Firm Commonality and Inference in Corporate Finance

Jiyuan Huang*

Job Market Paper

November 8, 2023

Abstract

In this paper, I explore latent connections among firms and their implications for empirical work. These connections can be motivated by competition, peer effects, supply chains, or common factors. I introduce a spatial framework that captures these relations in a corporate landscape, using product similarity (Hoberg and Philips, 2016) as a proxy for firm commonality. I find that firm commonality has significant explanatory power of corporate outcomes such as capital expenditure and cash holdings, altering the interpretation of commonly used explanatory variables. Further, omitting firm commonality leads to significantly correlated error terms. I show that the widely used firm-clustered standard errors reject up to 95%, which is dramatically higher than the designed 5%. Finally, I provide a bootstrap solution of standard errors to address the over-rejection problem caused by firm commonality.

*Department of Banking and Finance, University of Zurich; Swiss Finance Institute; Email: jiyuan.huang@bf.uzh.ch

1 Introduction

Firms are not isolated islands; rather they share a web of visible and invisible relationships that link their activities. There is extensive literature documenting connections due to supply chain relationships (e.g. Kelly, Lustig, and Van Nieuwerburgh (2013), Cohen and Frazzini (2008), Menzly and Ozbas (2010)), competition effects (e.g. Campello (2003), Chevalier (1995)), peer effects (e.g. Dougal, Parsons, and Titman (2015), Leary and Roberts (2014), Gao, Ng, and Wang (2011)) and common responses to factors (e.g. Huang and Östberg (2023), DeAngelo and Roll (2015), Korajczyk and Levy (2003)). All these connections result in commonality in widely used corporate outcomes. This paper characterizes these connections using spatial modeling with a novel measure of firm commonality and documents the implications of these connections on variable interpretations and statistical inference.

A canonical model of corporate outcomes focuses on relationships of characteristics within a firm. However, the prevalent commonalities between firms indicate that firm behavior should be understood within the context of a much broader interconnected corporate landscape. To capture firm commonality, I use the product specialty measure of Hoberg and Phillips (2010, 2016) to characterize pairwise distance between all Russell 3000 firms. This measurement assesses the product similarity of firms based on text analysis and firms that operate on comparable business models have higher commonality scores. The commonality score is a desired variable to characterize the corporate landscape, since it encapsulates the similarity of firm outcomes well.

The prevailing existence and the importance of firm commonality mean that ignoring this structure potentially leads to biased estimates and inflated t statistics. To illustrate these problems, I develop spatial models of firms with commonality scores and evaluate the importance of spatial dependence for corporate outcomes such as capital expenditure and cash holdings. It is appropriate to model firm connections with spatial modeling, as the corporate landscape intrinsically parallels the geographic landscapes. Just as lands compose the physical world, firms compose the corporate landscape.

Using the spatial model, I find that firm commonality is highly significant and economically important in all firm outcomes and I show the conventional regression potentially suffers bias from omitting firm commonality. For example, in the regression of capital expenditure, after controlling for firm commonality, the economic magnitude of the impact from the conventionally used variable, firm cash flows, is reduced by 50%. Moreover, the effect of firm commonality is significant and as large as two-thirds of the cash flow impact.

Besides the bias problem, I also find that ignoring firm commonality causes over-rejection in hypothesis testing and this problem still exists even under clustered standard errors. I show that the residuals of fixed effects regressions exhibit strong spatial dependence and this renders the widely used firm-clustered standard errors invalid, with an over-reject rate up to 95%, which is much higher than the designed 5%. I continue to show that two-way clustered standard errors is also not an adequate solution in practice, as it still over-rejects up to sevenfold. To accommodate the commonality among firms, I propose a two-way resampling bootstrap method for standard errors and show it solves the over-rejection problem. The intuition is that two-way resampling preserves both the time series structure within firms and the spatial dependence across firms and thus correctly estimates the standard errors.

A central contribution of this paper is to incorporate spatial modeling in the empirical corporate finance framework. By evaluating firm outcomes in the context of corporate landscape, we can have a more complete picture of firm behaviors. This paper is generally related to studies about firm outcomes under other firms' effects. However, the majority of previous studies build links in an implicit local network [1], while I investigate a much broader corporate landscape with a spatial modeling framework. The broad corporate landscape used here has at least two advantages. On the one hand, contrary to previous studies that treat the connections between firms as a binary relation, I take the intensity of the connections into account since the firm commonality score is continuous. On the other hand, the connections between firms in this paper are not as sparse as in the previous studies, which matches the real-world data better.

The paper also contributes to recent studies that investigate the validity of current empirical methods. I evaluate the impact of firm commonality on corporate outcomes such as capital expenditure and cash holding, where I illustrate the significant explanatory power of firm commonality on firm characteristics and the potential bias in commonly used regressors. Moreover, I document the failure of widely used clustered standard

[1] For example, the commonly used linear-in-mean peer effect model can be expressed as a special case of the spatial model. In this case, the commonality metric is the row standardization of a matrix in which values are binary depending on whether two firms are defined as peers or not.

error in the presence of spatial dependence and propose a bootstrap solution to solve the over-rejection problem.

The rest of the paper is organized as follows: Section 2 reviews the related literature and highlights the differences between their results and my study. Section 3 presents the dataset and the measurement of firm commonality. Section 4 illustrates the relationship between firm commonality and firm characteristics from various perspectives and shows the spatial dependence of corporate outcomes. Section 5 presents empirical examples of spatial models with firm outcomes and illustrates the potential bias in conventional methods. Section 6 documents the failure of clustered standard errors and section 7 provides the bootstrap solution of standard errors. Section 8 draws the conclusion.

2 Related Literature

The is a large body of literature on firm decisions impacted by its peers, such as competitors, suppliers/customers, or firms in proximity. This literature suggests that firm outcomes have commonalities since connected firms are influenced by each other.

Chevalier (1995) investigates the impact of leveraged buyouts (LBOs) on supermarket pricing and finds that when local rivals are highly leveraged, the market price rises following a LBO, while when rivals have low leverages, the price falls. They interpret this difference as an outcome of different competition dynamics and show the characteristics of local rivals play an important role in the outcome. Campello (2003) also finds that a firm's performance in the product market, especially after economic shocks, depends on the relative position of its capital structure, i.e., if the firm is more leveraged compared to its competitors. These studies highlight that the outcomes of a firm are nuanced and depend on factors such as its competitors' characteristics.

Gao, Ng, and Wang (2011) studies firms' financing policies and shows that firms exhibit conformity to their peers in the same metropolitan statistical area (MSA). They highlight this peer influence and conclude it is a significant factor in a firm's decision-making process. Besides financing policies, Dougal, Parsons, and Titman (2015) finds that a firm's investment choices are highly sensitive to the investments of other firms that are headquartered nearby even though they are not in the same industry. In addition to geographical proximity, empirical studies have considered abstract firm peers as well. One typical way of defining peer groups is to use the industry code. Leary and Roberts (2014) uses the idiosyncratic shocks in peer firms' stock return as an instrument and establishes a causal link between a firm's capital structure and its peer firms' financing decision. They show that the peer effects are more important for capital structure determination than most previously identified determinants. In addition to a number of empirical results indicating the importance of firm connections, Foucault and Fresard (2014) also provide a theoretical framework where peers' valuation matters for a firm's investment decisions.

Besides firms' operational and financial policies, the effects of firm relations on other firm outcomes have also been demonstrated by existing studies. Cohen and Frazzini (2008) and Menzly and Ozbas (2010) both find evidence of return predictability across economically linked firms, suggesting that supply chain dependencies can influence market behavior. These findings underscore the importance of supply chain connections in investment strategies.

Apart from being a direct result of peer firms' influences, firm commonality can also come from firms having similar exposure to similar risks. DeAngelo and Roll (2015) finds that factor structure explains around 30% of the variations of firm leverage, suggesting that factors have significant explanatory power of capital structure. The common exposure to risks can lead to commonalities across observations and thus statistical issues. For example, Huang and Östberg (2023) shows that factor structure plays an important role in real estate returns and causes bias in the difference-in-difference estimator.

