Kelejian and Prucha (1998), Kelejian and Prucha (1999), Lee (2007), and Kuersteiner and Prucha (2020) among others). We adopt the QML procedure proposed in Bailey et al. (2016) and Aquaro et al. (2020). Collect all parameters from the model into the $(n(K+3)) \times 1$ vector $\boldsymbol{\theta}_{full} = (\boldsymbol{\psi}^\intercal, \boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_K^\intercal, \boldsymbol{\sigma}^{2\intercal})^\intercal = (\boldsymbol{\psi}^\intercal, \boldsymbol{b}^\intercal, \boldsymbol{\sigma}^{2\intercal})^\intercal$, where $\boldsymbol{b} = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_K^\intercal)^\intercal$ collects all the parameters associated with the exogenous variables (we could also add weakly exogenous terms). The log-likelihood function of Equation 11 is

$$\mathcal{L}_T(\boldsymbol{\theta}_{full}) = -\frac{nT}{2}\ln(2\pi) - \frac{T}{2}\sum_i^n \ln(\sigma_i^2) + \frac{T}{2}\ln|\boldsymbol{S}(\boldsymbol{\Psi})^\intercal \boldsymbol{S}(\boldsymbol{\Psi})| - \frac{1}{2}\sum_{t=1}^T [\boldsymbol{S}(\boldsymbol{\Psi})\boldsymbol{r}_t - \underline{\boldsymbol{B}}\boldsymbol{x}_t]^{\intercal -1}[\boldsymbol{S}(\boldsymbol{\Psi})\boldsymbol{r}_t - \underline{\boldsymbol{B}}\boldsymbol{x}_t], \tag{12}$$

where $\boldsymbol{S}(\boldsymbol{\Psi}) = I_n - \boldsymbol{\Psi}W$, and $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{nt})^\intercal$. We stack the constant and all the exogenous variables for i at t in $\boldsymbol{x}_{it} = (1, \boldsymbol{f}_t^\intercal)^\intercal$, and $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}^\intercal, \ldots, \boldsymbol{x}_{nt}^\intercal)^\intercal$ is the $(1+K)n$ by 1 vector. Here, $\underline{\boldsymbol{B}}$ is the n by $(1+K)n$ block diagonal matrix with elements $\boldsymbol{b}_i^\intercal = (\alpha_i, \beta_{1,i}, \ldots, \beta_{K,i})$, $i = 1, \ldots, n$, on the main diagonal and zeros elsewhere.

An immediate extension of Equation 11 is to incorporate weakly exogenous spatial-temporal terms. The model is

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\Psi}_0 W\boldsymbol{r}_t + \sum_{l=1}^L \boldsymbol{\Psi}_l W\boldsymbol{r}_{t-l} + \boldsymbol{\epsilon}_t, \tag{13}$$

where we denote the contemporaneous spatial coefficients using $\boldsymbol{\Psi}_0$, and L is the number of spatial-temporal terms to incorporate. These dynamic terms could account for potential market microstructure effects, which might be important for applications that use high-frequency data. Similar modifications have been used in Eugene (1992), see also Dimson (1979). In such a case, in the likelihood function (Equation 12), $\boldsymbol{x}_{it} = (1, \boldsymbol{f}_t^\intercal, \sum_{j=1}^n w_{ij} r_{jt-1}, \ldots, \sum_{j=1}^n w_{ij} r_{jt-L})^\intercal$ collects all weakly exogenous for i at t, and $\boldsymbol{x_t} = (\boldsymbol{x}_{1t}^\intercal, \ldots, \boldsymbol{x}_{nt}^\intercal)^\intercal$ is the $(1+K+L)n$ by 1 vector. Here, $\underline{\boldsymbol{B}}$ is the n by $(1+K+L)n$ block diagonal matrix with elements $\boldsymbol{b}_i^\intercal = (\alpha_i, \beta_{1,i}, \ldots, \beta_{K,i}, \psi_{1,i}, \ldots, \psi_{L,i})$, $i = 1, \ldots, n$, on the main diagonal and zeros elsewhere. One may also include nonlinearities in the factors by adding basis functions of them into the regression and likelihood function, although the justification of such a semiparametric enterprise is beyond the scope of this work.

The quasi maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{full,QMLE}$ maximizes Equation 12. The error terms need not be Gaussian, but when they are, $\hat{\boldsymbol{\theta}}_{full,QMLE}$ is the maximum likelihood estimator of $\boldsymbol{\theta}_{full}$. Note that conditional on $\boldsymbol{\Psi}$ (or $\boldsymbol{\Psi}_0$ if we are dealing with the spatial-temporal model (Equation 13)), the system is linear, so that we can concentrate out the parameters $\underline{\boldsymbol{B}}, \boldsymbol{\sigma}$ (by standard linear regression arguments) to reduce the dimensionality and hence the computational burden. Aquaro et al. (2020) establish the consistency and the asymptotic normality of $\hat{\boldsymbol{\theta}}_{full,QMLE}$ in the case where T is large and n is fixed, and consistency and the asymptotic normality for the mean group estimators in the case where both T and n are large but $\sqrt{n}/T \to 0$. We provide their primitive conditions for identification and inference in Appendix B. Standard errors can be based on the first and second derivatives of the likelihood with respect to the parameters as usual for quasi-likelihood methods.

In the empirical section, we estimate two variants (multi-period versions of Equation 6 and Equation 9) of the baseline specification, for which we just need to modify the response variable and the exogenous variables.

4.2 Spatial Correlation

In this section we investigate how well the spatial factor model captures the remaining dependence in the de-factored component or rather how we can quantify this. Specifically, we can examine the changes in the correlation structure before and after adding the spatial component to factor models. There has been a literature on testing cross-sectional dependence, see Pesaran et al. (2004), and Juodis and Reese (2021). Instead of using a joint test, we focus on individual tests in this paper. If the spatial component does a good job explaining the remaining local dependence, we should see a reduction in the number of pairs with non-zero pair-wise error cross-correlations after adding the spatial component. In our application, we estimate the number of non-zero pair-wise cross-correlations of residuals from (1) a factor model and (2) its spatial-augmented model. For n cross-sectional units, the problem corresponds to testing $n(n-1)/2$ null hypotheses simultaneously. We use a multiple testing procedure to control for the overall size of the tests. We discuss this in more detail below.

Under the factor model settings, this task is relatively easy. From Pesaran et al. (2004) and Bailey et al. (2019b), we can establish the asymptotic distribution for the error correlation coefficient under the null $\mathcal{H}_{0,ij}$: $\rho_{ij} = 0$ for a panel data regression $y_{it} = \alpha_i + \boldsymbol{\beta}_i^\intercal \boldsymbol{x}_{it} + \epsilon_{it}$, where $\text{var}(\boldsymbol{\epsilon}_t) = \Sigma = (\sigma_{ij})$ is an $n \times n$ symmetric, positive definite matrix. Denote the correlation coefficient of ϵ_{it} and ϵ_{jt} by ρ_{ij}. To estimate the correlation coefficient of errors, one needs to first obtain residuals $\hat{\epsilon}_{it}$ by $\hat{\epsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\boldsymbol{\beta}}_i^\intercal \boldsymbol{x}_{it}$, where $\hat{\alpha}_i$ and $\hat{\boldsymbol{\beta}}_i$ are the OLS estimates for the ith entity. The sample estimate of ρ_{ij} is given by

$$\hat{\rho}_{ij} = \frac{\hat{\boldsymbol{\epsilon}}_i^\intercal \hat{\boldsymbol{\epsilon}}_j / T}{(\hat{\boldsymbol{\epsilon}}_i^\intercal \hat{\boldsymbol{\epsilon}}_i / T)^{1/2} (\hat{\boldsymbol{\epsilon}}_j^\intercal \hat{\boldsymbol{\epsilon}}_j / T)^{1/2}}.$$

When the regressors $\boldsymbol{x}_{it}$ are strictly exogenous, under the null $\mathcal{H}_{0,ij} : \rho_{ij} = 0$,

$$\sqrt{T}\hat{\rho}_{ij} \Longrightarrow \mathcal{N}(0,1) \text{ as } T \to \infty. \tag{14}$$

To test the hypothesis $\mathcal{H}_{0,ij} : \rho_{ij} = 0$, for p being the chosen nominal size, we can use $\frac{1}{\sqrt{T}}\Phi^{-1}(p/2)$ and $\frac{1}{\sqrt{T}}\Phi^{-1}(1-p/2)$ as the critical values, where Φ^{-1} is the inverse cdf of standard normal. However, to test $\rho_{ij} = 0$ for all $i < j$ jointly, we need to take the multiple testing issue into account. From Bonferroni (1935), given that there are $n_{test} = n(n-1)/2$ such tests, if the family-wise error rate (FWER) is p, it is sufficient to set the nominal size for each individual test as $p_i = p/n_{test}$ for $i = 1, \ldots, n_{test}$, so that the critical values for each $\hat{\rho}_{ij}$ becomes $\frac{1}{\sqrt{T}}\Phi^{-1}(p/2n_{test})$ and $\frac{1}{\sqrt{T}}\Phi^{-1}(1-p/2n_{test})$.[4]

The nice theoretical result from Equation 14 is derived under the assumption that all regressors are strictly exogenous. However, for the spatial factor model, this is not the case. In Equation 11, the spatial autoregressive term $W\boldsymbol{r}_t$ is endogenous, which makes the result from Equation 14 fail. Given that, we conduct inference using the residual wild bootstrap as in Mammen (1993). Again, we need to correct for multiple testing issue here. The critical values for each $\hat{\rho}_{ij}$ are $F^{-1}(p/2n_{test})$ and $F^{-1}(1-p/2n_{test})$, where F is the empirical null distribution of $\hat{\rho}_{ij}^b$ for all $i < j$, $b = 1, \ldots, B$.

4.3 Tests of Asset Pricing Restrictions

In this subsection, we study the econometric testing of the asset pricing restrictions derived in Section 3.

4.3.1 Risk Free Asset and Traded Asset Factors

The first case we consider is when there exists a risk-free asset with rate r_f, and all factors are traded portfolios (Equation 6). For exposition, we drop the notations (tildes and check), and write the multi-period spatial factor model in terms of excess-returns as

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{z}_t + \boldsymbol{\Psi}W\boldsymbol{r}_t + \boldsymbol{\epsilon}_t, \tag{15}$$

where $\boldsymbol{r}_t$ are asset excess returns and $\mathbf{z}_t$ are the factor excess returns. From corollary 1, our testing problem is $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}$.

Notice that equation Equation 15 can be written as a reduced form factor model:

$$\boldsymbol{r}_t = \boldsymbol{\alpha}^* + \mathbf{B}^*\mathbf{z}_t + \boldsymbol{\epsilon}_t^*, \tag{16}$$

where $\boldsymbol{\alpha}^* = G\boldsymbol{\alpha}$, $\mathbf{B}^* = G\mathbf{B}, \boldsymbol{\epsilon}_t^* = G\boldsymbol{\epsilon}_t$, and $G = (I_n - \boldsymbol{\Psi}W)^{-1}$. In particular, the error covariance matrix has the form $\Sigma_{\epsilon^*} = G(\boldsymbol{\Psi})D_{\boldsymbol{\sigma}}G(\boldsymbol{\Psi})^\intercal$, where $D_{\boldsymbol{\sigma}} = \text{diag}(\boldsymbol{\sigma}^2)$. It follows that $\boldsymbol{\alpha} = \mathbf{0}$ if and only if $\boldsymbol{\alpha}^* = \mathbf{0}$ under our condition that G is finite and positive definite. Thus we test the hypothesis that $\boldsymbol{\alpha}^* = \mathbf{0}$. We consider the Wald statistic

$$W = T\frac{\hat{\boldsymbol{\alpha}}^{*\intercal}\hat{\Sigma}_{\epsilon^*}^{-1}\hat{\boldsymbol{\alpha}}^*}{1 + \hat{\boldsymbol{\mu}}_z^\intercal\hat{\Sigma}_z^{-1}\hat{\boldsymbol{\mu}}_z}, \tag{17}$$

where $\hat{\boldsymbol{\alpha}}^*$ is the multivariate least squares estimator of $\boldsymbol{\alpha}^*$, and

$$\hat{\boldsymbol{\alpha}}^* = \widehat{\boldsymbol{\mu}}_r - \widehat{\mathbf{B}}^*\hat{\boldsymbol{\mu}}_z, \quad \widehat{\mathbf{B}}^* = \widehat{\Sigma}_{rz}\hat{\Sigma}_z^{-1}$$

[4] There are more advanced methods of choosing threshold values, like Bailey et al. (2019b). However, the theory does not go through for testing error correlation of the spatial factor model. To have a fair comparison, we consider a simple Bonferroni type of correction for both models.

$$\hat{\boldsymbol{\mu}}_z = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t,\ \hat{\boldsymbol{\mu}}_r = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t,\ \hat{\Sigma}_{rz} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_r)(\mathbf{z}_t - \hat{\boldsymbol{\mu}}_z)^\intercal,\ \hat{\Sigma}_z = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{z}_t - \hat{\boldsymbol{\mu}}_z)(\mathbf{z}_t - \hat{\boldsymbol{\mu}}_z)^\intercal,$$

$$\hat{\Sigma}_{\epsilon^*}^{-1} = (I_n - \widehat{\boldsymbol{\Psi}}W)^\intercal D_{\hat{\boldsymbol{\sigma}}}^{-1}(I_n - \widehat{\boldsymbol{\Psi}}W),$$

where $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2)$ are estimated from the structural model. The nice feature here is that $\hat{\Sigma}_{\epsilon^*}^{-1}$ only involves the inversion of the diagonal matrix $D_{\hat{\boldsymbol{\sigma}}}$ and so given the structural parameter estimates $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\sigma}}^2$ no further regularization would be called for. One could also replace $\hat{\boldsymbol{\alpha}}^*$ by $\hat{\boldsymbol{\alpha}}^* = G(\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}}$ is estimated from the structural model.

Proposition 1. *Suppose there is a risk free rate r_f and all factors are tradable portfolios. The spatial factor model and its reduced form model are Equation 15 and Equation 16, respectively, and suppose that Assumptions A1-A4 hold. Suppose further that Assumptions A5-A7 stated in the Appendix B hold. Then, under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}$, the Wald statistic (Equation 17) satisfies for every n, as $T \to \infty$*

$$W \Longrightarrow \chi^2(n). \tag{18}$$

Remark 6. With the spatial interaction structure, the error covariance matrix Σ_{ϵ^*} only has $\mathcal{O}(n)$ free parameters rather than $\mathcal{O}(n^2)$ in the unrestricted case. Furthermore, $\hat{\Sigma}_{\epsilon^*}^{-1}$ is a quadratic function of $\widehat{\boldsymbol{\Psi}}$, which is a diagonal matrix whose elements are chosen from a compact set $\boldsymbol{\Theta}_\psi$ by the structural model estimation. Unlike the sample covariance matrix-based reduced form test, this tests works well for large n, as long as T is large. When n is large, one can consider the following standardised Wald statistic

$$SW = \frac{W - n}{\sqrt{2n}}.$$

Under the null hypothesis, $SW \Longrightarrow \mathcal{N}(0, 1)$ as $T \to \infty$ and then $n \to \infty$ (sequential asymptotics).

4.3.2 Risk Free Asset and Some Non-Traded Asset Factors

Next, we consider the testing of APT restrictions when some factors are traded portfolios and others are not (Equation 9). Similar to the previous case, we drop the notations (the tildes and lower bar) for exposition, and write the multi-period structural model in this case as

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}_1\boldsymbol{z}_{1,t} + \mathbf{B}_2\boldsymbol{f}_{2,t} + \boldsymbol{\Psi}W\boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{19}$$

The reduced form is

$$\boldsymbol{r}_t = \boldsymbol{\alpha}^* + \mathbf{B}_1^*\boldsymbol{z}_{1,t} + \mathbf{B}_2^*\boldsymbol{f}_{2,t} + \boldsymbol{\epsilon}_t^*, \tag{20}$$

where $\boldsymbol{\alpha}^* = G\boldsymbol{\alpha}, \mathbf{B}_1^* = G\mathbf{B}_1, \mathbf{B}_2^* = G\mathbf{B}_2, \boldsymbol{\epsilon}_t^* = G\boldsymbol{\epsilon}_t$, and $G = (I_n - \boldsymbol{\Psi}W)^{-1}$. From corollary 3, the asset pricing restriction implies that our null is $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2$ for some $\boldsymbol{\lambda}_2 \in \mathbb{R}^{k_2}$. Note that this is equivalent to $\mathcal{H}_0 : \boldsymbol{\alpha}^* = \mathbf{B}_2^*\boldsymbol{\lambda}_2$ for some $\boldsymbol{\lambda}_2 \in \mathbb{R}^{k_2}$. This restriction may be written in the form

$$\boldsymbol{g}(\boldsymbol{\theta}_s) = \mathbf{0}, \tag{21}$$

where $\boldsymbol{\theta}_s = (\boldsymbol{\alpha}^{*\intercal}, \mathrm{vec}(\mathbf{B}_2^*)^\intercal)^\intercal$ and $\boldsymbol{g}(\boldsymbol{\theta}_s) = M_{\mathbf{B}_2^*}\boldsymbol{\alpha}^*, M_{\mathbf{B}_2^*} = I_n - \mathbf{B}_2^*(\mathbf{B}_2^{*\intercal}\mathbf{B}_2^*)^{-1}\mathbf{B}_2^{*\intercal}$. That is, the restriction vector is nonlinear in a subset of the factor model parameters, and furthermore, since $M_{\mathbf{B}_2^*}$ is of rank $n - k_2$ (there are only $n - k_2$ restrictions), there is a redundancy in the $n \times 1$ vector $\boldsymbol{g}$, which will affect the limiting null distribution of $\boldsymbol{g}(\widehat{\boldsymbol{\theta}}_s)$. The next task is to find the $n \times (n(k_2+1))$ Jacobian matrix $\mathbf{J}(\boldsymbol{\theta}_s) = \partial \boldsymbol{g}(\boldsymbol{\theta}_s)/\partial \boldsymbol{\theta}_s^\intercal$. In particular, we show that this is

$$\mathbf{J}(\boldsymbol{\theta}_s) = (M_{\mathbf{B}_2^*}, -(\boldsymbol{\alpha}^{*\intercal}\mathbf{B}_2^*(\mathbf{B}_2^{*\intercal}\mathbf{B}_2^*)^{-1} \otimes M_{\mathbf{B}_2^*})) + (M_{\mathbf{B}_2^*}, -(\boldsymbol{\alpha}^{*\intercal}M_{\mathbf{B}_2^*} \otimes \mathbf{B}_2^*(\mathbf{B}_2^{*\intercal}\mathbf{B}_2^*)^{-1})\mathbb{K}_{n,k_2}),$$

where $\mathbb{K}_{n,k_2}$ is the commutator matrix of dimensions $nk_2 \times nk_2$ (see Magnus and Neudecker (2019)). More details can be found in Appendix C.