Although there is a big strand of empirical corporate finance studying firm connections, I contribute to this area in three ways. First, instead of focusing on local peer firms such as firms in proximity or firms in the same industry, I study a broader corporate landscape where each firm has a commonality score with respect to others. By incorporating the general landscape, I illustrate the spatial structure of various firm characteristics. Second, in my study, the commonality among firms takes continuous values from 0 to 1, which takes into account the intensity of connections compared to a simple binary relation. Third, I illustrate the statistical issues in the current method and provide a solution to the over-rejection problem.

My work also relates to recent studies that investigate the validity of current empirical methods (e.g. Mitton (2021, 2022), Baker, Larcker, and Wang (2022)). Berg, Reisinger, and Streitz (2021) discusses the impact of spillover effects on the treatment effect estimates and recommends including the average peer outcomes. Kelly (2019) finds there is severe spatial autocorrelation in regressions residuals in the literature regarding the persistence of a place's modern outcomes and its characteristics in the distant past. The author demonstrates this overlooked spatial autocorrelation is linked to the unusually high t statistics in these studies and concludes that in most cases, the results of existing literature can be driven by spatial noises. The paper emphasizes that the results of regressions that may have spatial structures should be treated with caution without noise simulation. However, the paper does not provide a general solution to fix the t statistic inflation. I show in my paper that a bootstrapped standard error has the desired property and can be used to improve the hypothesis testing.

3 Data Description

The dataset consists of two parts. The first part is the firm characteristic variables and the second part is the measurement of commonality. For firm characteristics, I use the variables that have been widely used in the existing finance literature. For commonality measurement, I use the product similarity index developed in Hoberg and Phillips (2016). This measurement is simple yet sufficient and publicly accessible.

3.1 Firm characteristics data

In this section, I provide details about the firm characteristic variables. Firm fundamentals are from Compustat from 1962 to 2022 at a quarterly frequency. I focus on firms that are included in the Russell3000 index (2,970 firms) in December 2022. By filtering out the small firms, I keep the landscape structure simple without losing much economic meaning. By excluding firms that are not constituents of the Russell 3000 index, I have maintained a dataset comprising 268,672 observations across 2,887 firms. I then merge the dataset with the TNIC (Text-based Network Industry Classification) dataset in Hoberg and Phillips (2016). Now we have a dataset of 264,792 observations consisting of 2,766 firms.

Table 1: Summary Statistics

Variable	Obs	Mean	s.d.	1%	Median	99%
Panel A: Russell 3000 constituents						
Capital expenditure	217,211	0.028	0.047	0.000	0.013	0.211
Cash to asset ratio	243,056	0.148	0.197	0.001	0.067	0.915
Cash flow to asset	216,587	0.000	0.761	-0.232	0.012	0.108
Market-to-book ratio	234,654	2.204	19.993	0.743	1.416	10.353
Book Leverage	225,382	0.514	46.105	0.000	0.367	1.562
Size	244,841	7.263	2.083	2.104	7.295	12.205
Net working capital	198,668	1.567	415.596	-0.256	0.590	1.085
R&D expenditure	270,391	2.013	162.724	0.000	0.000	4.942
Financing cash flow	209,878	0.026	0.543	-0.248	-0.002	0.736
Dividend payment	270,391	0.462	0.499	0.000	0.000	1.000
Acquisition expense	270,391	0.011	0.045	-0.003	0.000	0.233
Δ Cash	236468	0.112	30.290	-0.167	0.000	0.417
Net equity issuance	201,895	0.014	0.143	-0.170	0.000	0.637
Net bond issuance	205,820	0.015	0.277	-0.184	0.000	0.339
Other cash	270,314	0.029	0.157	0.000	0.000	0.486
Tobin's Q	234,654	-0.971	15.114	-9.222	-0.317	0.924
Panel B: Balanced panel for capital expenditure						
Capital expenditure	31,880	0.026	0.035	0.001	0.016	0.167
Cash to asset ratio	31,880	0.139	0.147	0.002	0.088	0.680
Cash flow to asset	31,880	0.013	0.038	-0.108	0.016	0.087
Market-to-book ratio	31,880	2.296	1.883	0.766	1.731	10.283
Book Leverage	31,880	0.425	1.600	0.000	0.391	1.647
Size	31,880	7.903	1.708	4.125	7.828	12.248
Net working capital	31,880	0.596	0.248	-0.111	0.627	1.032
R&D expenditure	31,880	0.048	0.631	0.000	0.000	0.406
Financing cash flow	31,880	-0.010	0.099	-0.250	-0.013	0.354
Dividend payment	31,880	0.011	0.026	0.000	0.003	0.099
Acquisition expense	31,880	0.019	0.055	-0.001	0.000	0.294
Panel C: Balanced panel for cash holding						
Δ Cash	22640	0.004	0.069	-0.142	0.000	0.190
Net equity issuance	22,640	-0.013	0.074	-0.238	-0.002	0.211
Net bond issuance	22,640	0.017	0.069	-0.126	0.000	0.296
Other cash	22,640	0.016	0.063	0.000	0.000	0.296
Market-to-book ratio	22,640	2.340	2.127	0.788	1.714	11.080
Size	22,640	7.990	1.782	4.050	7.891	12.216
Net working capital	22,640	0.595	0.229	-0.085	0.629	1.026
Capital expenditure	22,640	0.029	0.038	0.001	0.018	0.176
Cash flow to asset	22,640	0.013	0.039	-0.117	0.015	0.087
Book Leverage	22,640	0.444	1.386	0.000	0.410	1.654
Dividend payment	22,640	0.012	0.026	0.000	0.004	0.105
R&D expenditure	22,640	0.060	1.370	0.000	0.000	0.361
Acquisition expense	22,640	0.018	0.054	-0.000	0.000	0.287

Later in section 5, I implement the spatial regression model where a fully balanced panel dataset without missing value is required. Therefore, I created a balanced subset of our dataset from 2013Q1 to 2022Q4. After dropping missing values, we obtain a balanced panel of 797 firms with 43,440 observations for the capital expenditure regression or 566 firms with 22,640 observations for the cash holding regression.

The definition of variables can be found in table A.1. All variables are winsorized at 1%. Table 1 shows the summary statistics of the whole dataset and two balanced subsets.

3.2 Measuring commonality

As mentioned above, the measurement of firm commonality follows Hoberg and Phillips (2016). In their papers, firm commonality is measured by the cosine similarity of firms' business descriptions in their 10K filings, hereafter referred to as the commonality score. The idea is that when firms pick similar words to describe themselves, they are likely to have similar business models and thus higher commonality. This method puts every firm at a specific virtual location in the corporate landscape and tells us the interconnectedness among firms.

For a given firm i at year t, its virtual location can be represented by a vocabulary vector P_{it}, with each element equal to 1 if firm i uses the given word in its business description at year t, and zero if it does not[2]. The vocabulary vectors are normalized to have a unit length as follows:

$$V_{it} = \frac{P_{it}}{\sqrt{P_{it} \cdot P_{it}}}$$

V_{it} can be interpreted as the virtual location of firm i at year t.

The commonality score $w_{ij,t}$ between firm i and firm j at year t is therefore defined as

$$w_{ij,t} = V_{it} \cdot V_{jt} \tag{1}$$

where $\cdot$ represents the inner product of two vectors. In fact, we can see that

$$w_{ij,t} = \frac{P_{it} \cdot P_{jt}}{\sqrt{\|P_{it}\| \times \|P_{jt}\|}} \tag{2}$$

[2]The whole dictionary of vocabularies is constructed by all words appear in at least one business descriptions at year t. The dictionary excludes words other than nouns or proper nouns, and also excludes the words that appear in more than 25% of all business descriptions in the given year.

which is the definition of cosine similarity of vocabulary vector P_{it} and P_{jt}.

The formula assigns each firm a virtual location based on its business description. Each firm has a unique location and connects with other firms in the corporate landscape based on the commonality score. A higher commonality score implies a higher correlation in firm characteristics.

For simplicity, in the following analysis, I use a static version of the commonality score. The static score between firm i and firm j (w_{ij}) is taken as the median value of commonality scores $w_{ij,t}$ across all year t[3]. This simplification is appropriate since $w_{ij,t}$ does not vary a lot across t. Table 2 shows the descriptive statistics of commonality scores and we see that the within-firm-pair variance contributes only 15% to 25% of the total variance.

Table 2: Describe Statistics of Commonality Scores

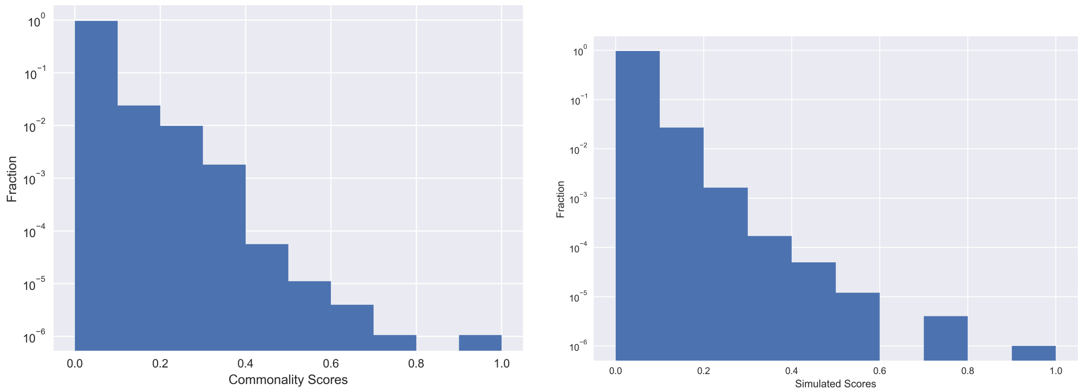
This table presents the descriptive statistics of commonality scores. The results illustrate three different samples: the whole Compustat universe, constituents of Russell 3000 index, and constituents of S&P 1500 index. Column 2 to 6 display the number of firm-pair × year observations, the number of firms, the mean value of the commonality scores, the total variance of the commonality scores, the between-firm-pair variance of the commonality scores, and the percentage of variance explained by between variance, respectively.

Sample range	Observations	Firms	Mean	Total Var.	Btw. Var.	Between %
Compustat	983,570,310	18633	0.0178	0.0014	0.0012	86.19%
Russell 3000	79,009,590	2766	0.0178	0.0013	0.0011	84.39%
S&P 1500	66,746,114	1929	0.0167	0.0010	0.0008	77.48%

The left panel of figure 1 displays the distribution of pair-wise commonality scores of Russell 3000 constituents. As one would expect, most firms do not have a very high commonality. 49.30% of the firm pairs have a median commonality score of 0 and 96.44% of the firm pairs' median commonality score is under 0.1. The distribution of commonality scores decreases exponentially with the exception of 4 firm pairs with commonality scores higher than 0.9. To show that this distribution is feasible under reasonable conditions, I run a simulation with 2000 firms and a dictionary consisting of 1000 words, where firms randomly choose 1 to 30 words from the dictionary to describe their business. The simulation gives us a similar distribution of commonality scores.

A natural question is how the commonality score relates to the SIC code, which reflects

[3]Taking the mean value of commonality scores does not lead to a result change.



(a) Real Commonality Scores (b) Simulated Commonality Scores

Figure 1: Histogram of Commonality Scores

whether firms are in the same industry or not. Figure 2 compares the two-digit SIC code and the commonality score.

As one would expect, along the diagonal of the graph, firms within the same SIC industry have higher commonality scores. Particularly, some industries such as Chemicals and allied products (SIC code 28) and Depository Institutions (SIC 60) have high within-industry commonality scores at around 0.4. This does not come as a surprise, since these industries are highly specialized and firms are likely tightly connected.

Across SIC industries, we can see that firms with SIC code 60 to 67 have high commonality scores with each other. This is also aligned with our expectation since Finance, Insurance, and Real Estate are generally seen as deeply linked industries.

Last but not least, the dark areas in the heat map are scattered around the whole graph which means firms that are in the same industries do not necessarily have the highest commonality scores and vice versa. This tells us the two-digit SIC code is not a reliable replacement for firm commonality.

4 Firm Characteristics exhibits spatial dependence

4.1 Pair-wise correlation of firm characteristics

In this section, I illustrate that the clustering of firms' characteristics indeed can be reflected by the commonality score between them. Firms with high commonality scores with each other show similar development with respect to their characteristics.

Empirically, I divide firm pairs into 10 decile groups based on their commonality

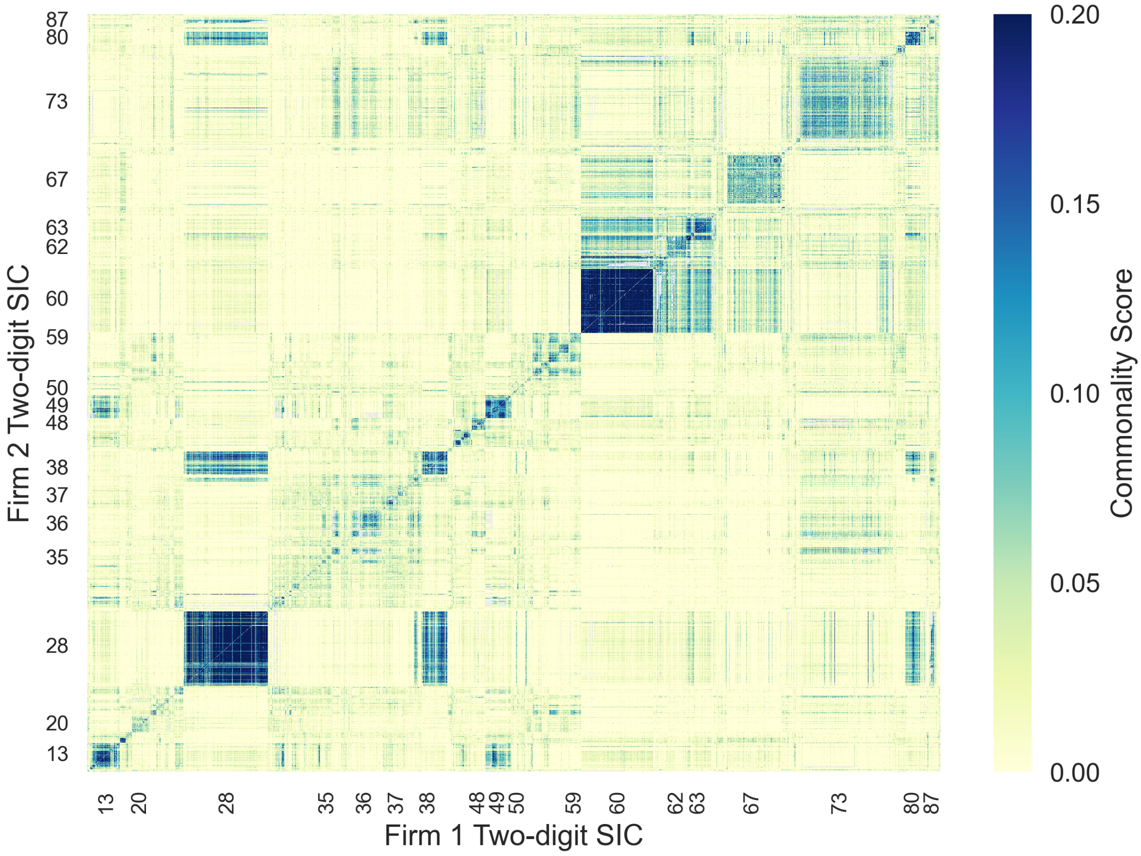
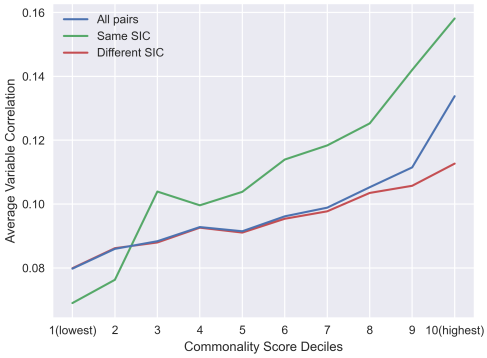