To construct the test statistics, notice that the reduced form model (Equation 20) could be written as a seemingly unrelated regression (SUR) $\boldsymbol{r}_t = \Theta^* \boldsymbol{x}_t + \boldsymbol{\epsilon}_t^*$, where $\Theta^* = (\boldsymbol{\alpha}^*, \mathbf{B}_1^*, \mathbf{B}_2^*)$ is the $n \times (K+1)$ parameter matrix, and $\boldsymbol{x}_t = (1, \boldsymbol{z}_{1,t}^\intercal, \boldsymbol{f}_{2,t}^\intercal)^\intercal$. Let X be the $T \times (K+1)$ matrix of covariate, and R be the $T \times n$ matrix of responses, the QML estimator of Θ^* is $\hat{\Theta}^* = R^\intercal X(X^\intercal X)^{-1}$. Denote $\boldsymbol{\theta}^* = \text{vec}(\Theta^*) = (\boldsymbol{\alpha}^{*\intercal}, \text{vec}(\mathbf{B}_1^*)^\intercal, \text{vec}(\mathbf{B}_2^*))^\intercal$ and $\hat{\boldsymbol{\theta}}^* = \text{vec}(\widehat{\Theta}^*)$. Then we have as $T \to \infty$:

$$\sqrt{T}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) \Longrightarrow \mathcal{N}(\mathbf{0}, V_{\theta^*}), \tag{22}$$

$$V_{\theta^*} = A^{-1} \otimes \Sigma_{\epsilon^*}, \quad A = \lim_{T\to\infty} E(\frac{X^\intercal X}{T}), \tag{23}$$

$$\Sigma_{\epsilon^*} = G(\boldsymbol{\Psi})D_{\boldsymbol{\sigma}}G(\boldsymbol{\Psi})^\intercal, \quad D_{\boldsymbol{\sigma}} = \text{diag}(\boldsymbol{\sigma}^2). \tag{24}$$

Letting $\hat{\boldsymbol{\theta}}_s = (\hat{\boldsymbol{\alpha}}^{*\intercal}, \text{vec}(\widehat{\mathbf{B}}_2^*)^\intercal)^\intercal$, we have as $T \to \infty$

$$\sqrt{T}\boldsymbol{g}(\hat{\boldsymbol{\theta}}_s) \Longrightarrow \mathcal{N}(\mathbf{0}, \Omega_g) \text{ under the null}, \tag{25}$$

where $\Omega_g = \mathbf{J}(\boldsymbol{\theta}_s)V_{\theta_s}\mathbf{J}(\boldsymbol{\theta}_s)^\intercal$ and V_{θ_s} is the submatrix of V_{θ^*} that corresponds to the $\boldsymbol{\theta}_s$ components. The matrix Ω_g is singular because even though V_{θ_s} is nonsingular, the matrix $\mathbf{J}(\boldsymbol{\theta}_s)$ is of deficient rank as discussed above.

We next discuss how to estimate Ω_g consistently. We may firstly estimate the Jacobian matrix $\mathbf{J}(\boldsymbol{\theta}_s)$ by

$$\widehat{\mathbf{J}} = (M_{\widehat{\mathbf{B}}_2^*}, -(\hat{\boldsymbol{\alpha}}^{*\intercal}\widehat{\mathbf{B}}_2^*(\widehat{\mathbf{B}}_2^{*\intercal}\widehat{\mathbf{B}}_2^*)^{-1} \otimes M_{\widehat{\mathbf{B}}_2^*})) + (M_{\widehat{\mathbf{B}}_2^*}, -(\hat{\boldsymbol{\alpha}}^{*\intercal}M_{\widehat{\mathbf{B}}_2^*} \otimes \widehat{\mathbf{B}}_2^*(\widehat{\mathbf{B}}_2^{*\intercal}\widehat{\mathbf{B}}_2^*)^{-1})\mathbb{K}_{n,k_2}),$$

where $M_{\widehat{\mathbf{B}}_2^*} = I_n - \widehat{\mathbf{B}}_2^*(\widehat{\mathbf{B}}_2^{*\intercal}\widehat{\mathbf{B}}_2^*)^{-1}\widehat{\mathbf{B}}_2^{*\intercal}$. We then estimate V_{θ^*} by $\hat{V}_{\theta^*} = (X^\intercal X/T)^{-1} \otimes \hat{\Sigma}_{\epsilon^*}$, and $\hat{\Sigma}_{\epsilon^*} = G(\hat{\boldsymbol{\Psi}})D_{\hat{\boldsymbol{\sigma}}}G(\hat{\boldsymbol{\Psi}})^\intercal$, where $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2)$ are estimated from the structural model. With $\hat{V}_{\theta^*}$, we can easily get $\hat{V}_{\theta_s}$ by only keeping the submatrix of $\hat{V}_{\theta^*}$ that corresponds to the $\boldsymbol{\theta}_s$ component. Then, Ω_g could be consistently estimated by

$$\hat{\Omega}_g = \widehat{\mathbf{J}}\hat{V}_{\theta_s}\widehat{\mathbf{J}}^\intercal.$$

However, since Ω_g itself is not invertible, we must define a suitable generalized inverse of $\hat{\Omega}_g$ to construct the Wald statistic. The key issue identified in Andrews (1987) is to make sure that the estimator of Ω_g has the same rank as Ω_g itself, i.e., has rank $n - k_2$. In that case, the estimator of Ω_g^+ will be consistent. We proceed as follows. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n \geq 0$ be the ordered eigenvalues of the positive semidefinite matrix $\hat{\Omega}_g$, and let $\hat{Q}$ be an orthonormal matrix such that $\hat{\Omega}_g = \hat{Q}\hat{\Lambda}\hat{Q}^\intercal$, where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$. We can construct a suitable reduced rank estimator of $\hat{\Omega}_g$ by using the principal components associated with the largest $n - k_2$ eigenvalues of the estimated covariance matrix under a guarantee that they are positive. Specifically, let

$$\widetilde{\lambda}_j = \hat{\lambda}_j 1(\hat{\lambda}_j > 0) + \delta_T 1(\hat{\lambda}_j = 0), \ j = 1, \ldots, n - k_2,$$

where $\delta_T + T^{-1/2}\delta_T^{-1} \to 0$ is a regularization sequence.[5] Let $\widetilde{\Lambda}^+ = \text{diag}(\widetilde{\lambda}_1^{-1}, \ldots, \widetilde{\lambda}_{n-k_2}^{-1}, 0, \ldots, 0)$ and

$$\hat{\Omega}_g^+ = \hat{Q}\widetilde{\Lambda}^+\hat{Q}^\intercal. \tag{26}$$

[5]In practice we have found regularization to be unnecessary, i.e., the largest $n - k_2$ empirical eigenvalues in our application are all strictly positive.

This satisfies the Andrews (1987) conditions. Then define the Wald statistic

$$W = T\boldsymbol{g}(\hat{\boldsymbol{\theta}}_s)^\intercal \hat{\Omega}_g^{+} \boldsymbol{g}(\hat{\boldsymbol{\theta}}_s). \tag{27}$$

Proposition 2. *Suppose that there is a risk free rate r_f. Out of K common risk factors, the first k_1 factors are tradable and the rest k_2 factors are not. The spatial factor model and its reduced form model are Equation 19 and Equation 20, respectively, and suppose that Assumptions A1-A4 hold. Suppose further that Assumptions A5-A7 stated in the Appendix hold. Then, under the null $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2$ for some $\boldsymbol{\lambda}_2 \in \mathbb{R}^{k_2}$, for $T \to \infty$, the Wald statistic (Equation 27)*

$$W \Longrightarrow \chi^2(n - k_2). \tag{28}$$

Remark 7. The test could also be conducted using structural model (Equation 19), and we test the null $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2$ for some $\boldsymbol{\lambda}_2 \in \mathbb{R}^{k_2}$. Aquaro et al. (2020) has established the asymptotic normality of the QMLE of the structural parameters, and they provided the consistent estimation of the asymptotic variance-covariance matrix of the QML estimator. Using their theoretical results (details could be found in Appendix B), we can easily construct the test using the structural spatial factor model. In the case without a risk free rate, the reduced form test of the restrictions $\boldsymbol{\alpha}^* = \mathbf{1}\lambda_0 + \mathbf{B}_2^*\boldsymbol{\lambda}_2$ can be treated similarly.

Remark 8. Unlike in the case with only traded factors, the estimator $\hat{\Omega}_g^{+}$ is derived from $\widehat{\mathbf{J}}\hat{V}_{\theta_s}\widehat{\mathbf{J}}^\intercal$, which does not exploit the invertibility of $\hat{V}_{\theta^*}$. In fact, it is possible to derive an alternative generalized inverse based on the singular value decomposition of $\widehat{\mathbf{J}}$ and[6]

$$\hat{V}_{\theta^*}^{-1} = \frac{X^\intercal X}{T} \otimes \widehat{\Sigma}_{\epsilon^*}^{-1} = \frac{X^\intercal X}{T} \otimes (I_n - \widehat{\boldsymbol{\Psi}}W)^\intercal D_{\hat{\boldsymbol{\sigma}}}^{-1}(I_n - \widehat{\boldsymbol{\Psi}}W).$$

See details of how to derive the generalized inverse using this second method in Appendix C. In simulation studies we have found the two methods perform similarly and in particular Equation 26 is computationally convenient.

Proofs of proposition 1 and proposition 2 can be found in Appendix C.

5 Simulation Study

We examine the performance of those tests by Monte Carlo experiments. We consider a five-factor model:

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i^\intercal \boldsymbol{f}_t + \psi_i \sum_{j=1}^{N} w_{ij} r_{jt} + \epsilon_{it}.$$

As in Fan et al. (2015), we simulate $\{\boldsymbol{\beta}_i\}_{i=1}^n$, $\{\boldsymbol{f}_t\}_{t=1}^T$, and $\{\epsilon_{it}\}_{i,t=1}^{n,T}$ independently from $\mathcal{N}_5(\boldsymbol{\mu}_b, \Sigma_b)$, $\mathcal{N}_5(\boldsymbol{\mu}_f, \Sigma_f)$, and $\mathcal{N}(0, \sigma^2)$, respectively. The parameters $\boldsymbol{\mu}_b, \Sigma_b, \boldsymbol{\mu}_f, \Sigma_f, \sigma^2$ are calibrated using Fama-French five factors and $S\&P$ 500 stocks from January 2004 to December 2015. For the spatial component, we simulate $\psi_i \sim \mathcal{N}(0.4, 0.1)$ independently for $i = 1, \ldots, n$. And to construct the adjacency matrix W, we simulate a random network where each pair of nodes has a linking probability $p = 0.1$. The Leontief inverse matrix is then calculated as $G = (I_n - \boldsymbol{\Psi}W)^{-1}$, where $\boldsymbol{\Psi} = diag(\psi_1, \ldots, \psi_n)$.

[6] We are grateful to Alexei Onatskiy for pointing this out.

When all factors are traded, the null is $\boldsymbol{\alpha} = \mathbf{0}$. To evaluate the size of the test, $\alpha_i = 0$ for $i = 1, \ldots, n$. To evaluate the power of the test, we generate $\alpha_i \sim \mathcal{N}(0,1)$ independently for $i = 1, \ldots, int(\frac{1}{20}n)$, where $int(x)$ gives the integer part of x. This is a sparse alternative, with 5% of the individual α_i different from zero. After obtaining $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\intercal, \mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_5), \boldsymbol{f}_t$, $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{nt})^\intercal$, and G, $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{nt})^\intercal$ can be calculated as $\boldsymbol{r}_t = G\boldsymbol{\alpha} + G\mathbf{B}\boldsymbol{f}_t + G\boldsymbol{\epsilon}_t$, for $t = 1, \ldots, T$.

In the case where there is a combination of tradable and non-tradable factors, we assume that the last two factors are not tradable with risk premiums $\boldsymbol{\lambda}_2 = (1,1)^\intercal$. To evaluate the size of the test, we generate the null using $\boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2$, where $\mathbf{B}_2 = (\boldsymbol{\beta}_4, \boldsymbol{\beta}_5)$. And to evaluate the power of the test, we generate $v_i \sim \mathcal{N}(0,1)$ independently for $i = 1, \ldots, int(\frac{1}{20}n)$, and calculate $\boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2 + \boldsymbol{v}$. Similar to the previous case, we calculate $\boldsymbol{r}_t$ by $\boldsymbol{r}_t = G\boldsymbol{\alpha} + G\mathbf{B}\boldsymbol{f}_t + G\boldsymbol{\epsilon}_t$ for $t = 1, \ldots, T$.

Table 1 and Table 2 present the empirical size and power of each testing method. When T is not large enough, the tests exhibit excessive size distortions and over-reject. This is expected as the consistent estimation of the heterogeneous spatial model requires large T. For example, when $T = 100$, the size of both tests is always higher than 10% for all different values of n. And the size distortions are bigger for larger n due to the accumulation of estimation errors. When we increase T to 500, the size distortions drop substantially. And when T is sufficiently large, the empirical size of both tests is very close to the 5% nominal level for all n values. When it comes to the power of the tests, both tests have very high empirical power for all (n, T) combinations that we consider. For the (n, T) combination that is close to our empirical application, both tests exhibit the correct size, and have power of one.

	Size (%)				Power (%)			
(n/T)	100	500	1000	3000	100	500	1000	3000
50	16.10%	6.00%	6.50%	6.20%	97.80%	99.80%	99.88%	100.00%
200	31.70%	8.45%	5.95%	6.00%	100.00%	100.00%	100.00%	100.00%
500	52.20%	11.55%	6.00%	5.80%	100.00%	100.00%	100.00%	100.00%

Table 1: Size and Power of tests for $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}$ when all factors are tradable portfolios. Tests are conducted at 5% significance level and results are based on 2000 replications.

	Size (%)				Power (%)			
(n/T)	100	500	1000	3000	100	500	1000	3000
50	14.50%	6.00%	7.00%	6.80%	80.60%	96.80%	99.88%	100.00%
200	26.55%	8.40%	6.70%	6.00%	99.40%	100.00%	100.00%	100.00%
500	46.60%	9.40%	6.00%	5.80%	100.00%	100.00%	100.00%	100.00%

Table 2: Size and Power of tests for $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{B}_2\boldsymbol{\lambda}_2$ when the the last two factors are not tradable portfolios. Tests are conducted at 5% significance level and results are based on 2000 replications.

6 Empirical Study

We use daily returns of $S\&P$ 500 stocks for our application. All the stock market related data are from the Center for Research in Security Prices (CRSP). Daily factor returns and industry classification information are obtained from Kenneth French's website. Accounting data are from the merged CRSP/Compustat database. Data used to construct alternative networks are described in detail in Section 6.1.3.

The news data are obtained from RavenPack Equity files Dow Jones Edition from January 2004 to December 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which produce the most actively monitored streams of news articles in the financial system. Each unique news story (identified by unique story ID) tags the companies mentioned in the news by their unique and permanent entity identifier codes (RP_ENTITY_ID), by which we link to stock identifier TICKER and PERMNO.

Inspired by Scherbina and Schlusche (2015) and Schwenkler and Zheng (2019), we identify links by news co-mentioning. To be more specific, if a piece of business news reports two and only two companies together, then the two firms share a link. Although news that mentions more than two companies together may carry potential information about links, they provide noisier information. We also remove news with topics including analyst recommendations, rating changes, and index movements as these types of news might stack multiple companies together when they actually do not imply any real links. Table 14 provides descriptive statistics for RavenPack Equity files Dow Jones Edition dataset during the sample period. Since our comprehensive news dataset combines several sources, given a similar length of sample period, the number of unique news stories is more than ten times larger than that from Scherbina and Schlusche (2015) and more than eight hundred times than that from Schwenkler and Zheng (2019). For link identification purposes, we only use sample news that (1) is not about the topics mentioned above, (2) tags $S\&P$ 500 companies, and (3) mentions exactly two companies, which gives a subsample of $1,637,256$ unique news stories.

From all the links identified using this methodology, some links are transitory while some are more long-lasting. To gauge the persistency of links, we split full sample news data into 12 yearly link identification windows. Table 15 is the frequency distribution table of the number of yearly link identification windows that a pair gets identified as economic neighbors for all possible pairs (i, j) in our sample. 72.80% of the pairs never get co-mentioned during the sample period. For all the linked pairs (i, j) identified throughout the sample period, 49.6% of them are only mentioned in one yearly window. We consider those pairs as temporarily linked. They could get co-mentioned multiple times within a yearly window. But out of that one-year window, they are never mentioned together. To further reduce noise, we say a pair (i, j) has persistent economic relationships if they are identified in more than a certain number $(1 \leq m \leq 11)$ of yearly identification windows. For the construction of full sample adjacency matrix W, we set w_{ij} to the number of times i and j are co-mentioned throughout the sample if the pair (i, j) gets co-mentioned in more than m yearly identification windows (i.e., their link is persistent), and to zero otherwise.

Table 16 presents the number of identified pairs aggregated at industry level for threshold $m = 1$. Results for higher threshold values are shown in Table 17, and Table 18. We classify stocks into Fama-French 12 industries based on their Standard Industrial Classification (SIC) code. Compared with companies from other industries, financial companies, hi-tech companies, and manufacturing companies are more connected. Another important feature is that there are a lot of intra-industry links. Except for some industries with very few stocks like Durables and Telecommunication, whose statistics should be interpreted with care, other industries all have a high percentage of intra-industry links. Comparing tables of adjacency matrices with different threshold values m, we can tell that although higher threshold values reduce the absolute number of identified pairs, the relative industry level network remains quite stable.

6.1 Results

6.1.1 Main Results

For full sample estimation, we keep $S\&P$ 500 stocks that have no missing observations from 2004 to 2015, which leaves us $n = 394$ stocks. Adjacency matrix W contains all the persistent links (for different thresholds m) identified throughout the sample. As a convention in spatial econometrics, we apply row-normalization to the raw adjacency matrix so that $\sum_j^n w_{ij} = 1$ for all $i = 1, \ldots, n$. We investigate several models under the framework where there are only tradable factors, and asset returns and factor returns are in excess of the risk-free rate (Equation 15).

- Model 1: Spatial CAPM model

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{29}$$

- Model 2: Spatial factor model with Fama-French three factors

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\beta_2} z_{SMB,t} + \boldsymbol{\beta_3} z_{HML,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{30}$$

- Model 3: Spatial factor model with Fama-French five factors

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\beta_2} z_{SMB,t} + \boldsymbol{\beta_3} z_{HML,t} + \boldsymbol{\beta_4} z_{RMW,t} + \boldsymbol{\beta_5} z_{CMA,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{31}$$

- Model 4: Spatial factor model with Fama-French five plus Momentum factor

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\beta_2} z_{SMB,t} + \boldsymbol{\beta_3} z_{HML,t} + \boldsymbol{\beta_4} z_{RMW,t} + \boldsymbol{\beta_5} z_{CMA,t} + \boldsymbol{\beta_6} z_{MOM,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{32}$$

- Model 5: Spatial factor model with Fama-French five plus Momentum factor and Media Attention factor[7]

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\beta_2} z_{SMB,t} + \boldsymbol{\beta_3} z_{HML,t} + \boldsymbol{\beta_4} z_{RMW,t} + \boldsymbol{\beta_5} z_{CMA,t} + \boldsymbol{\beta_6} z_{MOM,t} + \boldsymbol{\beta_7} z_{MA,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{33}$$

For each model, the $(n(K+3))$ parameters are estimated using quasi maximum likelihood (QML). Given the huge amount of parameters in the model, here we only report some important summary statistics of the estimates in Table 3.[8] For a heterogeneous coefficient panel model, what is often of interest to empirical researchers is the mean of the individual-specific parameters across all entities (or entities within some sub-groups). If we assume that individual-specific coefficients are randomly distributed around their common means as follows:

$$\beta_{k,i} = \beta_k + \zeta_{k,i}, \psi_i = \psi + \varsigma_i \text{ for } k = 1, \ldots, K, \text{ and } i = 1, \ldots, n.$$
$$\boldsymbol{\eta}_i = (\boldsymbol{\zeta}_i^\intercal, \varsigma_i)^\intercal \sim IID(\mathbf{0}, \Omega_\eta).$$

The common mean parameters β_k for $k = 1, \ldots, K$ and ψ are the the objects of interest, and they can be consistently estimated with the following mean group (MG) estimator given n and T are both large, with

[7] We thank an anonymous referee for pointing out that there might be a competing channel that drives our empirical results. Stocks that attract media attention behave differently than stock that are not covered by the media (Fang and Peress (2009), Barber and Odean (2008)). We include a media-attention factor that is constructed using the methodology from Fang and Peress (2009) to control for that competing channel.