Figure 2: Heatmap of Commonality Scores

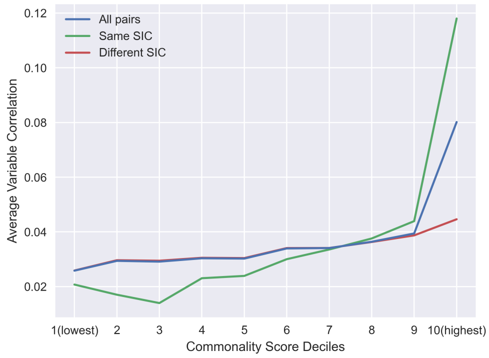
scores. Then I calculate the average pair-wise correlation of firm characteristics in each decile group. Figure 3 plots the average correlation of different firm characteristics for each commonality score group.

As we can see from the blue line, the correlation of firm characteristics increases as their commonality score becomes higher. Especially, for variables that closely reflect firm decisions, such as capital expenditure, the correlation between firms is always higher than 0.2 and increases steadily with firms' commonality scores.

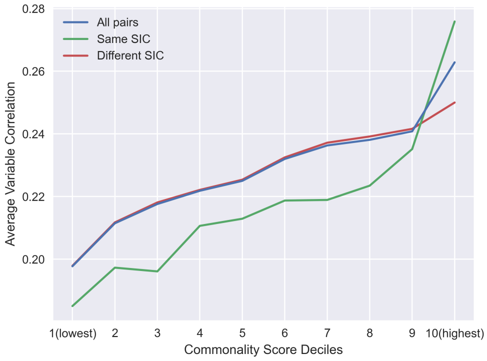
For each firm pair, I categorize it into the "same SIC" group and the "different SIC" group and then I divide them into commonality score decile groups and perform the same exercise as described in the last paragraph. This analysis yields two findings. On one hand, the SIC code is not a reliable measurement of firm commonality since firm pairs within the same industry do not always have a higher correlation of firm characteristics. On the other hand, even when firms are from different SIC industries, their commonality score mirrors their correlation in important firm characteristic variables.



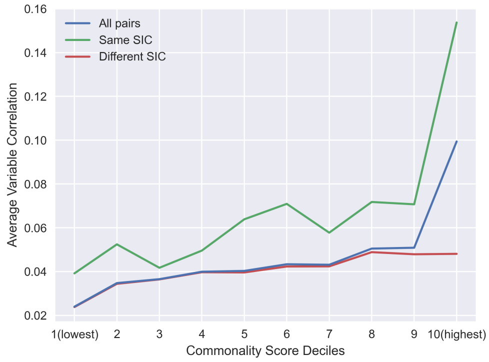
(a) Book-to-Market Ratio



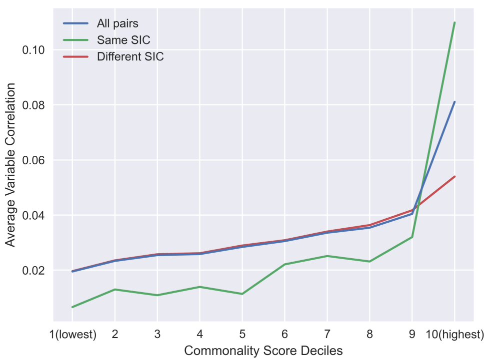
(b) Cash Holding



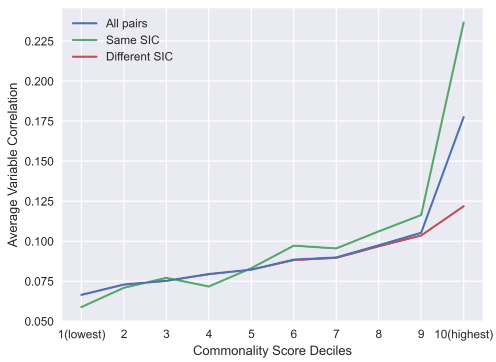
(c) Capital Expenditure



(d) Book Leverage



(e) Return of Asset



(f) Tobin's Q

Figure 3: Firm Characteristics Correlation Increases with Commonality Score

4.2 Interpretation of pair-wise correlation

So far, we have seen correlations among firm characteristics, but it remains a question of how to interpret these correlations. In this section, I illustrate this question with a simulation of a spatial autoregressive model of the corporate landscape.

An alternative way to interpret the pair-wise correlation is to directly link the spatial autoregression coefficient ρ with the average pair-wise correlation. Specifically, we assume variable X_i follows a spatial autoregressive model:

$$\begin{aligned} X_i &= \sum_j \rho W_{ij} X_j + Z_i \\ Z_i &\overset{\text{iid}}{\sim} \mathcal{N}(0,1) \end{aligned} \tag{3}$$

where X_i is the characteristic of firm i, X_j is the characteristic of firm j, W is the matrix of commonality scores and ρ is the spatial autoregression coefficient with respect to commonality metrics W.

Equation 3 can be rewritten in matrix form,

$$\boldsymbol{X} = \rho \boldsymbol{W} \boldsymbol{X} + \boldsymbol{Z}$$

where $\boldsymbol{X}$ is a $N \times 1$ vector, $\boldsymbol{W}$ is a $N \times N$ matrix and $\boldsymbol{Z}$ is a $N \times 1$ vector.

Solving for $\boldsymbol{X}$, we have,

$$\begin{aligned} \boldsymbol{X} &= \left(\boldsymbol{I_n} - \rho \boldsymbol{W}\right)^{-1} \boldsymbol{Z} \\ &= \sum_{i=1}^{\infty} \left(\rho \boldsymbol{W}\right)^i \boldsymbol{Z} \\ &= \rho \boldsymbol{W} \boldsymbol{Z} + \rho^2 \boldsymbol{W}^2 \boldsymbol{Z} + \rho^3 \boldsymbol{W}^3 \boldsymbol{Z} + \cdots \end{aligned}$$

as long as the inverse of $(\boldsymbol{I_n} - \rho \boldsymbol{W})$ exists, i.e., $\sum_{i=1}^{\infty} \left(\rho \boldsymbol{W}\right)^i$ converges.

I conduct a simulation according to the spatial autoregressive model described above, using the real commonality score matrix W. Figure 4 shows the relationship between the coefficient ρ and average pair-wise correlation. As ρ approaches 1, the average correlation increases rapidly towards 1. For example, when we take $\rho = 0.8$, we have an average pairwise correlation of 0.028, but when $\rho = 0.95$, the average correlation rises to 0.27.

4.3 Moran's I statistics

Now that we have an intuitive understanding of spatial autocorrelation and its relationship with the correlation of firm characteristics, we examine the coefficient ρ in our dataset formally using Moran's I statistics (Moran (1950)). Moran's I is a widely used global index that measures the similarity of values in neighboring places from an overall mean value and reflects a spatially weighted form of Pearson's correlation coefficient. The

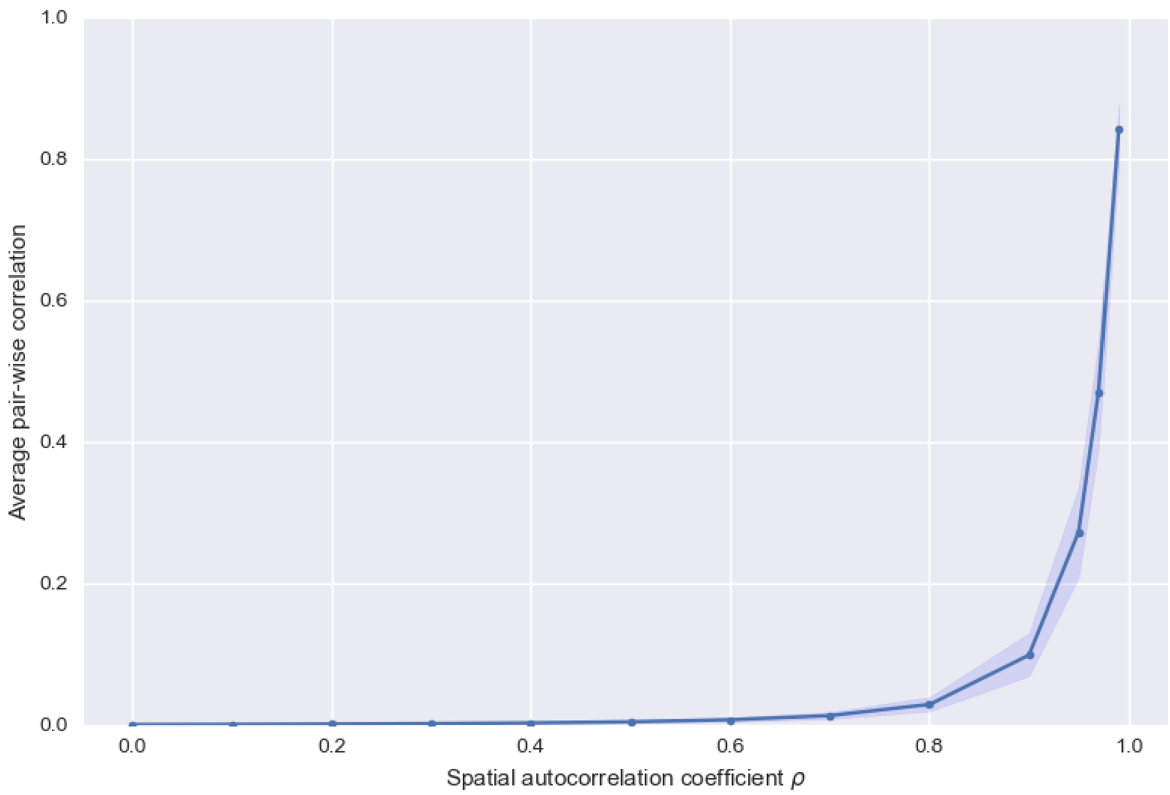


Figure 4: Average pair-wise correlation in Corporate Landscape

formula of the test statistics is,

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \left(X_i - \bar{X}\right) \left(X_j - \bar{X}\right)}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}$$

where N is the number of firms (spatial units) indexed by i and j, X is the variable of interest, $\bar{X}$ is the mean of the variable, w_{ij} is the i,j-th element of the matrix $\boldsymbol{W}$ and W is the sum of all w_{ij}, i.e., $W = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$.

The basic idea of the test is that if $\rho = 0$ is true, then the permutation between X_i's does not change the distribution of the test statistic. Therefore, I permute the X_i's uniformly at random and obtain the distribution of the test statistic under the null hypothesis. Using this distribution, the Moran's I value can be transformed into a Z-score by taking the difference between the sample Moran's I and the mean of the distribution and dividing it by the variance.

Table 3 presents the test results of various widely-used firm characteristics. Surprisingly yet aligned with our expectation, we find that all these variables have a highly significant spatial structure.

Table 3: Moran's I statistics

This table presents Moran's I statistics of commonality scores. The mean, standard deviation, Z-score, and P-value of Moran's I statistics are calculated by conducting a random permutation procedure 999 times.

Variable	Moran's I	Mean	Std dev	Z-score	P-value
Book-to-Market Ratio	0.0550	-0.0004	0.0012	44.6896	0.0000
Cash Holding	0.6301	-0.0004	0.0010	612.8756	0.0000
Capital Expenditure	0.1907	-0.0004	0.0010	185.1099	0.0000
Book Leverage	0.0628	-0.0004	0.0011	58.5502	0.0000
Return of Asset	0.3022	-0.0004	0.0010	294.0209	0.0000
Tobin's Q	0.2495	-0.0004	0.0011	242.9228	0.0000

5 Empirical Examples with Firm Commonality

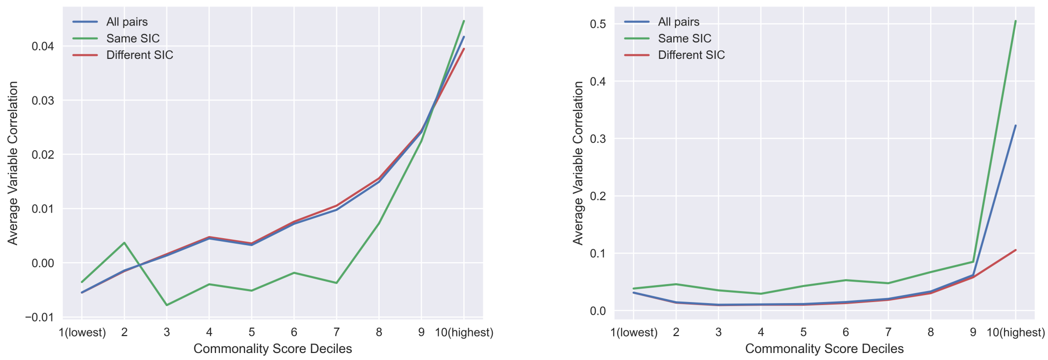
Simulations have shown concerning results regarding the validity of the regression. In this section, I use the real-world dataset and provide two empirical examples of capital expenditure and cash holding, showing that fixed effects estimators without accounting for firm commonality lead to potentially biased estimates.

5.1 Firm Commonality in TWFE Residuals

Figure 5 shows that the residual of a two-way fixed effect (TWFE) regression does not have a pattern of independent and identically distributed variables. To be more specific, had the residuals been independent and identically distributed, the lines in the graph should be horizontal and should be at the 0 level.

On the opposite, the residuals show a strong spatial dependence pattern. In other words, when a firm pair has a higher commonality score, these two firms' residuals also have a higher correlation. This correlation should not be overlooked. Take the capital expenditure graph as an example, as the firm commonality score increases to its highest decile, the correlation between firms' residuals increases notably to 0.5.

In the following sections, I first introduce three spatial models that can accommodate firm commonalities and then focus on two empirical examples of capital expenditure and



(a) Cash Holding (b) Capital Expenditure

Figure 5: Residuals of TWFE Estimator exhibit spatial dependence

cash holding. I show that taking the spatial structure into account significantly changes the economic meaning of widely used explanatory variables in current literature and provides a more complete picture of the relationship among firms.

5.2 Spatial Regression Model

Different types of spatial dependence can be modeled in different ways. The simplest form of spatial dependence is modeled by the spatial autoregressive model (SAR). Equation 4 defines the regression form:

$$\begin{aligned} Y_{it} &= \rho \sum_j W_{ij} Y_{jt} + X_{it}\beta + \varepsilon_{it} \\ \mathbb{E}\left[\varepsilon_{it} \middle| \boldsymbol{X}, \boldsymbol{W}\right] &= 0 \end{aligned} \tag{4}$$

In addition to the spatial autoregressive term of the dependent variable, if there is also spatial spillover of the covariates, then we extend the SAR to the spatial Durbin model (SDM). The model form is:

$$\begin{aligned} Y_{it} &= \rho \sum_j W_{ij} Y_{jt} + X_{it}\beta + \theta \sum_j W_{ij} X_{jt} + \varepsilon_{it} \\ \mathbb{E}\left[\varepsilon_{it} \middle| \boldsymbol{X}, \boldsymbol{W}\right] &= 0 \end{aligned} \tag{5}$$

On the other hand, if there is spatial autocorrelation in the disturbance term, then the basic model extends to the spatial autoregressive combined model (SAC) which follows:

$$\begin{aligned} Y_{it} &= \rho \sum_j W_{ij} Y_{jt} + X_{it}\beta + \varepsilon_{it} \\ \varepsilon_{it} &= \lambda \sum_j W_{ij} \varepsilon_{jt} + u_{it} \\ \mathbb{E}\left[u_{it} \right| \boldsymbol{X}, \boldsymbol{W}] &= 0 \end{aligned} \tag{6}$$

In the following empirical examples, I estimate all three forms of spatial regression models.