[8] Full estimation results can be requested from the authors.

$\sqrt{n}/T \to 0$. [9]

$$\hat{\beta}_k^{MG} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{k,i} \text{ for } k = 1,\dots,K, \text{ and } \hat{\psi}^{MG} = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i.$$

For a heterogeneous coefficient spatial model, we can only identify the spatial coefficients of those units with at least one link. Spatial coefficients of those units with zero links cannot be identified, and we need to restrict them to be zeros. If we apply the full sample adjacency matrix W discussed above with threshold value $m = 1$, only $n_0 = 7$ out of $n = 394$ companies do not have any long-run links. $N_p = n - n_0 = 387$ units have unrestricted spatial coefficients. In contrast, individual-specific factor coefficients and intercepts are identified for all units, with $N_p = n = 394$.

We estimate Model 1 - Model 5 (Equation 29 to Equation 33). In Table 3, we report the mean group estimates with their standard errors, and the percentages of entities with statistically significant parameters at 5% level (with multiple testing correction) when the threshold value $m = 1$. In Appendix D, Results for alternative thresholds m are reported in Table 19 and Table 20. We also report sub-sample estimation results in Table 21.

To address the multiple testing issue, we employ two methods. For factor loadings and spatial parameters, we use a simple Bonferroni correction (Bonferroni (1935)).[10] As a complement to the joint test of $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}$ proposed in Section 4.3, we also perform individual statistical tests for alphas. Finding alphas that significantly deviate from zero is a central question in finance. Since we are conducting individual tests with many test assets, a non-negligible number of alphas could be non-zero purely due to chance. To address this issue, we adopt the technique developed in Barras et al. (2010). With this approach, we could estimate the proportion of zero alphas with the false discovery rate (FDR) under control. To be more specific, We conduct the following procedure:

1. Given the collection of $p-$values $\{p_i, i = 1,\dots,n\}$, for a sufficiently high threshold λ^*, the proportion of alphas that are zero can be estimated with $\hat{\pi}_0(\lambda^*) = \frac{\hat{W}(\lambda^*)}{n}\frac{1}{1-\lambda^*}$, where $\hat{W}(\lambda^*)$ gives the number of alphas with p values that are larger than λ^*. And the optimal threshold value λ^* is selected using a bootstrap procedure that minimizes the estimated mean squared error of $\hat{\pi}_0(\lambda)$ (see Storey (2002)).

2. With an estimate of the proportion of zero-alpha stocks $\hat{\pi}_0$, for any significance level γ, we can calculate the percentage of false discoveries by $\hat{\pi}_0 * \gamma$.

3. For any significance level γ, the percentage of individually significant alphas without multiple testing corrections is S_γ. Subtracting the estimate of the percentage of false discoveries from that, we may get the percentage of individually significant alphas with multiple testing corrections as $T_\gamma = S_\gamma - \hat{\pi}_0 * \gamma$.

[9]See Pesaran and Smith (1995) for proofs of the consistency when individual-specific coefficients are independently distributed. Recent developments by Chudik and Pesaran (2019) prove the consistency under weakly correlated individual-specific estimators. In both cases, T and n are required to be large with $\sqrt{n}/T \to 0$. Intuitively, big T is required for the consistent estimation of individual-specific coefficients, and n needs to be big enough for the consistent estimation of the means. To see how the mean group estimators behave in the context of heterogeneous coefficient spatial model, see Aquaro et al. (2020).

[10]In the main text, we show the results with Bonferroni correction for simplicity. Other multiple testing techniques that are less conservative could also be applied.

	(1) Factor Component								(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ
(1) Spatial CAPM									
Mean Group Estimates	0.014	0.564							0.447
	(0.001)	(0.021)							(0.019)
% significant (at 5% level)	3.1%	79.4%							73.3%
N_p	394	394							387
(2) Spatial Factor Model (with FF3)									
Mean Group Estimates	0.013	0.528	0.127	-0.141					0.491
	(0.001)	(0.021)	(0.014)	(0.023)					(0.019)
% significant (at 5% level)	1.7%	74.4%	51.0%	69.0%					75.1%
N_p	394	394	394	394					387
(3) Spatial Factor Model (with FF5)									
Mean Group Estimates	0.012	0.541	0.142	-0.140	0.139	0.176			0.496
	(0.001)	(0.021)	(0.014)	(0.023)	(0.022)	(0.021)			(0.019)
% significant (at 5% level)	2.0%	75.1%	51.5%	68.0%	50.3%	47.5%			76.4%
N_p	394	394	394	394	394	394			387
(4) Spatial Factor Model (with FF5+MOM)									
Mean Group Estimates	0.012	0.549	0.143	-0.148	0.137	0.181	-0.018		0.487
	(0.001)	(0.022)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)		(0.019)
% significant (at 5% level)	1.4%	75.4%	51.3%	62.4%	49.7%	47.5%	30.5%		74.9%
N_p	394	394	394	394	394	394	394		387
(5) Spatial Factor Model (with FF5+MOM+MA)									
Mean Group Estimates	0.014	0.550	0.103	-0.104	0.118	0.146	-0.030	-0.185	0.492
	(0.001)	(0.022)	(0.013)	(0.019)	(0.021)	(0.022)	(0.007)	(0.023)	(0.019)
% significant (at 5% level)	2.5%	75.6%	41.6%	53.8%	45.7%	44.9%	26.9%	41.4%	74.4%
N_p	394	394	394	394	394	394	394	394	387

Table 3: **QML estimation results of Equation 29 to Equation 33 using full sample.** Note: threshold $m = 1$. For each panel, the first row gives the mean group (MG) estimates for the parameters with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level (with multiple testing correction), and the last row gives the number of unrestricted units N_p for each parameter.

The contemporaneous local dependence parameter ψ is highly statistically significant under all specifications. Local dependence also exhibit strong economic importance as the mean group estimates of ψ are around $0.45 - 0.50$ over the five models we consider. This magnitude is comparable to the average strength of the market factor, with the mean group estimate of market beta lying between $0.53 - 0.56$ across models. Comparing different models, we find that adding more common risk factors does not weaken the estimated strength of local dependence. The magnitude of the mean group estimate and the proportion of entities with statistically significant spatial parameter at 5% level do not change with the number of factors we include.

Multiple testing issue is taken into account when we calculate the proportion of individually significant parameters. Among 387 unrestricted contemporaneous spatial coefficients ψ_i, more than 70% of them are individually significant under all cases.[11] This high significance ratio implies that the linkage identification method based on news co-mentioning successfully detects the relevant links. If our data contained a lot of spurious links, we would see the individual specific spatial parameters to be insignificant for many entities.

[11] We report the results with a conservative Bonferroni correction (Bonferroni (1935)) for its simplicity. If we adopt a less conservative multiple testing correction, say the B-H procedure (Benjamini and Hochberg (1995)), we will have more than 85% of the individual ψ_i that is significant for all specifications that we consider.

Apart from those models with only tradable factors, we also investigate the below model with a mixture of both tradable and non-tradable factors (Equation 19).

- Model 6: Spatial factor model with 7 tradable factors (Fama-French five plus Momentum factor and Media Attention), and 4 macro factors (Default Spread, Term Spread, Trend, Dividend Yield)

$$\begin{aligned} \boldsymbol{r}_t = & \boldsymbol{\alpha} + \boldsymbol{\beta_1} z_{MRT,t} + \boldsymbol{\beta_2} z_{SMB,t} + \boldsymbol{\beta_3} z_{HML,t} + \boldsymbol{\beta_4} z_{RMW,t} + \boldsymbol{\beta_5} z_{CMA,t} + \boldsymbol{\beta_6} z_{MOM,t} + +\boldsymbol{\beta_7} z_{MA,t} \\ & + \boldsymbol{\beta_8} f_{DS,t} + \boldsymbol{\beta_9} f_{TS,t} + \boldsymbol{\beta_{10}} f_{TR,t} + \boldsymbol{\beta_{11}} f_{DY,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \end{aligned} \tag{34}$$

We use daily observations the same set of macro factors from Aït-sahali and Brandt (2001) and Connor et al. (2021).

- **Default Spread** (DS) is the yield difference between Moody's Baa and Aaa rated bonds.
- **Term Spread** (TS) is the yield difference between 10 and 1 year government bonds.
- **Trend** (TR) is the difference between the log of the current $S\&P$ 500 index level and the log of the average index level over the previous 12 months.
- **Dividend Yield** (DY), also called Dividend-to-Price, is the sum of dividends paid on the $S\&P$ 500 index over the past 12 months divided by the current level of the index.

	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	ψ
(6) Spatial Factor Model (with all traded factors and macro factors)													
MG	0.013	0.551	0.103	-0.103	0.119	0.146	-0.029	-0.186	-0.021	0.001	-0.058	3.891	0.491
	(0.001)	(0.022)	(0.013)	(0.019)	(0.021)	(0.022)	(0.007)	(0.023)	(0.005)	(0.002)	(0.032)	(1.135)	(0.019)
%sig	2.7%	76.4%	41.4%	53.8%	45.9%	44.7%	27.4%	41.4%	0.0%	0.0%	0.3%	0.3%	74.1%
N_p	394	394	394	394	394	394	394	394	394	394	394	394	387

Table 4: **QML estimation results of Equation 34 to using full sample.** Note: threshold $m = 1$. The first row gives the mean group (MG) estimates for the parameters with their standard errors in the parenthesis. The third row gives the percentages of unrestricted units with statistically significant parameters at 5% level (with multiple testing correction), and the last row gives the number of unrestricted units N_p for each parameter.

Full sample estimation results are reported in Table 4. One observation we make is that macro factors are not very significant in general. The mean group estimates of betas associated with default spread and dividend yield are only marginally significant. With a conservative multiple testing adjustments, the percentages of individually significant betas for macro factors are very low.[12] This result is not surprising given that we use daily data, and we expect these macro factors to play a more important role if we turn to data of lower frequency. Other results remain the same as from the previous case.

A large proportion of the news-implied links that we identify are intra-industry links. It has been documented widely that stocks within the same industry exhibit excess co-movement beyond what can be explained by common risk factors at the market level (Moskowitz and Grinblatt (1999), Fan et al. (2016), Engelberg et al. (2018)). In order to control for industry factors as an additional source of co-movement, we further augment

[12] A less conservative B-H procedure leads to a slight increase in the percentages.

Equation 29 to Equation 33 and Equation 34 with industry factors.

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{\alpha} + \mathbf{B}\mathbf{z}_t + \boldsymbol{\beta}_I f_{IND,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \\ \boldsymbol{r}_t &= \boldsymbol{\alpha} + \mathbf{B}_1\mathbf{z}_{1,t} + \mathbf{B}_2\mathbf{f}_{2,t} + \boldsymbol{\beta}_I f_{IND,t} + \boldsymbol{\Psi} W \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \end{aligned} \tag{35}$$

We use Fama French 12 equal-weighted industry portfolios. We choose to use broad industry classification and equal weighting. This is because we are dealing with $S\&P$ 500 stocks, and we do not want industry returns to be dominated by several large stocks within that industry.

Table 5 and Table 6 report the estimation results for models augmented with the industry factor. The Industry factor is highly significant in all cases, and the mean group estimate of industry beta lies between $0.39 - 0.43$. The introduction of the industry factor largely weakens the estimated effect of the market factor, and the average market beta is reduced to $0.21 - 0.26$. On the other hand, the magnitude of local dependence is only slightly reduced by the introduction of the industry factor. This shows that our results are not driven by exposure to common industry-level shocks but by granular interactions. Using other equal-weighted industry factors does not affect this finding.

	(1) Factor Component									(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_I	ψ
(1) Spatial CAPM+Industry										
MG	0.007	0.246							0.392	0.389
	(0.002)	(0.024)							(0.021)	(0.018)
% Sig	2.9%	65.0%							68.8%	67.5%
N_p	394	394							394	387
(2) Spatial Factor Model (with FF3+Industry)										
MG	0.004	0.217	-0.110	-0.157					0.427	0.419
	(0.002)	(0.025)	(0.018)	(0.019)					(0.026)	(0.018)
% Sig	2.8%	62.7%	53.6%	66.2%					66.0%	70.1%
N_p	394	394	394	394					394	387
(3) Spatial Factor Model (with FF5+Industry)										
MG	0.003	0.229	-0.100	-0.166	0.112	0.204			0.424	0.426
	(0.002)	(0.025)	(0.017)	(0.020)	(0.019)	(0.017)			(0.025)	(0.018)
% Sig	2.9%	61.9%	49.7%	65.2%	39.8%	42.4%			67.0%	69.5%
N_p	394	394	394	394	394	394			394	387
(4) Spatial Factor Model (with FF5+MOM+Industry)										
MG	0.003	0.237	-0.097	-0.155	0.109	0.197	0.006		0.417	0.425
	(0.002)	(0.026)	(0.018)	(0.018)	(0.019)	(0.017)	(0.007)		(0.025)	(0.018)
% Sig	2.6%	63.5%	49.7%	58.6%	39.6%	41.9%	23.9%		67.0%	67.8%
N_p	394	394	394	394	394	394	394		394	387
(5) Spatial Factor Model (with FF5+MOM+MA+Industry)										
MG	0.005	0.253	-0.108	-0.127	0.098	0.176	-0.002	-0.104	0.397	0.429
	(0.002)	(0.025)	(0.017)	(0.017)	(0.019)	(0.017)	(0.006)	(0.017)	(0.024)	(0.019)
% Sig	2.6%	62.7%	47.2%	47.7%	38.1%	39.8%	21.8%	28.7%	64.0%	70.1%
N_p	394	394	394	394	394	394	394	394	394	387

Table 5: **QML estimation results of spatial factor models augmented with the industry factor (Equation 35 where z includes only tradable factors) using full sample.** W is constructed using threshold $m = 1$.

	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_I	ψ
(6) Spatial Factor Model (with all traded factors and macro factors+Industry)														
MG	0.004	0.252	-0.109	-0.129	0.098	0.179	-0.003	-0.101	-0.022	0.008	-0.221	-1.847	0.401	0.428
	(0.002)	(0.026)	(0.017)	(0.017)	(0.019)	(0.017)	(0.006)	(0.017)	(0.005)	(0.002)	(0.032)	(1.126)	(0.025)	(0.018)
%sig	3.1%	62.9%	48.0%	48.2%	38.3%	39.8%	21.3%	28.4%	0.0%	0.0%	0.8%	0.5%	63.5%	69.0%
N_p	394	394	394	394	394	394	394	394	394	394	394	394	394	387

Table 6: **QML estimation results of spatial factor models augmented with the industry factor (Equation 35 where $\mathbf{z}_1$ includes all tradable factors and $\mathbf{f}_2$ includes all macro factors) using full sample.** W is constructed using threshold $m = 1$.

So far, we have shown the mean group estimation results for whole sample companies. It is also interesting to gauge heterogeneity at sub-group levels. It is reasonable to suspect that the average sensitivities to local risk spillovers are different for different industry groups. To explore this heterogeneity, we adopt the random coefficient assumptions at the industry level. Subscript g denotes industry membership, and we classify stocks into six broad industries. [13]

$$\beta_{k,i,g} = \beta_{k,g} + \zeta_{k,i,g}; \quad \psi_{i,g} = \psi_g + \varsigma_{i,g},$$
$$\boldsymbol{\eta}_{i,g} = (\boldsymbol{\zeta}_{i,g}^{\intercal}, \varsigma_{i,g})^{\intercal} \sim IID(\mathbf{0}, \Omega_\eta),$$

where $k = 1, \ldots, K$, $i = 1, \ldots, n$, and $g = 1, \ldots, G$. The industry-level common mean parameters for industry g can be consistently estimated when n_g, the number of cross-sectional units within that industry, is large by

$$\hat{\beta}_{k,g}^{MG} = \frac{1}{n_g} \sum_{i \in \mathbb{N}_g} \hat{\beta}_{k,i,g} \;\; ; \;\; \hat{\psi}_g^{MG} = \frac{1}{n_g} \sum_{i \in \mathbb{N}_g} \hat{\psi}_{i,g},$$

for $k = 1, \ldots, K$, and $g = 1, \ldots, G$.

To save space, we only report the mean group estimates by industry for the two largest spatial factor models (Equation 33 and Equation 34) in Table 7 and Table 8. Their counterparts with the industry factor are presented in Table 22 and Table 23 in Appendix D. The tables reveal that our main conclusions that equity returns are affected by that of their economic neighbors are very robust to the industry disaggregation. Local dependence is highly significant for stocks in all six industries, and the magnitudes of dependence show a considerable degree of heterogeneity across industries. The industrial mean group estimates of the spatial parameter are between $0.36 - 0.60$. After controlling for the industry factor, the estimates still range from $0.31 - 0.56$.

[13] We adopt broad industry classification to guarantee that there are a large number of stocks within each industry since mean group estimation requires large n_g to be consistent. We build the industry classification on top of the Fama-French five industry definitions where they classify all stocks according to their SIC code into five broad groups: "Consumer", "Health", "Hi-tech", "Manufacturing" and "Others". For the first four categories, we keep the same definitions as Fama and French. Since there are a large proportion of financial companies in the $S\&P500$ universe, it would be interesting to separate financial firms from those in the "Others" category. Among the stocks that fall into "Others", we categorize the stocks with a SIC in the range $6000 - 6799$ as "Finance" and put the remaining in the "Others" category.