5.3 Example 1: Capital Expenditure

There is a number of recent literature focusing on the relationship between firms' investment spending (proxied by capital expenditure) and cash flow (Güner, Malmendier, and Tate (2008), Gatchev, Pulvino, and Tarhan (2010)[4]). However, models that do not acknowledge the interdependent nature of firms' investments lead to an incomplete and potentially misleading view of firm behavior. In this section, I examine the results of commonly used regressions of capital expenditure and show the importance of the spatial structure.

Table 4 shows the results of fixed effects models and spatial models. One thing we have to note here is that the point estimates of TWFE models are not directly comparable with the spatial regression estimates. As I discuss in appendix B, the coefficient reported in the TWFE estimator measures the direct effect $\frac{\partial y_{it}}{\partial x_{it}}$, which is unequal to coefficient β_k in spatial regression. To make the β_k in spatial regression comparable with the TWFE estimator, we need to calculate the direct effects of the spatial regression as well. The direct effects of explanatory variable X_k is $\left(\boldsymbol{I}_N - \rho W\right)^{-1} \beta_k$ and the average direct effects is the mean of the main diagonal elements, i.e. $\frac{\mathrm{tr}(\boldsymbol{I}_N - \rho W)^{-1}}{N} \beta_k$. For example, in the fourth column, when $\rho = 0.95$, the multiplier is equal to 1.026.

Comparing the last three columns with the fixed effects models, we can see that after incorporating the spatial structure, the point estimates of our variable of interest, i.e., cash flow to asset, decrease by half. I also find significant comovement of firms' capital

[4] Gatchev, Pulvino, and Tarhan (2010) find a cash flow to asset coefficient of 0.0893, while Güner, Malmendier, and Tate (2008) report a slightly higher coefficient at 0.1199.

expenditure and cash flow which are reflected in ρ and θ.

The decrease in point estimates has significant economic implications. Specifically, within the firm fixed effects model, a one standard deviation shift in the cash flow to asset ratio is typically linked with a corresponding 0.14 standard deviation shift in capital expenditure. Conversely, in the Spatial Durbin Model (SDM), a similar one standard deviation variation in the cash flow to asset ratio is linked with a smaller average change of 0.066 in capital expenditure. Moreover, one standard deviation change in the capital expenditure of a peer firm is associated with an average change of 0.043 in capital expenditure, which is around two-thirds of the effect size of the impact of cash flow.

Table 4: Estimates of Spatial Regressions of Capital Expenditure

This table presents estimation results of capital expenditure in a spatial panel-data regression described in equation 4, 5, and 6. All variables are defined in table A.1. Standard errors are clustered at firm level. ***, **, * represents statistically significance at 10%, 5%, and 1% respectively.

	Capx FirmFE	Capx TWFE	Capx SAR	Capx SDM	Capx SAC
Main regression parameters β:					
Cash flow to asset	0.1267***	0.1127***	0.0603***	0.0615***	0.0600***
	(5.82)	(5.16)	(4.48)	(4.36)	(4.33)
Cash to asset ratio	-0.0306***	-0.0277***	-0.0241***	-0.0221***	-0.0225***
	(-6.37)	(-5.85)	(-5.33)	(-4.94)	(-4.96)
Market-to-book ratio	0.0013***	0.0018***	0.0016***	0.0014***	0.0015***
	(3.90)	(5.60)	(5.34)	(4.48)	(4.68)
Book Leverage	-0.0134***	-0.0107***	-0.0079***	-0.0075***	-0.0078***
	(-7.24)	(-6.08)	(-5.07)	(-4.79)	(-4.84)
Size	-0.0031***	-0.0011	-0.0013	-0.0027**	-0.0026**
	(-4.36)	(-1.20)	(-1.85)	(-3.17)	(-3.09)
Net working capital	-0.0121**	-0.0091*	-0.0081*	-0.0080*	-0.0082*
	(-3.12)	(-2.38)	(-2.24)	(-2.20)	(-2.24)
R&D expenditure	0.0612***	0.0441***	0.0440***	0.0454***	0.0452***
	(6.56)	(4.87)	(4.89)	(4.92)	(4.93)
Financing cash flow	0.0137*	0.0187**	0.0162**	0.0155**	0.0158**
	(2.39)	(3.30)	(3.19)	(2.97)	(3.05)
Dividend payment	0.519***	0.402***	0.375***	0.366***	0.363***
	(13.23)	(11.40)	(11.30)	(10.88)	(10.82)
Acquisition expense	0.0120*	-0.0120*	-0.0114*	-0.0091	-0.0090
	(2.23)	(-2.08)	(-2.26)	(-1.69)	(-1.73)
Firm Fixed Effect	✓	✓	✓	✓	✓
Quarter Fixed Effect		✓			
Spatial correlation parameters:					
Dep. Var. ρ			0.953***	0.940***	0.885***
			(70.53)	(79.76)	(40.08)
Error term λ					0.873***
					(41.27)
Covariates θ:					
Cash flow to asset				-0.154*	
				(-1.96)	
Other covariates				✓	
Direct effects (average of main diagonal $(\boldsymbol{I_N} - \rho \boldsymbol{W})^{-1}\beta_k$):					
Cash flow to asset			0.0610***	0.0581***	0.0596***
			(4.21)	(4.01)	(4.06)
Number of obs.	31,880	31,880	31,880	31,880	31,880

5.4 Example 2: Cash holding

Determinants of firm cash holding is another prevailing topic in empirical corporate finance. McLean (2011) divide the source of firm cash holding into four categories, including net equity issuance, net debt issuance, cash flow from operation, and other source. The paper then studies the share issuance–cash savings relation in a single equation static regression model. As we have seen in the previous sections, cash holding exhibits significant spatial dependence. Therefore, to provide a complete picture of cash holding motives, I apply the spatial regression model to cash holding and underscore its importance.

Table 5 shows the result. The point estimates of the main explanatory variables do not vary a lot. However, the spatial regression illustrates two important features of firm cash holding. First, comparing the SAC and SAR model, the spatial autocorrelation coefficient of Y_{jt} (i.e., ρ) becomes insignificant after adding spatial dependence of the disturbance term (i.e., λ). This suggests that firms do not affect each other's cash holding directly, rather it is more likely that firms experience common shocks to their cash holding. Second, although a firm's own net equity issuance and cash flow to asset have significant positive effects on its cash holding, we obtain significant negative θ for these two variables. This suggests a squeezing effect from competition, where other firms' equity issuance or the increase in their operating cash flow is negatively associated with my cash holding.

Table 5: Estimates of Spatial Regressions of Cash Holdings

This table presents estimation results of the first difference of cash holdings in a spatial panel-data regression described in equation 4, 5, and 6. All variables are defined in table A.1. Standard errors are clustered at firm level. ***, **, * represents statistically significance at 10%, 5%, and 1% respectively.