	(1) Factor Component								(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ
Panel A: Finance									
MG	0.013	0.490	0.134	0.359	0.041	-0.019	-0.070	0.306	0.567
	(0.003)	(0.058)	(0.032)	(0.060)	(0.060)	(0.049)	(0.019)	(0.061)	(0.047)
% Sig(at 5%)	0.0%	64.7%	42.6%	64.7%	38.2%	25.0%	29.4%	45.6%	79.4%
Non-zero coef.	68	68	68	68	68	68	68	68	68
Panel B: Consumer									
MG	0.016	0.514	0.149	-0.203	0.287	0.362	-0.060	-0.152	0.426
	(0.003)	(0.044)	(0.033)	(0.023)	(0.027)	(0.030)	(0.016)	(0.025)	(0.040)
% Sig(at 5%)	4.0%	76.0%	50.7%	54.7%	48.0%	52.0%	24.0%	17.3%	73.3%
Non-zero coef.	75	75	75	75	75	75	75	75	75
Panel C: Health									
MG	0.026	0.490	0.011	-0.387	-0.155	0.145	0.077	-0.067	0.368
	(0.007)	(0.067)	(0.043)	(0.037)	(0.054)	(0.048)	(0.022)	(0.040)	(0.059)
% Sig(at 5%)	14.3%	85.7%	28.6%	89.3%	25.0%	25.0%	32.1%	14.3%	82.1%
Non-zero coef.	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech									
MG	0.016	0.645	0.129	-0.381	-0.307	0.254	-0.048	-0.153	0.418
	(0.004)	(0.045)	(0.030)	(0.027)	(0.047)	(0.041)	(0.013)	(0.028)	(0.043)
% Sig(at 5%)	5.7%	87.1%	35.7%	72.9%	40.0%	38.6%	14.3%	18.6%	68.1%
Non-zero coef.	70	70	70	70	70	70	70	70	69
Panel E: Manufacturing									
MG	0.009	0.540	0.019	-0.057	0.366	0.020	0.013	-0.505	0.591
	(0.002)	(0.044)	(0.022)	(0.023)	(0.023)	(0.053)	(0.013)	(0.041)	(0.037)
% Sig(at 5%)	0.0%	71.7%	41.6%	27.4%	65.5%	61.1%	36.3%	72.6%	84.1%
Non-zero coef.	113	113	113	113	113	113	113	113	107
Panel F: Others									
MG	0.011	0.628	0.218	-0.155	0.173	0.190	-0.070	-0.320	0.433
	(0.004)	(0.069)	(0.041)	(0.047)	(0.044)	(0.057)	(0.020)	(0.055)	(0.067)
% Sig(at 5%)	0.0%	77.5%	42.5%	50.0%	22.5%	45.0%	20.0%	50.0%	77.5%
Non-zero coef.	40	40	40	40	40	40	40	40	40

Table 7: **QML estimation results of spatial factor model with Fama-French five factors plus Momentum and Media-Attention factor (Equation 33). Parameters summarized by industry.**

	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	ψ
Panel A: Finance													
MG	0.013	0.493	0.134	0.359	0.038	-0.018	-0.071	0.309	-0.011	0.010	-0.289	-3.400	0.566
	(0.003)	(0.058)	(0.032)	(0.060)	(0.060)	(0.049)	(0.019)	(0.060)	(0.011)	(0.004)	(0.094)	(3.158)	(0.047)
%sig	0.0%	64.7%	42.6%	64.7%	38.2%	25.0%	29.4%	45.6%	0.0%	0.0%	1.5%	1.5%	79.4%
N_p	68	68	68	68	68	68	68	68	68	68	68	68	68
Panel B: Consumer													
MG	0.015	0.514	0.149	-0.202	0.288	0.362	-0.059	-0.152	-0.016	-0.001	-0.037	3.421	0.426
	(0.003)	(0.044)	(0.033)	(0.023)	(0.026)	(0.030)	(0.015)	(0.025)	(0.010)	(0.003)	(0.068)	(2.507)	(0.040)
%sig	4.0%	76.0%	50.7%	54.7%	48.0%	52.0%	25.3%	17.3%	0.0%	0.0%	0.0%	0.0%	73.3%
N_p	75	75	75	75	75	75	75	75	75	75	75	75	75
Panel C: Health													
MG	0.026	0.489	0.011	-0.388	-0.154	0.145	0.076	-0.068	-0.015	0.008	0.110	3.031	0.369
	(0.007)	(0.067)	(0.043)	(0.038)	(0.054)	(0.048)	(0.021)	(0.039)	(0.023)	(0.006)	(0.094)	(4.975)	(0.059)
%sig	17.9%	85.7%	25.0%	89.3%	25.0%	25.0%	32.1%	14.3%	0.0%	0.0%	0.0%	0.0%	82.1%
N_p	28	28	28	28	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech													
MG	0.016	0.646	0.128	-0.380	-0.308	0.253	-0.047	-0.154	-0.035	0.010	-0.096	7.012	0.417
	(0.004)	(0.045)	(0.029)	(0.027)	(0.047)	(0.041)	(0.013)	(0.028)	(0.013)	(0.004)	(0.078)	(2.607)	(0.043)
%sig	5.7%	87.1%	37.1%	72.9%	40.0%	38.6%	14.3%	17.1%	0.0%	0.0%	0.0%	0.0%	66.7%
N_p	70	70	70	70	70	70	70	70	70	70	70	70	69
Panel E: Manufacturing													
MG	0.009	0.539	0.020	-0.055	0.369	0.019	0.015	-0.508	-0.020	-0.012	0.113	8.233	0.591
	(0.002)	(0.044)	(0.022)	(0.023)	(0.023)	(0.053)	(0.012)	(0.041)	(0.008)	(0.002)	(0.051)	(1.833)	(0.037)
%sig	0.0%	73.5%	40.7%	27.4%	66.4%	60.2%	36.3%	73.5%	0.0%	0.0%	0.0%	0.0%	84.1%
N_p	113	113	113	113	113	113	113	113	113	113	113	113	107
Panel F: Others													
MG	0.011	0.630	0.219	-0.156	0.173	0.193	-0.071	-0.318	-0.031	0.003	-0.236	0.045	0.433
	(0.004)	(0.069)	(0.041)	(0.047)	(0.044)	(0.057)	(0.020)	(0.055)	(0.015)	(0.005)	(0.095)	(3.408)	(0.067)
%sig	0.0%	80.0%	42.5%	50.0%	22.5%	45.0%	22.5%	50.0%	0.0%	0.0%	0.0%	0.0%	77.5%
N_p	40	40	40	40	40	40	40	40	40	40	40	40	40

Table 8: **QML estimation results of spatial factor model with all seven tradable factors and all four macro factors (Equation 34). Parameters summarized by industry.**

Financial companies have the largest exposure to their neighbors' shocks. And this high level of sensitivity to local shocks cannot be explained by exposures to common industry shocks as the mean group estimate of common spatial parameter stays unchanged with the introduction of the industry factor. After controlling for the industry factor, the mean group estimate of ψ_g for the financial industry is still as large as 0.55 (0.05). By contrast, introducing the industry factor reduces the estimated magnitudes of local dependence for companies in the consumer industry, the health industry, and the manufacturing industry by a much larger margin. For example, the mean group estimate of ψ_g for the manufacturing industry is as large as 0.59 (0.04) before the introduction of the industry factor, and it drops to 0.41 (0.04) if we include the industry factor.

In Appendix D (see Table 24 and Table 25), we also report the estimation results when the spatial factor models are augmented with weakly exogenous spatial-temporal terms (Equation 13). For simplicity, we pre-specify $L = 5$ to control within week dynamics. We find that dynamic spatial dependence terms are also statistically significant, although much smaller in economic magnitude. For example, for the first lag ψ_1, even with a conservative multiple testing adjustment, there are around 20% of $\psi_{1,i}$ are individually significant under all cases.[14] This is consistent with the finding of Scherbina and Schlusche (2015) that a trading strategy based on this lead-lag relationship may not be profitable in reality.

Next, we examine how the spatial factor model captures the remaining dependence in the de-factored returns. Using the method described in Section 4.2, we compute the number of non-zero pair-wise cross-correlations of the residuals from (1) factor model with FF5+MOM+MA that use Fama-French five factors, Momentum factor, and Media-Attention factor; (2) spatial factor model with FF5+MOM+MA that uses Fama-French five factors, Momentum factor, and Media-Attention factor.[15]

To test $\mathcal{H}_{0,ij} : \rho_{ij} = 0$ for $n_{test} = n(n-1)/2$ pairs of (i, j), for a given family-wise error rate (FWER) p, the critical values are $\pm\frac{1}{\sqrt{T}}\Phi^{-1}(1 - p/2n_{test})$ for the factor model, and $F^{-1}(p/2n_{test})$ and $F^{-1}(1 - p/2n_{test})$ for the spatial factor model. F is the empirical null distribution from $B = 1000$ bootstrap samples. Figure 1 shows the histogram of bootstrapped $\hat{\rho}^b_{ij}$ for all $i < j, b = 1, \ldots, 1000$ for the spatial factor model. Table 9 presents the degree of cross-sectional dependence in the factor model and its spatial-augmented version under different family-wise error rates. The table shows that adding the spatial component reduces the number of non-zero pair-wise cross-correlations by a huge margin.[16] The spatial component constructed with news-implied linkages is successful at eliminating the remaining correlations from the de-factored returns.

[14] There are more than 40% of $\psi_{1,i}$ are individually significant if we apply a less conservative B-H procedure.

[15] Here, we only present the results for the models with most factors (tradable factors only, as we have seen that macro factors do not have much explanatory power for daily returns) that are supposed to have the least residual correlations among all competing models. We can do more sets factor models and their spatial-augmented versions at the cost of bootstrap inference for each spatial factor model.

[16] This is not only because we have larger critical values (in absolute values) under the spatial factor specification. Even if we do not consider the distortion in the limiting distribution of $\hat{\rho}_{ij}$ brought by the spatial component and still use the limiting distribution for factor model residual correlation coefficients (Equation 14), the percentage of non-zero pair-wise cross correlations is still reduced by half.

	Critical values	# Non-zero pair-wise cross correlations	Density
(1) $p = 0.05$			
Factor Model with FF5+MOM+MA	-0.091,0.091	9024	11.66%
Spatial Factor Model with FF5+MOM+MA	-0.252,0.266	494	0.64%
(2) $p = 0.1$			
Factor Model with FF5+MOM+MA	-0.088,0.088	9582	12.38%
Spatial Factor Model with FF5+MOM+MA	-0.238,0.238	632	0.82%

Table 9: Degree of cross-sectional dependence in the residuals. Note: Density gives the percentage of non-zero pair-wise cross correlations (i.e., density=number of non-zero pair-wise cross-correlations/n_{test}).

6.1.2 APT Test and the Degree of Mispricing

In this section, following Kou et al. (2018), we use the procedures established in Section 4.3 to test the implications of no asymptotic arbitrage for the following factor models and their spatial-augmented versions (Model 1-6 from Equation 29 to Equation 33 and Equation 34).

- Model 1.1: CAPM model (CAPM)
- Model 1.2: Spatial CAPM model (CAPM(S))
- Model 2.1: Factor model with Fama-French three factors (FF3)
- Model 2.2: Spatial factor model with Fama-French three factors (FF3(S))
- Model 3.1: Factor model with Fama-French five factors (FF5)
- Model 3.2: Spatial factor model with Fama-French five factors (FF5(S))
- Model 4.1: Factor model with Fama-French five factors, and momentum factor (FF5+MOM)
- Model 4.2: Spatial factor model with Fama-French five factors, and momentum factor (FF5+MOM(S))
- Model 5.1: Factor model with Fama-French five factors, momentum factor, and media attention factors (FF5+MOM+MA)
- Model 5.2: Spatial factor model with Fama-French five factors, momentum factor, and media attention factors (FF5+MOM+MA(S))
- Model 6.1: Factor model with all tradable factors and macro factors ($F_{traded} + F_{macro}$)
- Model 6.2: Spatial Factor model with all tradable factors and macro factors ($F_{traded} + F_{macro}$(S))

Table 10 shows the standardised Wald (SW) statistics and the $p-$values of the tests. All factor models are rejected. On the other hand, all spatial factor models, except the one with only market factor (Model 1.2), cannot be rejected at 5% significance level.

	SW	$p-$value
Panel A: Factor Models		
Model 1.1: CAPM	3.2816	0.0005
Model 2.1: FF3	3.2537	0.0006
Model 3.1: FF5	3.0766	0.0010
Model 4.1: FF5+MOM	3.2566	0.0006
Model 5.1: FF5+MOM+MA	3.8500	0.0000
Model 6.1: $F_{traded} + F_{macro}$	2.2918	0.0109
Panel B: Spatial Factor Models		
Model 1.2: CAPM(S)	1.688	0.0457
Model 2.2: FF3(S)	-0.42689	0.6653
Model 3.2: FF5(S)	-0.1725	0.5685
Model 4.2: FF5+MOM(S)	-0.3999	0.6554
Model 5.2: FF5+MOM+MA(S)	0.3944	0.3466
Model 6.2: $F_{traded} + F_{macro}$(S)	-0.8251	0.7954

Table 10: Testing the implications of no asymptotic arbitrage for different factor models and their spatial-augmented versions. SW column gives the standardised wald statistics, and $p-$value column gives the associated $p-$values of the tests.

As a complement to the joint test, for models with only tradable factors, we also employ the following statistics to evaluate the performance of different models.

1. The percentage of individually significant alphas. Corollary 2 suggests that the number of large deviations from the null for individual alphas is important. Early literature such as Jensen (1968) and Jensen et al. (1972) also look at n individual tests. As elaborated in Section 6.1.1, we apply the False Discovery Rate(FDR) approach from Barras et al. (2010) to deal with the multiple testing issue.

2. Average L_1 norm of intercepts $A(\mid \alpha_i \mid)$.

3. Average L_2 norm of intercepts $A(\alpha_i^2)$ (Fama and French (2015)).

Those three statistics are implied by theorem 1 and the corollaries to be useful in measuring how good the approximations are. In addition to those three measures of mispricing, we also report the average adjusted R^2 of different models as a measure of goodness of fit. The adjusted R^2 of the ith asset is defined as $R_i^2 = 1 - \frac{T-1}{T-K-1}\frac{Var(\epsilon_i)}{Var(r_i)}$, where K is the number of coefficients in the equation, excluding the constant.

	% of significant α_i	$A(\mid \alpha_i \mid)$	$A(\alpha_i^2)$	Adjusted R^2
Model 1.1: CAPM	5.37%	2.51	10.91	0.4059
Model 1.2: CAPM(S)	3.06%	2.43	10.41	0.4655
Model 2.1: FF3	4.89%	2.49	10.66	0.4294
Model 2.2: FF3(S)	1.74%	2.31	9.39	0.4823
Model 3.1: FF5	4.15%	2.45	11.17	0.4397
Model 3.2: FF5(S)	1.99%	2.21	9.39	0.4920
Model 4.1: FF5+MOM	4.17%	2.49	11.18	0.4428
Model 4.2: FF5+MOM(S)	1.45%	2.11	8.44	0.4920
Model 5.1: FF5+MOM+MA	5.02%	2.49	11.18	0.4472
Model 5.2: FF5+MOM+MA(S)	2.48%	2.16	8.56	0.4956

Table 11: Summary of model performance for factor models and spatial factor models. Note: Each panel shows the performance statistics of a factor model, and its spatial-augmented version, when all factors are tradable portfolios . Note: $\boldsymbol{\alpha}$ used to compute $A(\mid \alpha_i \mid)$ and $A(\alpha_i^2)$ are in basis point. For each column, the best statistic is highlighted in red.

Table 11 shows that for all factor models, adding spatial interactions leads to noticeable improvements in reducing mispricing errors and boosting model fittings. In a case where models include macro factors, our null is no longer $\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}$, so the three mispricing measures discussed earlier are no longer valid.[17] Two versions of Model 6 that contain both tradable and non-tradable factors do not improve the model fitting (measured by the average adjusted R^2) further. And the proportions of individual deviations from null are both around 4%, which are also higher than other spatial factor models with tradable factors only. This is due to the fact that macro factors do not play an important role in explaining daily returns. And among all the competing models that we have considered, Model 4.2 performs the best in terms of the minimizing mispricing errors.

6.1.3 Alternative Networks

In this section, we gauge whether the news-implied links carry additional information on top of existing linkage datasets. We ask two questions. Firstly, does the spatial factor model with news-implied W outperform the models with W constructed using other existing linkage datasets? Secondly, conditional on other existing linkages, do local risk spillovers via our news-implied links continue to be significant? We consider the following competing networks:

- Industry-based adjacency matrices (based on Fama-French 12 Industry). This is motivated by Moskowitz and Grinblatt (1999), Engelberg et al. (2018) and Fan et al. (2016). The adjacency matrix is a block-diagonal matrix where companies within the same industry are fully connected.
- IBES analyst co-coverage networks. It has been documented that shared analyst coverage is a strong proxy for fundamental linkages between firms and reflects firm similarities along many dimensions (Ali and Hirshleifer (2020), Israelsen (2016), Kaustia and Rantala (2013)). We use the Institutional Brokers

[17] Similar to looking at individual t tests on alphas, in the case with macro factors, we could look at n individual tests $\mathcal{H}_0 : g_i = 0$ for $i = 1, \ldots, n$, where g_i is the ith element of $\boldsymbol{g}(\boldsymbol{\theta}_s)$ from Equation 21.

Estimate System (IBES) detail history files to construct the analyst co-coverage-based adjacency matrix. For each year in the sample, we consider a stock is covered by an analyst if the analyst issues at least one FY1 or FY2 earnings forecast for the stock during the year. And we consider two stocks as linked if there are common analysts during the year, weighted by the number of common analysts. We then add up the yearly adjacency matrices to get the full sample adjacency matrix.

- Customer-supplier links (Cohen and Frazzini (2008)) from Andrea Frazzini's data library. The strength of links is weighted by sales.
- Geographic links (Pirinsky and Wang (2006) and Parsons et al. (2020)). We obtain location information from CRSP Compustat merged files. We then merge the sample firms with the Metropolitan Statistical Areas (MSA) data using the ZIP-FIPS-MSA data from the US Department of Labor, which maps zip codes to MSAs. We follow Pirinsky and Wang (2006), and consider firms whose headquarters are in the same MSA as linked.
- Partial correlation network (Barigozzi and Hallin (2017)).[18] To construct the adjacency matrix, we use elastic net and BIC information criterion to estimate the following linear regressions

$$v_{it} = \sum_{j \neq i}^{n} \beta_{ij} v_{jt} + e_{it} \text{ for } i = 1, \ldots, n,$$

where v_{it} is the factor model (FF5+MOM) residual for stock i at t. For the adjacency matrix before normalizing, we let $w_{ij} = 1$ if $\beta_{ij} \neq 0$.

	% of significant α_i	$A(\mid \alpha_i \mid)$	$A(\alpha_i^2)$	Adjusted R^2
$W_{Industry}$	3.73%	2.12	8.91	0.4967
W_{IBES}	3.25%	2.25	9.45	0.4933
$W_{Customer-Supplier}$	4.15%	2.51	10.87	0.4450
$W_{Geographic}$	3.12%	2.46	10.64	0.4591
$W_{Partial}$	1.63%	2.01	8.04	0.5213
W_{News}	1.45%	2.11	8.40	0.4920

Table 12: Summary of model performance using competing networks. Note: This table shows the performance of the competing adjacency matrices under the spatial factor model with Fama-French five factors, and Momentum factor (the best model according to Table 11). $\boldsymbol{\alpha}$ used to compute $A(\mid \alpha_i \mid)$ and $A(\alpha_i^2)$ are in basis point. For each column, the best statistic is highlighted in red.

Table 12 shows the performance of the competing networks. Among adjacency matrices constructed using non-return data, the spatial factor model estimated with new-implied linkages outperforms other networks in terms of minimizing pricing errors. The partial correlation network constructed using penalising method is purely data-driven, and it very much reflects the true connectivity structure among the de-factored returns in the sample. It is not surprising that the partial correlation network performs the best among most of the criteria that we use.[19] The fact that our news-implied network performs reasonably close to the partial correlation

[18] We thank an anonymous referee for pointing out that we could compare our results with W estimated using penalising methods.

[19] It should be pointed out that despite the good in-sample performance, a spatial factor model with $W_{Partial}$ is not theoretically justified given that $W_{Partial}$ is estimated with return data.

network shows that it is a very nice proxy of firm-to-firm connectivity. And the advantage of W_{News} is that it is interpretable while $W_{Partial}$ is purely data-driven.

Next, we examine whether our news-implied linkages carry additional information on top of existing linkages documented? To do that, we estimate the two-W spatial factor models below, with W_1 being our news-implied networks and W_2 being a set of other candidate matrices. For simplicity, we only demonstrate the case where all factors are tradable.

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{z}_t + \boldsymbol{\Psi}_1 W_1 \boldsymbol{r}_t + \boldsymbol{\Psi}_2 W_2 \boldsymbol{r}_t + \boldsymbol{\epsilon}_t. \tag{36}$$

Table 13 shows the estimation results for Equation 36. Although the magnitude of local dependence among news-implied peers is weakened by the introduction of other networks, our new-implied links are still important channels of risk spillovers. The mean group estimates of ψ_1 are around $0.29 - 0.49$. Given a conservative Bonferroni correction, there are still more than 50% of parameters being individually significant across different specifications. The results confirm that the novel dataset carries additional information on top of existing networks. The statistically and economically significant local dependence among the news-implied peers cannot be explained by other existing networks.