	ΔCash FirmFE	ΔCash TWFE	ΔCash SAR	ΔCash SDM	ΔCash SAC
Main regression parameters β:					
Net equity issuance	0.146***	0.142***	0.142***	0.144***	0.145***
	(13.78)	(13.59)	(13.58)	(13.75)	(13.71)
Net bond issuance	0.133***	0.132***	0.131***	0.132***	0.133***
	(15.43)	(15.43)	(15.36)	(15.35)	(15.52)
Cash flow to asset	0.198***	0.215***	0.207***	0.225***	0.217***
	(5.81)	(6.37)	(6.14)	(6.46)	(6.15)
Other cash	0.00945	0.00234	0.00689	0.00441	0.00498
	(0.76)	(0.19)	(0.55)	(0.36)	(0.40)
Market-to-book ratio	-0.00264***	-0.00222***	-0.00244***	-0.00219***	-0.00249***
	(-5.46)	(-4.61)	(-5.14)	(-4.57)	(-4.99)
Size	0.00713***	0.00786***	0.00717***	0.00762***	0.00709***
	(7.09)	(6.48)	(7.19)	(7.08)	(6.07)
Net working capital	-0.0778***	-0.0788***	-0.0766***	-0.0765***	-0.0787***
	(-14.77)	(-14.65)	(-14.65)	(-14.59)	(-14.52)
Capital expenditure	-0.0301	-0.0628**	-0.0294	-0.0499**	-0.0482*
	(-1.84)	(-3.17)	(-1.83)	(-2.86)	(-2.48)
Book leverage	0.00405	0.00257	0.00326	0.00291	0.00288
	(1.95)	(1.20)	(1.57)	(1.40)	(1.32)
Dividend payment	0.256***	0.247***	0.256***	0.259***	0.249***
	(5.04)	(4.89)	(5.09)	(5.10)	(4.85)
R&D expenditure	-0.0186	-0.0195	-0.0181	-0.0179	-0.0206
	(-0.91)	(-0.96)	(-0.90)	(-0.89)	(-1.01)
Acquisition expense	-0.212***	-0.212***	-0.209***	-0.212***	-0.213***
	(-19.14)	(-19.07)	(-18.95)	(-19.03)	(-18.91)
Firm Fixed Effect	✓	✓	✓	✓	✓
Quarter Fixed Effect		✓			
Spatial correlation parameters:					
Dep. Var. ρ			0.602***	0.553***	0.0174
			(21.76)	(17.52)	(0.17)
Error term λ					0.666***
					(15.07)
Covariates θ:					
Net equity issuance				-0.125**	
				(-2.91)	
Net bond issuance				-0.0353	
				(-0.98)	
Cash flow to asset				-0.432***	
				(-5.51)	
Other cash				0.0969	
				(1.04)	
Number of obs.	22,640	22,640	22,640	22,640	22,640

6 Two-way Clustering and its Failure

A vast stream of finance literature has been taking care of the correlations in residuals by using firm-clustered standard errors. However, firm-clustered standard errors only allow residuals to be correlated within a firm, not across firms. Therefore, the previous problem still exists and leads to an over-rejection. An alternative is to cluster on time dimension, however, this method also fails under practical circumstances.

I illustrate the problem with two-way clustering standard errors through a simulation study that focuses on the panel data setting where both spatial dependence and fixed effects are present. Specifically, the data-generating process follows

$$\begin{aligned}
Y_{it} &= \beta X_{it} + \varepsilon_{it} \\
X_{it} &= \sum_j \phi W_{ij} X_{jt} + Z_{it} \\
Z_{it} &= \alpha_i + h_{it} \\
\varepsilon_{it} &= \sum_j \lambda W_{ij} u_{jt} + u_{it} \\
u_{it} &= \gamma_i + v_{it}
\end{aligned} \tag{7}$$

where $\alpha_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$, $h_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_h^2)$, $\gamma_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\gamma^2)$, $v_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_v^2)$ and W_{ij} is the commonality score I describe above. The default values of parameters are as following: $\beta = 0$, $\sigma_\alpha^2 = 0.5$, $\sigma_h^2 = 0.5$, $\sigma_\gamma^2 = 0.5$, $\sigma_v^2 = 0.5$. In the benchmark setting, the panel consists of a fully balanced panel with 1000 firms and 40 quarters. For the spatial autocorrelation coefficients, I assign three discrete values to ϕ and λ: 0, 0.8 and 0.9, respectively representing the cases of no connection, mild connection, and strong connection. A spatial autoregression coefficient of 0.8 may seem high at first glance, but it actually corresponds to an average pairwise correlation of only 0.028, mirroring the typical pairwise correlation observed in our dataset.

In the data-generating process 7, the independent variable Y_{it} and dependent variable X_{it} are spatially correlated and have unobserved heterogeneity components. It indicates that both variables have a correlation among firms and across time. As suggested by Petersen (2008), we can implement two-way clustering to address the correlation issue on these dimensions.

Table 6 presents the simulated rejection rates of the benchmark specifications. We can see that under the benchmark setting, clustering only at the firm dimension is not enough as it leads to higher rejection rates when spatial dependence exists. However, under the benchmark case, two-way clustered standard errors correctly reject roughly 5% of the samples.

Table 6: Rejection rates of Panel Simulation with Benchmark Specification

This table presents rejection rates of the coefficient β in the panel regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. Both ε_{it} and X_{it} follow the data-generating process described in equation 7. The true value of β is 0. The null hypothesis is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. Column (2) and (3) control for only firm fixed effects while column (4) and (5) control for both firm and quarter fixed effects. Column (2) and (4) use firm-clustered standard errors and column (3) and (5) use firm- and quarter-clustered standard errors. All simulations are run 1000 times. The significance level is 5%.

Parameters	FE[F]V[F]	FE[F]V[FT]	FE[FT]V[F]	FE[FT]V[FT]
$\phi = 0,\ \lambda = 0$	0.049	0.058	0.047	0.063
$\phi = 0,\ \lambda = 0.8$	0.047	0.059	0.046	0.064
$\phi = 0,\ \lambda = 0.9$	0.050	0.057	0.050	0.055
$\phi = 0.8,\ \lambda = 0$	0.046	0.061	0.049	0.068
$\phi = 0.8,\ \lambda = 0.8$	**0.109**	0.053	**0.086**	0.056
$\phi = 0.8,\ \lambda = 0.9$	**0.267**	0.061	**0.188**	0.054
$\phi = 0.9,\ \lambda = 0$	0.049	0.055	0.055	0.074
$\phi = 0.9,\ \lambda = 0.8$	**0.267**	0.060	**0.200**	0.063
$\phi = 0.9,\ \lambda = 0.9$	**0.519**	0.055	**0.424**	0.049

According to White (2014), Cameron and Miller (2015) and MacKinnon, Nielsen, and Webb (2023), two-way clustered standard errors are only consistent under three key assumptions:

- The number of clusters goes to infinity.
- The correlation within clusters is identical.
- The number of observations in each cluster are same.

If either of the assumptions is violated, we do not obtain a consistent estimate of standard errors. In the subsequent three subsections, I show the simulation results of two-way clustering standard errors when these assumptions are relaxed.

6.1 Short panel

Angrist and Pischke (2009) and Donald and Lang (2007) suggest that the number of clusters should be greater than 50 for good practical performance of clustering standard errors. Unfortunately, this is not always feasible in empirical corporate finance research since many firm fundamentals are only reported annually. In the simulation, I restricted the length of the dataset to 10 to mimic the situation empirical researchers often face.

Table 7 presents the results with a short panel dataset. Contrary to the benchmark case, two-way clustering standard errors also suffer the over-rejection problem now, especially when spatial autocorrelation is present.

Table 7: Rejection rates of Panel Simulation with 10 Quarters

This table presents rejection rates of the coefficient β in the panel regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. Compared to the benchmark specification, the length of the panel is shorten to 10 quarters. Both ε_{it} and X_{it} follow the data-generating process described in equation 7. The true value of β is 0. The null hypothesis is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. Column (2) and (3) control for only firm fixed effects while column (4) and (5) control for both firm and quarter fixed effects. Column (2) and (4) use firm-clustered standard errors and column (3) and (5) use firm- and quarter-clustered standard errors. All simulations are run 1000 times. The significance level is 5%.