	(1) Factor Component							**(2) Spatial Component**	
	α	β_1	β_2	β_3	β_4	β_5	β_6	ψ_1	ψ_2
Panel(1): $W_1 = W_{news}$ **and** $W_2 = W_{Industry}$									
Mean Group Estimates	0.007	0.291	0.097	-0.119	0.081	0.147	-0.011	0.347	0.385
	(0.001)	(0.024)	(0.013)	(0.018)	(0.018)	(0.017)	(0.007)	(0.018)	(0.020)
% significant (at 5% level)	1.1%	60.2%	39.1%	40.6%	31.5%	36.8%	20.6%	61.0%	71.1%
N_p	394	394	394	394	394	394	394	387	394
Panel(2): $W_1 = W_{news}$ **and** $W_2 = W_{IBES}$									
Mean Group Estimates	0.012	0.486	0.126	-0.115	0.093	0.165	-0.025	0.368	0.197
	(0.001)	(0.022)	(0.013)	(0.018)	(0.019)	(0.019)	(0.007)	(0.019)	(0.014)
% significant (at 5% level)	1.3%	72.3%	46.2%	51.5%	40.4%	38.6%	26.1%	67.4%	61.8%
N_p	394	394	394	394	394	394	394	387	340
Panel(3): $W_1 = W_{news}$ **and** $W_2 = W_{Customer-Supplier}$									
Mean Group Estimates	0.012	0.538	0.144	-0.145	0.132	0.178	-0.018	0.485	0.082
	(0.001)	(0.022)	(0.014)	(0.021)	(0.021)	(0.021)	(0.007)	(0.019)	(0.020)
% significant (at 5% level)	1.4%	74.6%	50.5%	61.4%	49.0%	46.4%	30.2%	77.3%	44.1%
N_p	394	394	394	394	394	394	394	387	59
Panel(4): $W_1 = W_{news}$ **and** $W_2 = W_{Geographic}$									
Mean Group Estimates	0.011	0.491	0.135	-0.141	0.124	0.192	-0.019	0.462	0.093
	(0.001)	(0.025)	(0.014)	(0.020)	(0.020)	(0.020)	(0.007)	(0.018)	(0.019)
% significant (at 5% level)	1.4%	65.7%	48.0%	59.4%	45.7%	45.4%	27.7%	74.7%	32.6%
N_p	394	394	394	394	394	394	394	387	307
Panel(5): $W_1 = W_{news}$ **and** $W_2 = W_{Partial}$									
Mean Group Estimates	0.008	0.286	0.073	-0.121	0.065	0.138	-0.004	0.292	0.421
	(0.001)	(0.025)	(0.012)	(0.018)	(0.016)	(0.017)	(0.006)	(0.015)	(0.021)
% significant (at 5% level)	0.9%	61.4%	34.8%	45.9%	30.7%	33.5%	22.6%	53.7%	83.5%
N_p	394	394	394	394	394	394	394	387	394

Table 13: **QML estimation results of two-W spatial factor model with Fama-French five factors, and Momentum Factor (Equation 36).**

7 Conclusion

We characterize how both strong and weak/local dependencies affect asset returns using a flexible heterogeneous coefficient spatial factor model. Theoretically, we derive the testable implications of no asymptotic arbitrage for such a model, and we also provide the associated Wald tests. Empirically, we focus on the weak/local dependency in equity returns, which is an area less explored in empirical financial studies due to data availability issues. Utilizing the novel business news-implied linkage data, we construct the channels through which the local shocks transmit. We find that stocks linked via business news co-mentioning exhibit excess co-movement beyond that can be explained by standard asset pricing models like CAPM and APT. Exposures to common risk factors and local interactions are two distinct mechanisms that jointly explain the comovement in asset returns. It is important for investors and policymakers to separately analyze the two types of dependencies to fully understand what type of risk they are exposed to.

References

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Y. Aït-sahali and M. W. Brandt. Variable selection for portfolio choice. *The Journal of Finance*, 56(4):1297–1351, 2001.

U. Ali and D. Hirshleifer. Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics*, 136(3):649–675, 2020.

D. W. Andrews. Asymptotic results for generalized wald tests. *Econometric Theory*, 3(3):348–358, 1987.

M. Aquaro, N. Bailey, and M. H. Pesaran. Estimation and inference for spatial models with heterogeneous coefficients: an application to us house prices. *Journal of Applied Econometrics*, 2020.

N. Bailey, S. Holly, and M. H. Pesaran. A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31(1):249–280, 2016.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Exponent of cross-sectional dependence for residuals. *Sankhya B*, 81(1):46–102, 2019a.

N. Bailey, M. H. Pesaran, and L. V. Smith. A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, 208(2):507–534, 2019b.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Measurement of factor strenght: Theory and practice. 2020.

S. R. Baker, N. Bloom, and S. J. Davis. Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636, 2016.

B. M. Barber and T. Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

M. Barigozzi and C. Brownlees. Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34 (3):347–364, 2019.

M. Barigozzi and M. Hallin. A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):581–605, 2017.

L. Barras, O. Scaillet, and R. Wermers. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, 65(1):179–216, 2010.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

P. J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008.

C. E. Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. *Studi in onore del professore salvatore ortu carboni*, pages 13–60, 1935.

S. Bryzgalova. Spurious factors in linear asset pricing models. *LSE manuscript*, 1(3), 2015.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, pages 1281–1304, 1983.

J. Chen, Y. Shin, and C. Zheng. Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics*, 2021.

L. Chen, M. Pelger, and J. Zhu. Deep learning in asset pricing. *Available at SSRN 3350138*, 2020.

A. Chudik and M. H. Pesaran. Mean group estimation in presence of weakly cross-correlated estimators. *Economics Letters*, 175:101–105, 2019.

A. Chudik, M. H. Pesaran, and E. Tosetti. Weak and strong cross-section dependence and estimation of large panels, 2011.

L. Cohen and A. Frazzini. Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011, 2008.

G. Connor, S. Li, and O. B. Linton. A dynamic semiparametric characteristics-based model for optimal portfolio selection. *Michael J. Brennan Irish Finance Working Paper Series Research Paper*, (21-1), 2021.

E. Dimson. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2):197–226, 1979.

J. Engelberg, A. Ozoguz, and S. Wang. Know thy neighbor: Industry clusters, information spillovers, and market efficiency. *Journal of Financial and Quantitative Analysis*, 53(5):1937–1961, 2018.

F. Eugene. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320, 2011.

J. Fan, Y. Liao, and J. Yao. Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83 (4):1497–1541, 2015.

J. Fan, A. Furger, and D. Xiu. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics*, 34(4):489–503, 2016.

L. Fang and J. Peress. Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5): 2023–2052, 2009.

X. Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 2011.

P. Gagliardini, E. Ossola, and O. Scaillet. Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046, 2016.

P. Gagliardini, E. Ossola, and O. Scaillet. Estimation of large dimensional conditional factor models in finance. *Handbook of Econometrics*, 7:219–282, 2020.

D. Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

S. Giglio, Y. Liao, and D. Xiu. Thousands of alpha tests. *The Review of Financial Studies*, 34(7):3456–3496, 2021.

S. Gu, B. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33 (5):2223–2273, 2020.

G. Hale and J. A. Lopez. Monitoring banking system connectedness with big data. *Journal of Econometrics*, 212(1):203–220, 2019.

C. R. Harvey, Y. Liu, and H. Zhu. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

G. Hoberg and G. Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016.

J. E. Ingersoll Jr. Some results in the theory of arbitrage pricing. *The Journal of Finance*, 39(4):1021–1039, 1984.

R. D. Israelsen. Does common analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis*, 51(4):1193–1229, 2016.

M. C. Jensen. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, 23(2):389–416, 1968.

M. C. Jensen, F. Black, and M. S. Scholes. The capital asset pricing model: Some empirical tests. 1972.

E. Juárez-Ruiz, R. Cortés-Maldonado, and F. Perez-Rodriguez. Relationship between the inverses of a matrix and a submatrix. *Computación y Sistemas*, 20(2):251–262, 2016.

A. Juodis and S. Reese. The incidental parameters problem in testing for remaining cross-section correlation. *Journal of Business & Economic Statistics*, pages 1–13, 2021.

M. Kaustia and V. Rantala. Common analyst-based method for defining peer firms. *Available at SSRN*, 2013.

Z. T. Ke, B. T. Kelly, and D. Xiu. Predicting returns with text data. Technical report, National Bureau of Economic Research, 2019.

H. H. Kelejian and I. R. Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

H. H. Kelejian and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2):509–533, 1999.

S. Kou, X. Peng, and H. Zhong. Asset pricing with spatial interaction. *Management Science*, 64(5):2083–2101, 2018.

G. M. Kuersteiner and I. R. Prucha. Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88(5):2109–2146, 2020.

L.-F. Lee. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925, 2004.

L.-f. Lee. Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137(2):489–514, 2007.

L.-f. Lee and J. Yu. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of econometrics*, 154(2):165–185, 2010.

J. P. LeSage. An introduction to spatial econometrics. *Revue d'économie industrielle*, (123):19–44, 2008.

J. P. LeSage and Y.-Y. Chih. A bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics*, 72:58–73, 2018.

J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2019.

E. Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, 21(1): 255–285, 1993.

T. J. Moskowitz and M. Grinblatt. Do industries explain momentum? *The Journal of finance*, 54(4):1249–1290, 1999.

C. A. Parsons, R. Sabbatucci, and S. Titman. Geographic lead-lag effects. *The Review of Financial Studies*, 33 (10):4721–4770, 2020.

H. Pesaran et al. General diagnostic tests for cross-sectional dependence in panels. *University of Cambridge, Cambridge Working Papers in Economics*, 435, 2004.

M. H. Pesaran and R. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, 68(1):79–113, 1995.

M. H. Pesaran and C. F. Yang. Estimation and inference in spatial models with dominant units. *Journal of Econometrics*, 221(2):591–615, 2021.

C. Pirinsky and Q. Wang. Does corporate headquarters location matter for stock returns? *The Journal of Finance*, 61(4):1991–2015, 2006.

S. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

A. Scherbina and B. Schlusche. Economic linkages inferred from news stories and the predictability of stock returns. *Available at SSRN 2363436*, 2015.

G. Schwenkler and H. Zheng. The network of firms implied by the news. *Available at SSRN 3320859*, 2019.

G. Schwenkler and H. Zheng. Peer co-movement in crypto markets. *Available at SSRN 3572471*, 2020.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

W. Shi and L.-f. Lee. Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 197(2):323–347, 2017.

J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

J. Yu, R. De Jong, and L.-f. Lee. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics*, 146(1):118–134, 2008.

Appendices

A APT Theory

A.1 Proof of Theorem 1

Under Assumption Equation 2.3, $(I - \mathbf{\Psi}W)$ is invertible, and we denote the inverse as $G(\boldsymbol{\psi}) = (I - \mathbf{\Psi}W)^{-1}$. We rewrite the spatial factor model as

$$\boldsymbol{r} = G\boldsymbol{\alpha} + G\mathbf{B}\mathbf{f} + G\boldsymbol{\epsilon}.$$

We let

$$\boldsymbol{\alpha}^* = G\boldsymbol{\alpha}, \;\; \mathbf{B}^* = G\mathbf{B}, \;\; \boldsymbol{\epsilon}^* = G\boldsymbol{\epsilon}.$$

The spatial factor model can be written as a reduced-form factor model

$$\boldsymbol{r} = \boldsymbol{\alpha}^* + \mathbf{B}^*\mathbf{f} + \boldsymbol{\epsilon}^*. \tag{37}$$

In particular, the covariance matrix of the reduced form error is

$$\Omega^* = E(\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\intercal}) = G\Omega G^\intercal.$$

We follow Ingersoll Jr (1984), and factor the positive definite covariance matrix Ω^* as $\Omega^* = CC^\intercal$, where C is a nonsingular matrix. Now consider a subsequences of assets. For the nth economy, consider the orthogonal projection of the vector $(C^n)^{-1}\boldsymbol{\alpha}^{*n}$ into the space spanned by $(C^n)^{-1}\mathbf{1}^n$ and the columns of $(C^n)^{-1}\mathbf{B}^{*n}$ as follows:

$$(C^n)^{-1}\boldsymbol{\alpha}^{*n} = (C^n)^{-1}\mathbf{1}^n\lambda_0^n + (C^n)^{-1}\mathbf{B}^{*n}\boldsymbol{\lambda}^n + \boldsymbol{u}^n.$$

By the nature of orthogonal projection,

$$0 = (\mathbf{B}^{*n})^\intercal((C^n)^\intercal)^{-1}\boldsymbol{u}^n = (\mathbf{1}^n)^\intercal((C^n)^\intercal)^{-1}\boldsymbol{u}^n.$$

Define pricing error as $\boldsymbol{v}^n = \boldsymbol{\alpha}^n - (G^n)^{-1}\mathbf{1}^n\lambda_0^n - \mathbf{B}^n\boldsymbol{\lambda}^n$, the reduced form pricing error from the reduced form factor model Equation 37 is $\boldsymbol{v}^{*n} = G\boldsymbol{v}^n$

$$\boldsymbol{v}^{*n} = \boldsymbol{\alpha}^{*n} - \mathbf{1}_n\lambda_0^n - \mathbf{B}^{*n}\boldsymbol{\lambda}^n = G^n(\boldsymbol{\alpha}^n - (G^n)^{-1}\mathbf{1}^n\lambda_0^n - \mathbf{B}^n\boldsymbol{\lambda}^n) = G^n\boldsymbol{v}^n.$$

Given that, the reduced form pricing error $\boldsymbol{v}^{*n}$ can be written as $\boldsymbol{v}^{*n} = C^n\boldsymbol{u}^n$. Using the orthogonal conditions and the factorization $\Omega^{*n} = C^n(C^n)^\intercal$, we have:

$$(\mathbf{B}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n} = (\mathbf{1}^n)^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n} = 0.$$

Consider a zero cost portfolio $\boldsymbol{c}^n = (\Omega^{*n})^{-1}\boldsymbol{v}^{*n}[(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}$

$$(\mathbf{1}^n)^\intercal\boldsymbol{c}^n = (\mathbf{1}^n)^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}[(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1} = 0,$$

with expected return

$$E((\boldsymbol{c}^n)^\intercal\boldsymbol{r}^n) = (\boldsymbol{c}^n)^\intercal\boldsymbol{\alpha}^{*n} = [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}(\mathbf{1}^n\lambda_0^n + \mathbf{B}^{*n}\boldsymbol{\lambda}^n + \boldsymbol{v}^{*n}) = 1,$$

and variance

$$\begin{aligned}
Var((\boldsymbol{c}^n)^\intercal\boldsymbol{r}^n) &= (\boldsymbol{c}^n)^\intercal Var(\boldsymbol{r}^n)\boldsymbol{c}^n \\
&= [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}(\mathbf{B}^{*n}(\mathbf{B}^{*n})^\intercal + \Omega^{*n})(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}[(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1} \\
&= [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1} = [(G^n\boldsymbol{v}^n)^\intercal(G^n\Omega^n(G^n)^\intercal)^{-1}G^n\boldsymbol{v}^n]^{-1} \\
&= [(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n)]^{-1}.
\end{aligned}$$

If the weighted sum of squared pricing errors $(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n)$ is not uniformly bounded, then the variance of this portfolio would go to zero along some subsequence, and the asymptotic arbitrage opportunity described in Section 3 exists.

We have established that no asymptotic arbitrage requires that

$$(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n) \le V_1 < \infty \text{ for all } n.$$

Under Assumption A1, we have

$$(\boldsymbol{v}^n)^\intercal(\boldsymbol{v}^n) \le \bar{\sigma}^2(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n) \le \bar{\sigma}^2 V_1 = V < \infty \text{ for all } n.$$

So we have the sum of squared pricing errors being uniformly bounded. Q.E.D.

A.2 Proof of Corollary 1.1

When there is a risk-free asset with rate r_f, we can write Equation 3 as:

$$\boldsymbol{r} - r_f \mathbf{1} = \boldsymbol{\alpha} - r_f \mathbf{1} + r_f \boldsymbol{\Psi} W \mathbf{1} + \mathbf{B}\mathbf{f} + \boldsymbol{\Psi} W(\boldsymbol{r} - r_f \mathbf{1}) + \boldsymbol{\epsilon},$$

which is equivalent to

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\mathbf{f} + \boldsymbol{\Psi} W \tilde{\boldsymbol{r}} + \boldsymbol{\epsilon}, \tag{38}$$

where $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I - \boldsymbol{\Psi} W)\mathbf{1} r_f$. We let

$$\tilde{\boldsymbol{\alpha}}^* = G\tilde{\boldsymbol{\alpha}}, \quad \mathbf{B}^* = G\mathbf{B}, \quad \boldsymbol{\epsilon}^* = G\boldsymbol{\epsilon}.$$

The spatial factor model can be written as a reduced-form factor model

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}}^* + \mathbf{B}^*\mathbf{f} + \boldsymbol{\epsilon}^*. \tag{39}$$

We follow Ingersoll Jr (1984), and factor the positive definite covariance matrix Ω^* as $\Omega^* = CC^\intercal$, where C is a nonsingular matrix. Now consider a subsequences of assets. For the nth economy, consider the orthogonal projection of the vector $(C^n)^{-1}\tilde{\boldsymbol{\alpha}}^{*n}$ into the space spanned by the columns of $(C^n)^{-1}\mathbf{B}^{*n}$ as follows:

$$(C^n)^{-1}\tilde{\boldsymbol{\alpha}}^{*n} = (C^n)^{-1}\mathbf{B}^{*n}\boldsymbol{\lambda}^n + \boldsymbol{u}^n.$$

By the nature of orthogonal projection,

$$(\mathbf{B}^{*n})^\intercal((C^n)^\intercal)^{-1}\boldsymbol{u}^n = 0.$$

Define the pricing errors as:

$$\boldsymbol{v}^{*n} = \tilde{\boldsymbol{\alpha}}^{*n} - \mathbf{B}^{*n}\boldsymbol{\lambda}^n = C^n\boldsymbol{u}^n = G^n\boldsymbol{v}^n.$$

Consider a zero cost portfolio with no factor risk $\boldsymbol{c}^n = (\Omega^{*n})^{-1}\boldsymbol{v}^{*n}[(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}$, which has expected return

$$E((\boldsymbol{c}^n)^\intercal\tilde{\boldsymbol{r}}^n) = (\boldsymbol{c}^n)^\intercal\tilde{\boldsymbol{\alpha}}^{*n} = [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}(\mathbf{B}^{*n}\boldsymbol{\lambda}^n + \boldsymbol{v}^{*n}) = 1,$$

and variance

$$\begin{aligned}
Var((\boldsymbol{c}^n)^\intercal\tilde{\boldsymbol{r}}^n) &= (\boldsymbol{c}^n)^\intercal Var(\tilde{\boldsymbol{r}}^n)\boldsymbol{c}^n \\
&= [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1}(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}(\mathbf{B}^{*n}(\mathbf{B}^{*n})^\intercal + \Omega^{*n})(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}[(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1} \\
&= [(\boldsymbol{v}^{*n})^\intercal(\Omega^{*n})^{-1}\boldsymbol{v}^{*n}]^{-1} = [(G^n\boldsymbol{v}^n)^\intercal(G^n\Omega^n(G^n)^\intercal)^{-1}G^n\boldsymbol{v}^n]^{-1} \\
&= [(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n)]^{-1}.
\end{aligned}$$

So we establish that

$$\tilde{\boldsymbol{\alpha}}^n \approx \mathbf{B}^n\boldsymbol{\lambda}^n. \tag{40}$$

If the weighted sum of squared pricing errors $(\boldsymbol{v}^n)^\intercal(\Omega^n)^{-1}(\boldsymbol{v}^n)$ is not uniformly bounded, then the variance of this portfolio would go to zero along some subsequence, and the asymptotic arbitrage opportunity described in Section 3 exists. As in theorem 1, Equation 40 also implies a unweighted pricing error bound.

From theorem 5 of Ingersoll Jr (1984), when the factor loading matrix $\mathbf{B}$ is a **complete factor representation**, the kth factor premium can be uniquely identified as the excess expected return on a well diversified portfolio with a loading of one on the kth risk factor, and a zero loading on all the rest. That is to say, $\boldsymbol{\lambda} = E(\tilde{\boldsymbol{f}}) - 1r_f$ when all factors are traded portfolios, where 1is a vector of ones that is conformable. Following their definition below:

Definition 1. A factor economy representation is **diagonal** if the correlation matrix of residuals has uniformly bounded spectral norm for all n. And a factor economy has **bounded residual variation** if the covariance matrix of residuals has uniformly bounded spectral norm for all n.

Lemma 1. The reduced form factor model (Equation 39) implied by the spatial factor model (Equation 3) has **bounded residual variation**, and is **diagonal** under Assumptions A1-A4.

Proof: for the reduced form factor model Equation 39, the sequence of reduced form error covariance matrix $\{\Omega^{*n}\}$ has uniformly bounded eigenvalues under Assumptions A1-A4 given that:

$$\|\Omega^*\| = \|G\Omega G^\intercal\| \leq \bar{\sigma}^2 \|G\| \|G^\intercal\| \leq \bar{\sigma}^2 \|G\|_\infty^2 \leq \bar{\sigma}^2 c^2 < \infty \text{ for all } n.$$

And under the Assumption A1 that residual variances are uniformly bounded away from zero, this factor economy is also diagonal given that:

$$\|R^*\| = \left\| D_{\Omega^*}^{-1/2} \Omega^* D_{\Omega^*}^{-1/2} \right\| \leq \frac{1}{\underline{\sigma}^2} \|\Omega^*\| \leq \frac{\bar{\sigma}^2}{\underline{\sigma}^2} c^2 < \infty \text{ for all } n.$$

Q.E.D.

Definition 2. A factor loading matrix $\mathbf{B}$ is a **complete factor representation** if it is **regular**, namely

$$\lim_{n\to\infty} \left\| (\mathbf{B}_n^\intercal \mathbf{B}_n)^{-1} \right\| = 0,$$

and has uniformly **bounded residual variation**.

Lemma 2. The reduced form factor model (Equation 39) implied by the spatial factor model (Equation 3) is a **complete factor representation** under Assumptions A1-A4.

Lemma 1 has established that the reduced form factor model (Equation 39) has bounded residual variation under Assumptions A1-A4. In addition, Assumption A2 guarantees that the factor representation is regular. Together, the complete factor representation condition is satisfied under our Assumptions A1-A4.

Re-write Equation 38 as:

$$\tilde{\boldsymbol{r}} = \breve{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{z}} + \boldsymbol{\Psi} W \tilde{\boldsymbol{r}} + \boldsymbol{\epsilon},$$

where $\breve{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}1r_f - \mathbf{B}E(\tilde{\boldsymbol{f}}) = \boldsymbol{\alpha} - (I - \boldsymbol{\Psi} W)\mathbf{1}r_f + \mathbf{B}1r_f - \mathbf{B}E(\tilde{\boldsymbol{f}})$, for $\mathbf{1}$ being the $n \times 1$ unit vector and 1 being a vector of ones that is conformable. Given Lemma 2, $\boldsymbol{\lambda} = E(\tilde{\boldsymbol{f}}) - 1r_f$ when factors are traded portfolios. Combining the results from Equation 40, we have the below results:

$$\breve{\boldsymbol{\alpha}}^n = \tilde{\boldsymbol{\alpha}}^n + \mathbf{B}^n 1 r_f - \mathbf{B}^n E(\tilde{\boldsymbol{f}}) \approx \mathbf{B}^n \boldsymbol{\lambda}^n + \mathbf{B}^n 1 r_f - \mathbf{B}^n E(\tilde{\boldsymbol{f}}) = \mathbf{0}.$$

Q.E.D.

A.3 Proof of Corollary 1.2

This proof is heavily borrowed from Kou et al. (2018). We first rewrite the spatial-factor model with the dependent variable being the excess returns (Equation 6) as

$$(I - \mathbf{\Psi}W)\tilde{\boldsymbol{r}} = G^{-1}\tilde{\boldsymbol{r}} = \breve{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{z}} + \boldsymbol{\epsilon}.$$

Suppose the asset returns from an infinite economy are generated by:

$$(G^n)^{-1}\tilde{\boldsymbol{r}}^n = \breve{\boldsymbol{\alpha}}^n + \mathbf{B}^n\tilde{\mathbf{z}} + \boldsymbol{\epsilon}^n.$$

For any fixed $\delta > 0$, assume $I(\mid \breve{\alpha}_j^n \mid > \delta) = 1$ for $j = 1, \ldots, N(n, \delta)$. For each of those $N(n, \delta)$ elements, we can construct a zero cost portfolio as following way.

Take the jth element for example. Denote the jth column of Identify matrix I_n by $\boldsymbol{e}_j$. If $\breve{\alpha}_j^n > \delta$, consider a zero-cost portfolio which takes a long position $\boldsymbol{e}_j^\intercal(G^n)^{-1}$ in excess returns $\tilde{\boldsymbol{r}}^n$, and short position $\boldsymbol{e}_j^\intercal\mathbf{B}^n$ in the zero-cost traded factors $\tilde{\mathbf{z}}$. If $\breve{\alpha}_j^n < -\delta$, consider a zero-cost portfolio which takes a short position $\boldsymbol{e}_j^\intercal(G^n)^{-1}$ in excess returns $\tilde{\boldsymbol{r}}^n$, and long position $\boldsymbol{e}_j^\intercal\mathbf{B}^n$ in the zero-cost traded factors $\tilde{\mathbf{z}}$. The portfolio is a zero-cost one because the long and short position are both zero-cost. The portfolio has expected return $\mid \breve{\alpha}_j^n \mid > \delta$, and variance $\sigma_j^2 < \bar{\sigma}^2$.

We can construct $N(n, \delta)$ such portfolios. Consider a new portfolio that takes equal weight in these $N(n, \delta)$ portfolios. This new portfolio is zero-cost, with expected return $\frac{\sum_{j=1}^{N(n,\delta)}|\breve{\alpha}_j^n|}{N(n,\delta)} > \delta > 0$. Under Assumption A1 that errors are uncorrelated, the variance of this portfolio is smaller than $\frac{\bar{\sigma}^2}{N(n,\delta)}$. If Equation 8 fails, and $N(n, \delta)$ is diverging, then we have asymptotic arbitrage. Q.E.D.

A.4 Proof of Corollary 1.3

When there is risk free asset with rate r_f, we establish earlier that we have:

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}} + \mathbf{Bf} + \mathbf{\Psi}W\tilde{\boldsymbol{r}} + \boldsymbol{\epsilon},$$

where $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I - \mathbf{\Psi}W)\mathbf{1}r_f$. And no asymptotic arbitrage implies that $\tilde{\boldsymbol{\alpha}}^n \approx \mathbf{B}^n\boldsymbol{\lambda}^n$.

Re-arrange the above spatial factor model Equation 3 in another way to have

$$\tilde{\boldsymbol{r}} = \underset{\sim}{\boldsymbol{\alpha}} + \mathbf{B}_1\tilde{\boldsymbol{z}}_1 + \mathbf{B}_2\boldsymbol{f}_2 + \mathbf{\Psi}W\tilde{\boldsymbol{r}} + \boldsymbol{\epsilon},$$

where $\underset{\sim}{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I - \mathbf{\Psi}W)\mathbf{1}r_f + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_1)$, for $\mathbf{1}$ being the $n \times 1$ unit vector and 1 being a vector of ones that is conformable. Comparing $\underset{\sim}{\boldsymbol{\alpha}}$ with $\tilde{\boldsymbol{\alpha}}$ from Equation 38, we have

$$\underset{\sim}{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_{\mathbf{1}}). \tag{41}$$

Same as corollary 1, when the first $k1$ factors $\boldsymbol{f}_1$ are traded, and under Assumptions A1-A4, $\boldsymbol{\lambda}_1 = E(\tilde{\boldsymbol{f}}_1) - 1r_f$. together with Equation 41 and $\tilde{\boldsymbol{\alpha}}^n \approx \mathbf{B}^n\boldsymbol{\lambda}^n$, they imply that:

$$\underset{\sim}{\boldsymbol{\alpha}}^n \approx \mathbf{B}^n\boldsymbol{\lambda}^n + \mathbf{B}_1^n 1 r_f - \mathbf{B}_1^n E(\tilde{\boldsymbol{f}}_1). \tag{42}$$

The right hand side of Equation 42 without the superscript n can be written as:

$$\begin{aligned}\mathbf{B}\boldsymbol{\lambda} + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_1) &= \mathbf{B}_1\boldsymbol{\lambda}_1 + \mathbf{B}_2\boldsymbol{\lambda}_2 + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_1) \\ &= \mathbf{B}_1(E(\tilde{\boldsymbol{f}}_1) - 1 r_f) + \mathbf{B}_2\boldsymbol{\lambda}_2 + \mathbf{B}_1 1 r_f - \mathbf{B}_1 E(\tilde{\boldsymbol{f}}_1) \\ &= \mathbf{B}_2\boldsymbol{\lambda}_2.\end{aligned}$$

Combining that with Equation 42, we can get

$$\underset{\sim}{\boldsymbol{\alpha}}^n \approx \mathbf{B}_2^n \boldsymbol{\lambda}_2^n.$$

Q.E.D.

B Identification and Inference of the Heterogeneous Spatial-Temporal Model

Aquaro et al. (2020) studies the conditions under which $\boldsymbol{\theta}_{full,0}$ is identified, and establishes consistency and asymptotic normality of the estimator. Focusing on the baseline model (Equation 11), write the $(n(K+3))$ by 1 vector $\boldsymbol{\theta}_{full} = (\boldsymbol{\psi}^\intercal, \boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_K^\intercal, \boldsymbol{\sigma}^{2\intercal})^\intercal = (\boldsymbol{\psi}^\intercal, \boldsymbol{b}^\intercal, \boldsymbol{\sigma}^{2\intercal})^\intercal$, where $\boldsymbol{b} = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_K^\intercal)^\intercal$ collects all the parameters associated with exogenous variables. To establish the identification and asymptotic result, in addition to Assumption A3 from Section 2.3, the following assumptions are made:

Assumption 5. The error terms $\{\epsilon_{it}, i = 1, \ldots, n; t = 1, \ldots, T\}$ are independently distributed over i and t. For filtration $\mathcal{F}_t = (\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, ...)$, $E(\epsilon_{it} \mid \mathcal{F}_t) = 0$, $E(\epsilon_{it}^2 \mid \mathcal{F}_t) = \sigma_{i0}^2$, for $i = 1, \ldots, n$, so there is no conditional heteroskedasticity. $inf_i \sigma_{i0}^2 > c > 0$ and $sup_i \sigma_{i0}^2 < \bar{\sigma}^2 < \infty$, and $E(\mid \epsilon_{it} \mid^p \mid \mathcal{F}_t) = E((\mid \epsilon_{it} \mid^p) = \bar{\omega}_{ip} < \bar{c}$, for all i and t, where $1 \leq p \leq 4 + \varepsilon$, for some $\varepsilon > 0$.

Assumption 6. (a) $\boldsymbol{x}_t$ are stationary processes, that satisfy the moment condition $sup_{i,t,l} E(\mid x_{it,l} \mid^{2+g}) < \bar{c}$, for some $g > 0$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, $l = 1, \ldots, (K+1)$. (b) $E(\boldsymbol{x}_t \boldsymbol{x}_t') = \Sigma_{xx}$, where entry $\Sigma_{ij} = E(\boldsymbol{x}_{it}\boldsymbol{x}_{jt}')$ exists for all i and j, such as $sup_{i,j}\|\Sigma_{ij}\| < \bar{c}$, and Σ_{ii} is a $k \times k$ non-singular matrix with $inf_i[\lambda_{min}(\Sigma_{ii})] > c > 0$, and $sup_i[\lambda_{min}(\Sigma_{ii})] < \bar{c} < \infty$. (c) $\frac{1}{T}\sum_{t=1}^T \boldsymbol{x}_t \boldsymbol{x}_t' \xrightarrow{a.s} \Sigma_{xx}$ as $T \to \infty$.

To investigate identification, one needs to introduce the following definition

Definition 3. The set $\mathcal{N}_c(\boldsymbol{\sigma}_0^2)$ in the closed neighbourhood of $\boldsymbol{\sigma}_0^2$ if:

$$\mathcal{N}_c(\boldsymbol{\sigma}_0^2) = \{\boldsymbol{\sigma}_0^2 \in \boldsymbol{\Theta}_\sigma, \mid \sigma_{i0}^2/\sigma_i^2 - 1 \mid < c_i, \text{for } i = 1, \ldots, n\},$$

for some $c_i > 0$, where $\boldsymbol{\Theta}_\sigma$ is a compact subset of $\mathcal{R}^n$.

Assumption 7. The $(n(K+3))$ by 1 vector $\boldsymbol{\theta}_{full} = (\boldsymbol{\psi}', \boldsymbol{b}', \boldsymbol{\sigma}^{2\prime})'$ belongs to $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}_{\boldsymbol{\psi}} \times \boldsymbol{\Theta}_{\boldsymbol{b}} \times \mathcal{N}_c(\boldsymbol{\sigma}_0^2)$. $\boldsymbol{\Theta}_{\boldsymbol{\psi}}$ and $\boldsymbol{\Theta}_{\boldsymbol{b}}$ are compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^{n(K+1)}$, respectively, and $\mathcal{N}_c(\boldsymbol{\sigma}_0^2)$ is defined in Definition 3. $\boldsymbol{\Theta}_c$ is a subset of the $(n(K+3))$ dimensional Euclidean space, $\mathbb{R}^{n(K+3)}$, and $\boldsymbol{\theta}_{full,0}$ is an interior point of $\boldsymbol{\Theta}_c$.

The identification results are given by the following proposition:

Proposition 3. Suppose that Assumption A3, A5-A7 hold, consider a heterogeneous coefficient spatial model given by Equation 11 and log-likelihood function given by Equation 12. For fixed n and K, the $(n(K+3))$ dimensional true parameter vector $\boldsymbol{\theta}_{full,0}$ is almost surely locally identified on $\boldsymbol{\Theta}_c$.

The main inference results are given by the following proposition:

Proposition 4. Suppose that Assumption A3, A5-A7 hold, consider a heterogeneous coefficient spatial model given by Equation 11 and log-likelihood function given by Equation 12. For fixed n and K, the $(n(K+3))$ dimensional QML estimator of $\boldsymbol{\theta}_{full,0}$ is denoted as $\hat{\boldsymbol{\theta}}_{full,QMLE}$, which is almost surely locally consistent for $\boldsymbol{\theta}_{full,0}$ on $\boldsymbol{\Theta}_c$, and has the following asymptotic distribution:

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{full,QMLE} - \boldsymbol{\theta}_{full,0}) \Longrightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{\theta_{full}}), \tag{43}$$

where $\boldsymbol{V}_{\theta_{full}}$ is the asymptotic covariance matrix, which has a standard sandwich form:

$$\boldsymbol{V}_{\theta_{full}} = \boldsymbol{H}^{-1}(\boldsymbol{\theta}_{full,0})\boldsymbol{J}(\boldsymbol{\theta}_{full,0}, \gamma)\boldsymbol{H}^{-1}(\boldsymbol{\theta}_{full,0}), \tag{44}$$

where $\boldsymbol{H}(\boldsymbol{\theta}_{full,0}) = \lim_{T\to\infty} E_0(-\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta}_{full})}{\partial \boldsymbol{\theta}_{full}\partial \boldsymbol{\theta}'_{full}})$ is the Hessian, and $\boldsymbol{J}(\boldsymbol{\theta}_{full,0}, \gamma)$ is the asymptotic variance of the score, which depends on the distribution of the errors. In the case of Gaussian errors, $\gamma = 2$, and $\boldsymbol{H}(\boldsymbol{\theta}_{full,0}) = \boldsymbol{J}(\boldsymbol{\theta}_{full,0}, 2)$.

C Tests of Asset Pricing Restrictions

C.1 Derivation of the Jacobian

This subsection show the details of deriving the Jacobian $\mathbf{J}$ and L_B, which is the derivative of matrix function $F(\mathbf{B}) = M_{\mathbf{B}} = I - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$.

For simplicity and exposition, for the following proof, drop all notations ($*$ and non-traded factor loadings subscript) in the reduced form model. To be specific, we replace $\boldsymbol{\theta}_s, \mathbf{B}_2^*$, and $\boldsymbol{\alpha}^*$ from the main text with $\boldsymbol{\theta}, \mathbf{B}$, and $\boldsymbol{\alpha}$, respectively. Then the testing problem from the main text (Equation 21) can be written as:

$$\boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}, \quad \boldsymbol{\theta} = (\boldsymbol{\alpha}', vec(\mathbf{B})')',$$

$$\boldsymbol{g}(\boldsymbol{\theta}) = M_{\mathbf{B}}\boldsymbol{\alpha}, M_{\mathbf{B}} = I - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'.$$

Our goal is to derive the Jacobian matrix

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}.$$

Having partitioned the vector $\boldsymbol{\theta}' = (\boldsymbol{\alpha}', vec(\mathbf{B})')$, the Jacobian can be partitioned in the same way, and we write

$$\mathbf{J}(\boldsymbol{\theta}) = (J_\alpha, J_{vec(B)}),$$

where J_α contains partial derivative with respect to $\boldsymbol{\alpha}$, and $J_{vec(B)}$ contains the partial derivative with respect to $vec(\mathbf{B})$. It can be easily seen that $J_\alpha = M_{\mathbf{B}}$, and our main task is to derive $J_{vec(B)}$. By the chain rule,

$$J_{vec(B)} = \frac{\partial \boldsymbol{g}}{\partial vec(M_{\mathbf{B}})'}\frac{\partial vec(M_{\mathbf{B}})}{\partial vec(\mathbf{B})'}.$$

To obtain $\frac{\partial \boldsymbol{g}}{\partial vec(M_{\mathbf{B}})'}$, holding $\boldsymbol{\alpha}$ as fixed,

$$d\boldsymbol{g} = (dM_{\mathbf{B}})\boldsymbol{\alpha} = (\boldsymbol{\alpha}' \otimes I_n)d(vec(M_{\mathbf{B}})),$$

and thus $\frac{\partial \boldsymbol{g}}{\partial vec(M_{\mathbf{B}})'} = \boldsymbol{\alpha}' \otimes I_n$. Denote $\frac{\partial vec(M_{\mathbf{B}})}{\partial vec(\mathbf{B})'}$ as L_B. From Magnus and Neudecker (2019), L_B is the derivative of matrix function $F(\mathbf{B}) = I - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$, and we have

$$d(vec(M_{\mathbf{B}})) = L_B d(vec(\mathbf{B})).$$

Given $F(\mathbf{B})$,

$$\begin{aligned}
dF(\mathbf{B}) &= -d(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}') \\
&= -(d\mathbf{B})[(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] - \mathbf{B}d[(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] \\
&= -(d\mathbf{B})[(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] - \mathbf{B}[(d(\mathbf{B}'\mathbf{B})^{-1})\mathbf{B}' + (\mathbf{B}'\mathbf{B})^{-1}d\mathbf{B}'] \\
&= -(d\mathbf{B})[(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] - \mathbf{B}[-(\mathbf{B}'\mathbf{B})^{-1}[(d\mathbf{B})'\mathbf{B} + \mathbf{B}'d\mathbf{B}](\mathbf{B}'\mathbf{B})^{-1}]\mathbf{B}' - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}(d\mathbf{B})'.
\end{aligned}$$

Denote $d(vec(\mathbf{B})) = vec(d\mathbf{B}) = \delta_B$,

$$\begin{aligned}
& d(vec(M_{\mathbf{B}})) = d(vec(F)) = vec(dF) \\
& = -(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes I_n)\delta_B - (I_n \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1})\mathbb{K}_{n,k2}\delta_B \\
& + [\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}]\mathbb{K}_{n,k2}\delta_B + [\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}']\delta_B \\
& = L_B\delta_B,
\end{aligned}$$

where

$$\begin{aligned}
L_B &= -(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes I_n) - (I_n \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1})\mathbb{K}_{n,k2} \\
&\quad + [\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}]\mathbb{K}_{n,k2} + [\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] \\
&\quad = -(I_{n^2} + \mathbb{K}_{n^2})((\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes I_n) - (\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')) \\
&\quad = -(I_{n^2} + \mathbb{K}_{n^2})(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes (I_n - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')) \\
&\quad = -(I_{n^2} + \mathbb{K}_{n^2})(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes M_{\mathbf{B}}).
\end{aligned}$$

And by the property of commutator matrix,

$$\begin{aligned}
L_B &= -(I_{n^2} + \mathbb{K}_{n^2})(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes M_{\mathbf{B}}) \\
&= -(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \otimes M_{\mathbf{B}}) - (M_{\mathbf{B}} \otimes \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1})\mathbb{K}_{n,k2},
\end{aligned}$$

which is computationally attractive when n is large.

C.2 Alternative Method of Deriving the Generalized Inverse of Ω_g

To implement the Wald statistic one needs to find a suitable generalized inverse of the singular $n \times n$ covariance matrix $\Omega_g = \mathbf{J}(\boldsymbol{\theta}_s)V_{\theta_s}\mathbf{J}(\boldsymbol{\theta}_s)^{\intercal}$, where $\mathbf{J}(\boldsymbol{\theta}_s) = \partial \boldsymbol{g}(\boldsymbol{\theta}_s)/\partial \boldsymbol{\theta}_s^{\intercal}$ is the $n \times n(k_2+1)$ Jacobian matrix, as outlined in Andrews (1987). We present one approach to that in the main text, and we present an alternative approach here. This second approach is based on the singular value decomposition of $\mathbf{J}$, namely suppose that

$$\mathbf{J}^{\intercal} = U_* D U^{\intercal} = (U_{*1}, U_{*2}) \begin{pmatrix} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^{\intercal} \\ U_2^{\intercal} \end{pmatrix}.$$

Then

$$\begin{aligned}
\mathbf{J}V_{\theta_s}\mathbf{J}^\intercal &= (U_1,U_2)\begin{pmatrix} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_{*1}^\intercal \\ U_{*2}^\intercal \end{pmatrix} V_{\theta_s}\,(U_{*1},U_{*2})\begin{pmatrix} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_1^\intercal \\ U_2^\intercal \end{pmatrix} \\
&= (U_1,U_2)\begin{pmatrix} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_{*1}^\intercal V_{\theta_s} U_{*1} & U_{*1}^\intercal V_{\theta_s} U_{*2} \\ U_{*2}^\intercal V_{\theta_s} U_{*1} & U_{*2}^\intercal V_{\theta_s} U_{*2} \end{pmatrix}\begin{pmatrix} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_1^\intercal \\ U_2^\intercal \end{pmatrix} \\
&= (U_1,U_2)\begin{pmatrix} D_{n-k_2}U_{*1}^\intercal V_{\theta_s} U_{*1} D_{n-k_2} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_1^\intercal \\ U_2^\intercal \end{pmatrix}.
\end{aligned}$$

Therefore

$$(\mathbf{J}V_{\theta_s}\mathbf{J}^\intercal)^+ = U_1 D_{n-k_2}^{-1}\left(U_{*1}^\intercal V_{\theta_s} U_{*1}\right)^{-1} D_{n-k_2}^{-1} U_1^\intercal.$$

Since $\left(U_*^\intercal V_{\theta_s} U_*\right)^{-1} = U_*^\intercal V_{\theta_s}^{-1} U_*$, the main issue is to express $\left(U_{*1}^\intercal V_{\theta_s} U_{*1}\right)^{-1}$ in terms of

$$\begin{pmatrix} U_{*1}^\intercal V_{\theta_s} U_{*1} & U_{*1}^\intercal V_{\theta_s} U_{*2} \\ U_{*2}^\intercal V_{\theta_s} U_{*1} & U_{*2}^\intercal V_{\theta_s} U_{*2} \end{pmatrix}^{-1} = U_*^\intercal V_{\theta_s}^{-1} U_*.$$

In addition, we have

$$\widehat{V}_{\theta^*}^{-1} = \left(\frac{X^\intercal X}{T}\right) \otimes \widehat{\Sigma}_{\epsilon^*}^{-1}, \quad \hat{\Sigma}_{\epsilon^*}^{-1} = (I_n - \widehat{\mathbf{\Psi}}W)^\intercal D_{\hat{\boldsymbol{\sigma}}}^{-1}(I_n - \widehat{\mathbf{\Psi}}W).$$

Juárez-Ruiz et al. (2016) propose an algorithm for obtaining the inverse of a submatrix in terms of the inverse of the matrix. Therefore, it is possible to express $(\mathbf{J}V_{\theta_s}\mathbf{J}^\intercal)^+$ in terms of $V_{\theta_s}^{-1}$ and the singular value decomposition of $\mathbf{J}$ via a finite sequence of steps.

C.3 Proof of Proposition 1 and Proposition 2

C.3.1 Proposition 1

Define the limits

$$\boldsymbol{\mu}_z = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(\mathbf{z}_t),\ \Sigma_z = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left((\mathbf{z}_t - \boldsymbol{\mu}_z)(\mathbf{z}_t - \boldsymbol{\mu}_z)^\intercal\right).$$

This certainly holds under Assumption A6. We also have, as $T \to \infty$,

$$\sqrt{T}(\hat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}^*) \Longrightarrow \mathcal{N}(\mathbf{0}, (1 + \boldsymbol{\mu}_z^\intercal \Sigma_z^{-1}\boldsymbol{\mu}_z)\Sigma_{\epsilon^*}). \tag{45}$$

Recall that $\Sigma_{\epsilon^*} = G(\mathbf{\Psi})D_{\boldsymbol{\sigma}}G(\mathbf{\Psi})^\intercal$, where $\mathbf{\Psi} = \text{diag}(\boldsymbol{\psi})$. Here, $\boldsymbol{\psi}$ and $\boldsymbol{\sigma}$ are parameters from the structural spatial factor model (Equation 15). Aquaro et al. (2020) established that under Assumptions A3, and A5-A7 from Appendix B, for fixed n and K, the $(n(K+3))$ dimensional QML estimator of $\boldsymbol{\theta}_{full,0}$, denoted $\hat{\boldsymbol{\theta}}_{full,QMLE}$, is almost surely locally consistent for $\boldsymbol{\theta}_{full,0}$ on $\mathbf{\Theta}_c$. In particular, $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\sigma}}$ are consistent estimators. It follows that as $T \to \infty$

$$\hat{\Sigma}_{\epsilon^*} = G(\hat{\mathbf{\Psi}})D_{\hat{\boldsymbol{\sigma}}}G(\hat{\mathbf{\Psi}})^\intercal \xrightarrow{p} \Sigma_{\epsilon^*}.$$

Combining the above equation with Equation 45, under the null that $\boldsymbol{\alpha}^* = \mathbf{0}$, we have by Slutsky's theorem

$$W = T\frac{\hat{\boldsymbol{\alpha}}^{*\intercal}\hat{\Sigma}_{\epsilon^*}^{-1}\hat{\boldsymbol{\alpha}}^*}{1 + \hat{\boldsymbol{\mu}}_z^\intercal \hat{\Sigma}_z^{-1}\hat{\boldsymbol{\mu}}_z} = T\frac{\hat{\boldsymbol{\alpha}}^{*\intercal}\Sigma_{\epsilon^*}^{-1}\hat{\boldsymbol{\alpha}}^*}{1+\boldsymbol{\mu}_z^\intercal\Sigma_z^{-1}\boldsymbol{\mu}_z} + o_P(1) \Longrightarrow \chi^2(n).$$

Q.E.D.

C.3.2 Proposition 2

Equation 22 implies that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s) \Longrightarrow \mathcal{N}(\mathbf{0}, V_{\theta_s}), \tag{46}$$

where V_{θ_s} is the submatrix of V_θ^* that corresponds to the $\boldsymbol{\theta}_s$ components. By the delta method,

$$\sqrt{T}\boldsymbol{g}(\hat{\boldsymbol{\theta}}_s) \Longrightarrow \mathcal{N}(\mathbf{0}, \Omega_g) \;\text{ under the null,} \tag{47}$$

where $\Omega_g = \mathbf{J}(\boldsymbol{\theta}_s)V_{\theta_s}\mathbf{J}(\boldsymbol{\theta}_s)^\intercal$ and $\mathbf{J}(\boldsymbol{\theta}_s) = \partial \boldsymbol{g}(\boldsymbol{\theta}_s)/\partial \boldsymbol{\theta}_s^\intercal$ is the Jacobian matrix. Given consistent QML estimator $\hat{\Theta}^* = (\hat{\boldsymbol{\alpha}}^{*\intercal}, \text{vec}(\hat{\mathbf{B}}_1^*)^\intercal, \text{vec}(\hat{\mathbf{B}}_2^*))^\intercal) = R^\intercal X(X^\intercal X)^{-1}$, we have

$$\hat{\mathbf{J}} = (M_{\widehat{\mathbf{B}}_2^*}, -(\hat{\boldsymbol{\alpha}}^{*\intercal}\widehat{\mathbf{B}}_2^*(\widehat{\mathbf{B}}_2^{*\intercal}\widehat{\mathbf{B}}_2^*)^{-1} \otimes M_{\widehat{\mathbf{B}}_2^*})) + (M_{\widehat{\mathbf{B}}_2^*}, -(\hat{\boldsymbol{\alpha}}^{*\intercal}M_{\widehat{\mathbf{B}}_2^*} \otimes \widehat{\mathbf{B}}_2^*(\widehat{\mathbf{B}}_2^{*\intercal}\widehat{\mathbf{B}}_2^*)^{-1})\mathbb{K}_{n,k_2}) \xrightarrow{p} \mathbf{J}.$$

In addition, we have

$$\widehat{V}_{\theta^*} = \left(\frac{X^\intercal X}{T}\right) \otimes \widehat{\Sigma}_{\epsilon^*} \xrightarrow{p} V_{\theta^*} \;\text{ and thus } \widehat{V}_{\theta_s} \xrightarrow{p} V_{\theta_s}.$$

Combining the above two equations gives

$$\hat{\Omega}_g = \hat{\mathbf{J}}\widehat{V}_{\theta_s}\hat{\mathbf{J}}^\intercal \xrightarrow{p} \mathbf{J}V_{\theta_s}\mathbf{J}^\intercal = \Omega_g.$$

Given a suitable generalized inverse $\hat{\Omega}_g^+$ that we have discussed in details in the main body of the paper, and in particular, with $\Pr\Big(\text{rank}(\hat{\Omega}_g) = \text{rank}(\Omega_g) = n - k_2\Big) \to 1$ as $T \to \infty$, using the result of Andrews (1987)

$$W = T\boldsymbol{g}(\hat{\boldsymbol{\theta}}_s)^\intercal\hat{\Omega}_g^+\boldsymbol{g}(\hat{\boldsymbol{\theta}}_s) \Longrightarrow \chi^2(n - k_2).$$

In fact, since $\hat{\Omega}_g \xrightarrow{p} \Omega_g$, the eigenvectors satisfy $\widehat{Q} \xrightarrow{p} Q$ and the eigenvalues satisfy $\widehat{\lambda}_j \xrightarrow{p} \lambda_j$, where $\lambda_j > 0$ for $j = 1, \ldots, n - k_2$ and $\lambda_j = 0$ for $j > n - k_2$. Therefore, for any $\varepsilon > 0$ we have

$$\lim_{T\to\infty} \Pr\left(\max_{1\le j\le n-k_2} \left|\widetilde{\lambda}_j^{-1} - \lambda_j^{-1}\right| > \varepsilon\right) = 0$$

and so $\widehat{Q}\widetilde{\Lambda}^+\widehat{Q}^\intercal \xrightarrow{p} Q\Lambda^+Q^\intercal$, where $\widetilde{\Lambda}^+ = \text{diag}(\widetilde{\lambda}_1^{-1}, \ldots, \widetilde{\lambda}_{n-k_2}^{-1}, 0, \ldots, 0)$. By construction the matrix $\widehat{Q}\widetilde{\Lambda}\widehat{Q}^\intercal$, with $\widetilde{\Lambda} = \text{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{n-k_2}, 0, \ldots, 0)$, has rank $n - k_2$. Therefore, the result follows. Q.E.D.

D Supplementary Figures and Tables

Number of unique news stories	88,316,898
Number of stories remaining after removing topics including analyst recommendations, ratings changes, and index movements	87,841,641
Of these:	
Number of stories tag sample companies	8,341,848
Of these:	
Number of stories that mention only one company	5,507,772 (66.03%)
Number of stories that mention exactly two companies	1,637,256 (19.63%)
Number of stories that mention more than two companies	1,196,820 (14.34%)

Table 14: Descriptive statistics for RavenPack Equity files Dow Jones Edition for the period January 2004 to December 2015.

Number of yearly window a pair gets identified	Frequency	Percentage	Cumulative Percentage
0	217024	72.80%	72.80%
1	40178	13.48%	86.28%
2	13302	4.46%	90.74%
3	7116	2.39%	93.13%
4	4522	1.52%	94.65%
5	3236	1.09%	95.74%
6	2506	0.84%	96.58%
7	2022	0.68%	97.26%
8	1804	0.61%	97.87%
9	1508	0.51%	98.38%
10	1350	0.45%	98.83%
11	1232	0.41%	99.24%
12	2316	0.78%	100%

Table 15: Frequency distribution table of the number of yearly link identification windows that a pair gets identified as economic neighbors for all possible pairs (i, j) in our sample. Note: A pair identified in k yearly windows could get multiple co-mentions within each window.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1840	81	256	777	315	529	273	573	568	116	235
	0.33	0.01	0.05	0.14	0.06	0.10	0.05	0.10	0.10	0.02	0.04
Durbs	81	12	14	67	16	72	27	49	45	10	13
	0.20	0.03	0.03	0.17	0.04	0.18	0.07	0.12	0.11	0.02	0.03
Energy	256	14	372	147	42	115	51	153	83	20	172
	0.18	0.01	0.26	0.10	0.03	0.08	0.04	0.11	0.06	0.01	0.12
Hi-tec	777	67	147	1376	227	419	182	439	403	126	86
	0.18	0.02	0.03	0.32	0.05	0.10	0.04	0.10	0.09	0.03	0.02
Health	315	16	42	227	370	111	71	134	143	28	19
	0.21	0.01	0.03	0.15	0.25	0.08	0.05	0.09	0.10	0.02	0.01
Manuf	529	72	115	419	111	470	134	287	211	43	62
	0.220	0.03	0.05	0.17	0.05	0.19	0.05	0.12	0.09	0.02	0.03
Nondur	273	27	51	182	71	134	196	152	244	42	25
	0.20	0.02	0.04	0.13	0.05	0.10	0.14	0.11	0.17	0.03	0.02
Other	573	49	153	439	134	287	152	344	295	63	138
	0.22	0.02	0.06	0.17	0.05	0.11	0.06	0.13	0.11	0.02	0.05
Shops	568	45	83	403	143	211	244	295	698	73	40
	0.20	0.02	0.03	0.14	0.05	0.08	0.09	0.11	0.25	0.03	0.01
Telcm	116	10	20	126	28	43	42	63	73	18	22
	0.21	0.02	0.04	0.22	0.05	0.08	0.07	0.11	0.13	0.03	0.04
Utilities	235	13	172	86	19	62	25	138	40	22	366
	0.20	0.01	0.15	0.07	0.02	0.05	0.02	0.12	0.03	0.02	0.31

Table 16: **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 1$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs identified, and the second row gives the proportion to total links firms in that industry have.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1496	65	193	566	233	377	193	451	397	84	173
	0.35	0.02	0.05	0.13	0.06	0.09	0.05	0.11	0.09	0.02	0.04
Durbs	65	12	8	41	8	43	14	36	29	8	7
	0.24	0.04	0.03	0.15	0.03	0.16	0.05	0.13	0.11	0.03	0.03
Energy	193	8	294	87	18	61	29	95	44	11	110
	0.20	0.01	0.31	0.09	0.02	0.06	0.03	0.10	0.05	0.01	0.12
Hi-tec	566	41	87	1040	123	254	103	311	254	85	40
	0.19	0.01	0.03	0.36	0.04	0.09	0.04	0.11	0.09	0.03	0.01
Health	233	8	18	123	288	64	40	86	75	17	4
	0.24	0.01	0.02	0.13	0.30	0.07	0.04	0.09	0.08	0.02	0
Manuf	377	43	61	254	64	264	84	205	109	26	20
	0.25	0.03	0.04	0.17	0.04	0.18	0.06	0.14	0.07	0.02	0.01
Nondur	193	14	29	103	40	84	144	97	158	30	12
	0.21	0.02	0.03	0.11	0.04	0.09	0.16	0.11	0.17	0.03	0.01
Other	451	36	95	311	86	205	97	256	177	52	84
	0.24	0.02	0.05	0.17	0.05	0.11	0.05	0.14	0.10	0.03	0.05
Shops	397	29	44	254	75	109	158	177	536	48	19
	0.22	0.02	0.02	0.14	0.04	0.06	0.09	0.10	0.29	0.03	0.01
Telcm	84	8	11	85	17	26	30	52	48	18	13
	0.21	0.02	0.03	0.22	0.04	0.07	0.08	0.13	0.12	0.05	0.03
Utilities	173	7	110	40	4	20	12	84	19	13	290
	0.22	0.01	0.14	0.05	0.01	0.03	0.02	0.11	0.02	0.02	0.38

Table 17: **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 2$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs identified, and the second row gives the proportion to total links firms in that industry have.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1250	50	153	415	187	289	160	380	315	61	136
	0.37	0.01	0.05	0.12	0.06	0.09	0.05	0.11	0.09	0.02	0.04
Durbs	50	8	7	29	5	31	10	30	20	5	4
	0.25	0.04	0.04	0.15	0.03	0.16	0.05	0.15	0.10	0.03	0.02
Energy	153	7	246	54	11	42	19	72	22	8	60
	0.22	0.01	0.35	0.08	0.02	0.06	0.03	0.10	0.03	0.01	0.09
Hi-tec	415	29	54	832	82	172	63	235	164	73	23
	0.19	0.01	0.03	0.39	0.04	0.08	0.03	0.11	0.08	0.03	0.01
Health	187	5	11	82	246	44	26	67	44	9	2
	0.26	0.01	0.02	0.11	0.34	0.06	0.04	0.09	0.06	0.01	0
Manuf	289	31	42	172	44	186	55	156	62	16	10
	0.27	0.03	0.04	0.16	0.04	0.17	0.05	0.15	0.06	0.02	0.01
Nondur	160	10	19	63	26	55	112	75	114	21	5
	0.24	0.02	0.03	0.10	0.04	0.08	0.17	0.11	0.17	0.03	0.01
Other	380	30	72	235	67	156	75	210	126	30	63
	0.26	0.02	0.05	0.16	0.05	0.11	0.05	0.15	0.09	0.02	0.04
Shops	315	20	22	164	44	62	114	126	394	32	10
	0.24	0.02	0.02	0.13	0.03	0.05	0.09	0.10	0.30	0.02	0.01
Telcm	61	5	8	73	9	16	21	30	32	16	8
	0.22	0.02	0.03	0.26	0.03	0.06	0.08	0.11	0.11	0.06	0.03
Utilities	136	4	60	23	2	10	5	63	10	8	214
	0.25	0.01	0.11	0.04	0	0.02	0.01	0.12	0.02	0.01	0.40

Table 18: **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 3$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs identified, and the second row gives the proportion to total links firms in that industry have.

	(1) Factor Component								(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ
(1) Spatial CAPM									
Mean Group Estimates	0.015	0.597							0.417
	(0.001)	(0.021)							(0.019)
% significant (at 5% level)	3.1%	79.4%							73.3%
N_p	394	394							387
(2) Spatial Factor Model (with FF3)									
Mean Group Estimates	0.013	0.569	0.127	-0.131					0.452
	(0.001)	(0.021)	(0.014)	(0.022)					(0.018)
% significant (at 5% level)	1.7%	74.4%	51.0%	69.0%					75.1%
N_p	394	394	394	394					387
(3) Spatial Factor Model (with FF5)									
Mean Group Estimates	0.013	0.584	0.142	-0.130	0.141	0.173			0.455
	(0.001)	(0.021)	(0.014)	(0.023)	(0.022)	(0.021)			(0.018)
% significant (at 5% level)	2.0%	75.1%	51.5%	68.0%	50.3%	47.5%			76.4%
N_p	394	394	394	394	394	394			387
(4) Spatial Factor Model (with FF5+MOM)									
Mean Group Estimates	0.013	0.591	0.143	-0.140	0.139	0.181	-0.022		0.447
	(0.001)	(0.021)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)		(0.019)
% significant (at 5% level)	1.4%	75.4%	51.3%	62.4%	49.7%	47.5%	30.5%		74.9%
N_p	394	394	394	394	394	394	394		387
(5) Spatial Factor Model (with FF5+MOM+MA)									
Mean Group Estimates	0.014	0.594	0.104	-0.097	0.120	0.146	-0.034	-0.182	0.450
	(0.001)	(0.021)	(0.013)	(0.019)	(0.021)	(0.022)	(0.007)	(0.023)	(0.019)
% significant (at 5% level)	2.5%	75.6%	41.6%	53.8%	45.7%	44.9%	26.9%	41.4%	74.4%
N_p	394	394	394	394	394	394	394	394	387

Table 19: **QML estimation results of Equation 29 to Equation 33 using full sample.** Note: threshold $m = 2$. For each panel, the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level (with multiple testing correction), and the last row gives the number of unrestricted units N_p for each parameter.

	(1) Factor Component								(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ
(1) Spatial CAPM									
Mean Group Estimates	0.015	0.628							0.392
	(0.001)	(0.021)							(0.019)
% significant (at 5% level)	2.8%	77.9%							73.6%
N_p	394	394							384
(2) Spatial Factor Model (with FF3)									
Mean Group Estimates	0.014	0.605	0.126	-0.120					0.422
	(0.001)	(0.020)	(0.014)	(0.022)					(0.018)
% significant (at 5% level)	1.5%	71.1%	51.3%	69.3%					76.4%
N_p	394	394	394	394					384
(3) Spatial Factor Model (with FF5)									
Mean Group Estimates	0.013	0.619	0.141	-0.119	0.140	0.170			0.424
	(0.001)	(0.020)	(0.014)	(0.023)	(0.022)	(0.021)			(0.018)
% significant (at 5% level)	1.7%	73.9%	51.5%	68.0%	50.3%	46.7%			76.1%
N_p	394	394	394	394	394	394			384
(4) Spatial Factor Model (with FF5+MOM)									
Mean Group Estimates	0.013	0.627	0.143	-0.133	0.138	0.179	-0.027		0.416
	(0.001)	(0.021)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)		(0.018)
% significant (at 5% level)	1.4%	73.1%	50.8%	61.9%	49.5%	47.2%	30.2%		74.4%
N_p	394	394	394	394	394	394	394		384
(5) Spatial Factor Model (with FF5+MOM+MA)									
Mean Group Estimates	0.015	0.631	0.104	-0.090	0.120	0.146	-0.038	-0.176	0.417
	(0.001)	(0.021)	(0.013)	(0.019)	(0.021)	(0.022)	(0.007)	(0.023)	(0.018)
% significant (at 5% level)	1.9%	72.6%	41.9%	54.8%	45.9%	44.9%	27.7%	41.4%	74.9%
N_p	394	394	394	394	394	394	394	394	384

Table 20: **QML estimation results of Equation 29 to Equation 33 using full sample.** Note: threshold $m = 3$. For each panel, the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level (with multiple testing correction), and the last row gives the number of unrestricted units N_p for each parameter.

	(1) Factor Component								(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ
(1)Sample 1: 04-07									
MG Estimates	0.024	0.507	0.100	-0.069	-0.004	0.067	0.086	-0.192	0.482
	(0.003)	(0.024)	(0.016)	(0.025)	(0.027)	(0.027)	(0.020)	(0.021)	(0.021)
% sig (at 5%)	5.9%	43.4%	12.2%	20.6%	13.2%	13.7%	25.4%	8.4%	51.3%
N_p	394	394	394	394	394	394	394	394	386
(1)Sample 2: 08-11									
MG Estimates	0.014	0.599	0.106	-0.093	0.124	0.100	-0.043	-0.126	0.460
	(0.002)	(0.023)	(0.016)	(0.020)	(0.024)	(0.032)	(0.009)	(0.025)	(0.020)
% sig (at 5%)	0.0%	63.5%	27.7%	23.1%	21.6%	30.2%	17.5%	16.5%	55.1%
N_p	394	394	394	394	394	394	394	394	385
(1)Sample 3: 12-15									
MG Estimates	0.003	0.550	0.047	-0.122	0.085	0.160	-0.068	-0.214	0.470
	(0.002)	(0.024)	(0.013)	(0.021)	(0.019)	(0.026)	(0.017)	(0.024)	(0.021)
% sig (at 5%)	2.8%	50.8%	12.7%	22.8%	10.9%	17.5%	24.9%	19.8%	55.8%
N_p	394	394	394	394	394	394	394	394	387

Table 21: **QML estimation results of the Equation 33 using sub-samples.** Note: threshold $m = 1$. Each panel gives the results for Equation 33 estimated using a four-year sub-sample. With each panel, the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level (with multiple testing correction), and the last row gives the number of unrestricted units N_p for each parameter.

	(1) Factor Component									(2) Spatial Component
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_I	ψ
Panel A: Finance										
MG Estimates	0.007	0.357	0.017	0.293	0.065	-0.002	-0.050	0.266	0.229	0.552
	(0.003)	(0.062)	(0.051)	(0.050)	(0.057)	(0.048)	(0.018)	(0.056)	(0.057)	(0.048)
% Sig(at 5%)	2.9%	69.1%	44.1%	48.5%	38.2%	25.0%	26.5%	38.2%	36.8%	76.5%
N_p	68	68	68	68	68	68	68	68	68	68
Panel B: Consumer										
MG Estimates	0.001	0.097	-0.215	-0.199	0.232	0.295	0.008	-0.141	0.551	0.376
	(0.003)	(0.054)	(0.025)	(0.022)	(0.024)	(0.028)	(0.012)	(0.026)	(0.052)	(0.038)
% Sig(at 5%)	1.3%	57.3%	44.0%	52.0%	42.7%	42.7%	13.3%	17.3%	78.7%	72.0%
N_p	75	75	75	75	75	75	75	75	75	75
Panel C: Health										
MG Estimates	0.009	0.288	-0.215	-0.250	0.105	0.213	0.067	-0.096	0.347	0.315
	(0.007)	(0.055)	(0.036)	(0.037)	(0.041)	(0.046)	(0.022)	(0.040)	(0.034)	(0.061)
% Sig(at 5%)	3.6%	64.3%	57.1%	53.6%	17.9%	39.3%	32.1%	21.4%	92.9%	71.4%
N_p	28	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech										
MG Estimates	0.017	0.627	0.118	-0.370	-0.311	0.238	-0.043	-0.139	-0.005	0.420
	(0.004)	(0.042)	(0.038)	(0.028)	(0.049)	(0.042)	(0.012)	(0.026)	(0.034)	(0.041)
% Sig(at 5%)	4.3%	74.3%	37.1%	67.1%	41.4%	37.1%	17.1%	14.3%	34.3%	68.1%
N_p	70	70	70	70	70	70	70	70	70	69
Panel E: Manufacturing										
MG Estimates	-0.002	0.017	-0.269	-0.137	0.240	0.146	0.040	-0.221	0.690	0.417
	(0.003)	(0.045)	(0.028)	(0.018)	(0.020)	(0.032)	(0.011)	(0.025)	(0.046)	(0.036)
% Sig(at 5%)	3.5%	57.5%	65.5%	30.1%	41.6%	46.9%	25.7%	36.3%	87.6%	74.8%
N_p	113	113	113	113	113	113	113	113	113	107
Panel F: Others										
MG Estimates	0.002	0.357	0.016	-0.170	0.218	0.204	-0.038	-0.269	0.306	0.445
	(0.005)	(0.082)	(0.045)	(0.046)	(0.044)	(0.056)	(0.020)	(0.050)	(0.049)	(0.063)
% Sig(at 5%)	0.0%	55.0%	17.5%	50.0%	27.5%	45.0%	20.0%	42.5%	47.5%	75.0%
N_p	40	40	40	40	40	40	40	40	40	40

Table 22: **QML estimation results of spatial factor model with Fama-French five factors plus Momentum, and Media-Attention factors, and industry factor (Equation 35). Parameters summarized by industry.**

	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_I	ψ
Panel A: Finance														
MG	0.007	0.361	0.018	0.292	0.063	0.000	-0.050	0.269	-0.018	0.010	-0.300	-2.722	0.229	0.550
	(0.003)	(0.062)	(0.051)	(0.051)	(0.057)	(0.048)	(0.018)	(0.055)	(0.011)	(0.004)	(0.093)	(3.138)	(0.057)	(0.048)
%sig	2.9%	69.1%	44.1%	50.0%	38.2%	25.0%	23.5%	38.2%	0.0%	0.0%	1.5%	1.5%	35.3%	76.5%
N_p	68	68	68	68	68	68	68	68	68	68	68	68	68	68
Panel B: Consumer														
MG	0.001	0.091	-0.221	-0.203	0.229	0.300	0.006	-0.135	-0.001	0.006	-0.291	-11.496	0.563	0.374
	(0.003)	(0.054)	(0.025)	(0.022)	(0.024)	(0.028)	(0.012)	(0.026)	(0.010)	(0.003)	(0.069)	(2.493)	(0.053)	(0.038)
%sig	2.7%	56.0%	46.7%	54.7%	42.7%	45.3%	13.3%	16.0%	0.0%	0.0%	0.0%	1.3%	78.7%	70.7%
N_p	75	75	75	75	75	75	75	75	75	75	75	75	75	75
Panel C: Health														
MG	0.009	0.284	-0.219	-0.252	0.112	0.222	0.063	-0.092	-0.037	0.015	-0.214	-5.586	0.356	0.314
	(0.007)	(0.055)	(0.037)	(0.037)	(0.041)	(0.045)	(0.021)	(0.039)	(0.023)	(0.006)	(0.093)	(4.856)	(0.034)	(0.061)
%sig	7.1%	64.3%	60.7%	53.6%	21.4%	39.3%	28.6%	21.4%	0.0%	0.0%	0.0%	0.0%	92.9%	71.4%
N_p	28	28	28	28	28	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech														
MG	0.017	0.630	0.118	-0.369	-0.313	0.237	-0.042	-0.140	-0.033	0.012	-0.129	5.507	-0.006	0.419
	(0.004)	(0.042)	(0.038)	(0.028)	(0.049)	(0.041)	(0.012)	(0.027)	(0.013)	(0.004)	(0.075)	(2.622)	(0.034)	(0.041)
%sig	4.3%	75.7%	37.1%	67.1%	41.4%	35.7%	15.7%	15.7%	0.0%	0.0%	0.0%	0.0%	32.9%	65.2%
N_p	70	70	70	70	70	70	70	70	70	70	70	70	70	69
Panel E: Manufacturing														
MG	-0.002	0.017	-0.269	-0.137	0.240	0.147	0.040	-0.221	-0.027	0.003	-0.145	2.920	0.691	0.417
	(0.003)	(0.046)	(0.028)	(0.019)	(0.020)	(0.031)	(0.011)	(0.025)	(0.007)	(0.002)	(0.055)	(1.676)	(0.046)	(0.036)
%sig	3.5%	58.4%	65.5%	29.2%	41.6%	46.0%	26.5%	36.3%	0.0%	0.0%	0.9%	0.0%	87.6%	73.8%
N_p	113	113	113	113	113	113	113	113	113	113	113	113	113	107
Panel F: Others														
MG	0.001	0.352	0.012	-0.175	0.219	0.211	-0.039	-0.263	-0.029	0.007	-0.333	-5.989	0.316	0.444
	(0.005)	(0.082)	(0.045)	(0.046)	(0.044)	(0.056)	(0.020)	(0.050)	(0.015)	(0.005)	(0.095)	(3.184)	(0.049)	(0.063)
%sig	0.0%	55.0%	17.5%	50.0%	27.5%	45.0%	22.5%	40.0%	0.0%	0.0%	2.5%	0.0%	47.5%	75.0%
N_p	40	40	40	40	40	40	40	40	40	40	40	40	40	40

Table 23: **QML estimation results of spatial factor model with both tradable factors and macro factors, and industry factor (Equation 34). Parameters summarized by industry.**

	(1) Factor Component								(2) Spatial-temporal Component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM														
MG	0.015	0.564							0.446	0.002	-0.008	0.001	-0.003	0.004
	(0.001)	(0.022)							(0.020)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	3.1%	78.4%							76.5%	21.7%	5.9%	2.3%	5.4%	1.3%
N_p	394	394							387	387	387	387	387	387
(2) Spatial Factor Model (with FF3)														
MG	0.013	0.529	0.129	-0.137					0.489	0.002	-0.008	-0.001	-0.001	0.003
	(0.001)	(0.021)	(0.014)	(0.022)					(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	3.1%	74.6%	50.5%	68.8%					78.6%	23.3%	4.7%	2.3%	5.4%	1.3%
N_p	394	394	394	394					387	387	387	387	387	387
(3) Spatial Factor Model (with FF5)														
MG	0.011	0.544	0.144	-0.137	0.140	0.179			0.493	0.007	-0.007	-0.001	-0.002	0.002
	(0.001)	(0.021)	(0.014)	(0.023)	(0.022)	(0.021)			(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	3.1%	75.1%	52.0%	68.5%	50.3%	44.4%			79.6%	22.5%	4.1%	2.8%	4.1%	1.6%
N_p	394	394	394	394	394	394			387	387	387	387	387	387
(4) Spatial Factor Model (with FF5+MOM)														
MG	0.015	0.471	0.124	-0.116	0.074	0.113	-0.010		0.545	0.002	-0.014	0.002	-0.008	0.005
	(0.001)	(0.022)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)		(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	3.1%	75.9%	52.3%	61.9%	50.5%	45.9%	30.5%		78.0%	20.4%	4.1%	3.1%	4.1%	1.6%
N_p	394	394	394	394	394	394	394		387	387	387	387	387	387
(5) Spatial Factor Model (with FF5+MOM+MA)														
MG	0.014	0.552	0.105	-0.102	0.118	0.147	-0.030	-0.182	0.489	0.003	-0.006	0.000	-0.001	0.002
	(0.001)	(0.022)	(0.013)	(0.019)	(0.021)	(0.022)	(0.007)	(0.023)	(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	3.1%	75.6%	43.1%	53.8%	47.0%	44.2%	27.4%	40.4%	77.3%	19.6%	4.4%	3.4%	4.7%	1.6%
N_p	394	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 24: **QML estimation results of heterogeneous coefficients spatial-temporal model using full sample (all factors are traded).** Note: $m = 1$, $L = 5$. For each panel , the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level, and the last row gives the number of unrestricted units N_p for each parameter.

	(1) Factor Component						(2) Spatial-temporal Component					
	α	...	β_8	β_9	β_{10}	β_{11}	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
MG	0.014	...	-0.021	0.001	-0.047	4.230	0.489	0.003	-0.006	0.000	-0.001	0.002
	(0.001)	...	(0.005)	(0.002)	(0.027)	(1.067)	(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	3.1%	...	0.0%	0.0%	0.0%	0.3%	77.3%	17.8%	3.9%	3.4%	4.4%	1.6%
N_p	394	...	394	394	394	394	387	387	387	387	387	387

Table 25: **QML estimation results of heterogeneous coefficients spatial-temporal model using full sample (mixture of tradable and macro factors).** Note: $m = 1$, $L = 5$. For each panel , the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The third row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level, and the last row gives the number of unrestricted units N_p for each parameter. To save space, we only report the estimates for non-tradable factors and spatial-temporal coefficients.

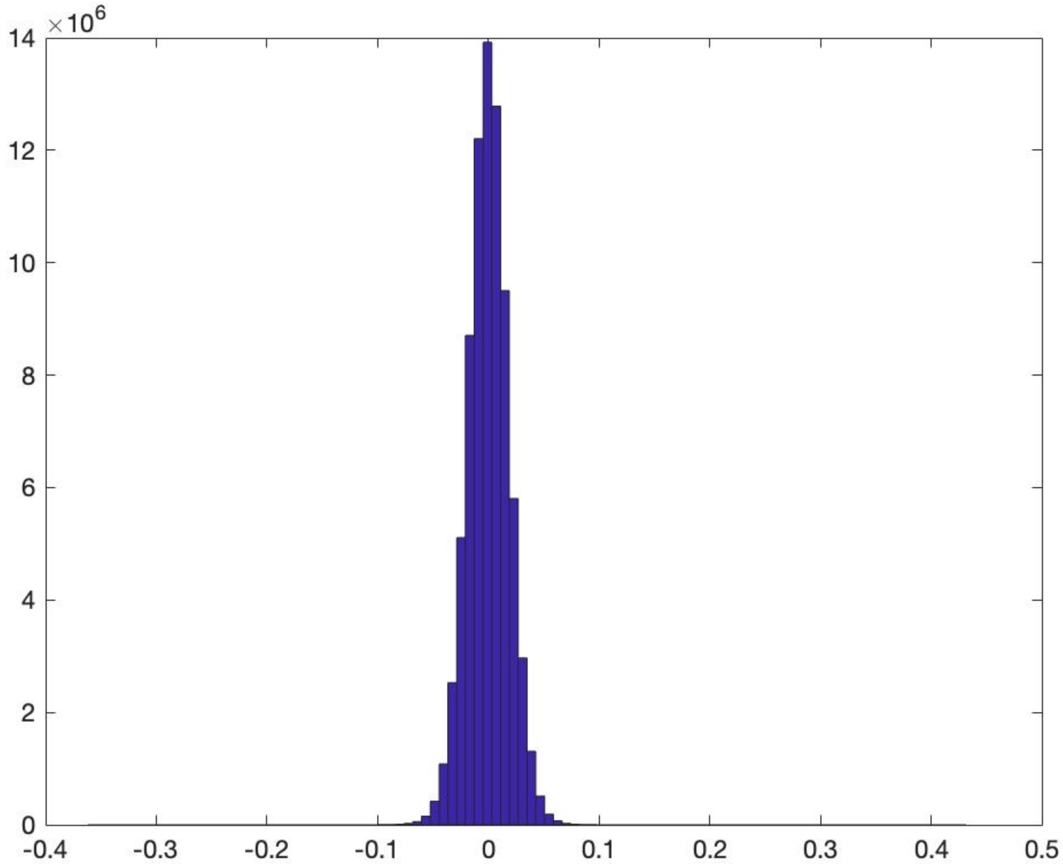

Figure 1: Histogram of bootstrapped $\hat{\rho}_{ij}^{b}$ for all $i \neq j, b = 1, \ldots, 1000$.