Parameters	FE[F]V[F]	FE[F]V[FT]	FE[FT]V[F]	FE[FT]V[FT]
$\phi = 0, \lambda = 0$	0.048	**0.075**	0.049	**0.065**
$\phi = 0, \lambda = 0.8$	0.053	**0.076**	0.049	**0.085**
$\phi = 0, \lambda = 0.9$	0.047	**0.075**	0.050	**0.077**
$\phi = 0.8, \lambda = 0$	0.053	**0.082**	0.048	**0.073**
$\phi = 0.8, \lambda = 0.8$	**0.109**	**0.090**	**0.094**	**0.089**
$\phi = 0.8, \lambda = 0.9$	**0.252**	**0.189**	**0.187**	**0.180**
$\phi = 0.9, \lambda = 0$	0.048	**0.084**	0.050	**0.084**
$\phi = 0.9, \lambda = 0.8$	**0.256**	**0.199**	**0.178**	**0.201**
$\phi = 0.9, \lambda = 0.9$	**0.522**	**0.397**	**0.436**	**0.393**

6.2 Unbalanced panel

Besides the short panel, another common feature in the empirical corporate finance dataset is the unbalanced panel. Table 8 lists the number of Russell 3000 constituents in the Compustat database at the end of each year from 2013 to 2022. We can see that the number of firms shows considerable variation and the number of firms in 2013 is only 72% of the number in 2022.

In the cross-sectional case, MacKinnon and Webb (2017) and Djogbenou, MacKinnon, and Nielsen (2019) show that unequal size of clusters can lead to over-rejection of clustered standard errors. I extend their simulation to the panel setting with the corporate dataset. Table 9 presents my simulation results with an unbalanced panel. The specification in

Table 8: Number of Russell 3000 constituents available in Compustat

This table presents the number of Russell 3000 constituents in Compustat database at the end of each year from 2013 to 2022.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
No. Firms	1977	2041	2122	2209	2302	2424	2622	2696	2764	2743

table 9 is identical to the baseline specification except that the number of firms for each quarter now follows table 8.

As one would expect, neither firm clustered standard errors nor two-way clustering standard errors are correct under the setting. Notably, even when one of the spatial autocorrelation coefficients is zero, these standard errors still lead to the over-rejection problem, even though the magnitude is not large.

Table 9: Rejection rates of Panel Simulation with unbalanced dataset

This table presents rejection rates of the coefficient β in the panel regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. Compared to the benchmark specification, the number of firms for each quarter follows table 8. Both ε_{it} and X_{it} follow the data-generating process described in equation 7. The true value of β is 0. The null hypothesis is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. Column (2) and (3) control for only firm fixed effects while column (4) and (5) control for both firm and quarter fixed effects. Column (2) and (4) use firm-clustered standard errors and column (3) and (5) use firm- and quarter-clustered standard errors. All simulations are run 1000 times. The significance level is 5%.

Parameters	FE[F]V[F]	FE[F]V[FT]	FE[FT]V[F]	FE[FT]V[FT]
$\phi = 0,\ \lambda = 0$	0.049	0.060	0.046	0.052
$\phi = 0,\ \lambda = 0.8$	0.055	**0.081**	0.057	**0.077**
$\phi = 0,\ \lambda = 0.9$	0.057	**0.081**	0.059	**0.084**
$\phi = 0.8,\ \lambda = 0$	0.046	0.055	0.045	0.067
$\phi = 0.8,\ \lambda = 0.8$	**0.120**	**0.118**	**0.118**	**0.109**
$\phi = 0.8,\ \lambda = 0.9$	**0.270**	**0.140**	**0.198**	**0.137**
$\phi = 0.9,\ \lambda = 0$	0.049	0.062	0.053	0.055
$\phi = 0.9,\ \lambda = 0.8$	**0.272**	**0.150**	**0.201**	**0.153**
$\phi = 0.9,\ \lambda = 0.9$	**0.510**	**0.152**	**0.420**	**0.133**

6.3 Time-varying correlation

Another problem with two-way clustering standard errors is that the intra-cluster correlation may not be constant (Carter, Schnepel, and Steigerwald (2017)). Take the corporate landscape as an example, the commonality score among firms is actually time-varying

(Hoberg and Phillips (2010, 2016)), and thus the correlation is time-varying as well. This suggests that the traditional two-way clustering standard errors may lead to a biased estimate of the standard error.

Table 10 presents the results when the commonality score is time-varying. The specification in table 10 is identical to the baseline specification except that I replace the median commonality score W_{ij} by the commonality score for each year $W_{ij,t}$. Theover-rejection problem here is so severe that firm-clustered standard errors reject up to 95% and the two-way clustering standard errors over-reject up to seven fold.

Table 10: Rejection rates of Panel Simulation with time-varying Commonality

This table presents rejection rates of the coefficient β in the panel regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. Compared to the benchmark specification, the median firm commonality W_{ij} is replaced by firm commonality for each year $W_{ij,t}$. Both ε_{it} and X_{it} follow the data-generating process described in equation 7. The true value of β is 0. The null hypothesis is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. Column (2) and (3) control for only firm fixed effects while column (4) and (5) control for both firm and quarter fixed effects. Column (2) and (4) use firm-clustered standard errors and column (3) and (5) use firm- and quarter-clustered standard errors. All simulations are run 1000 times. The significance level is 5%.

Parameters	FE[F]V[F]	FE[F]V[FT]	FE[FT]V[F]	FE[FT]V[FT]
$\phi = 0,\ \lambda = 0$	0.060	0.076	0.060	0.066
$\phi = 0,\ \lambda = 0.8$	0.045	0.068	0.045	0.069
$\phi = 0,\ \lambda = 0.9$	0.048	0.097	0.051	0.092
$\phi = 0.8,\ \lambda = 0$	0.055	0.045	0.055	0.044
$\phi = 0.8,\ \lambda = 0.8$	**0.456**	**0.242**	**0.471**	**0.251**
$\phi = 0.8,\ \lambda = 0.9$	**0.648**	**0.313**	**0.635**	**0.294**
$\phi = 0.9,\ \lambda = 0$	0.045	0.097	0.046	0.092
$\phi = 0.9,\ \lambda = 0.8$	**0.618**	**0.282**	**0.632**	**0.278**
$\phi = 0.9,\ \lambda = 0.9$	**0.957**	**0.341**	**0.957**	**0.353**

In conclusion, the simulation results show the unreliability of the clustering standard errors in the panel setting with the presence of spatial dependence. Unfortunately, these three cases commonly exist in real-world datatsets, and this calls for a better solution for estimating the standard errors.

7 Potential Solutions

7.1 Robustness check with spatial regressions

To address the potential bias in the estimates stemming from the spatial structure, we can run a spatial regression as a robustness check. If the results of these two regressions do not differ much, we are then confident that our estimates are reliable.

Note that for the spatial regression coefficient to be consistent, we need to specify the spatial structure correctly. In this paper, I use the structure measured by Hoberg and Phillips (2010) and Hoberg and Phillips (2016) which is generally accepted as a reasonable characterization of firm connections. If we are willing to assume this is the correct structure then our spatial regression estimates are correct.

After all, the true structure is extremely difficult to measure and the approximation of it is itself another strand of research (De Paula, Rasul, and Souza (2018)). I leave this topic for further exploration.

7.2 Bootstrapping standard errors

Another approach to resolve the concerns about spatial dependence is to use the resampling methods, with bootstrapping being a notable example. Webb (2023) find that wild bootstrapping can achieve the correct p-value even if the number of clusters is only 12. MacKinnon, Nielsen, and Webb (2021) and Djogbenou, MacKinnon, and Nielsen (2019) suggest that wild bootstrapping still works correctly even if the dataset has unequal size of clusters.

Wild bootstrap is a bootstrapping method that randomly assigns weights to residuals in regression models. By resampling on residuals, the structure of variables is kept untainted.

In this exercise, I obtain the fitted values and residuals from the restricted regression. Specifically, I impose the restriction that $\beta = 0$ in the regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. The goal of the bootstrapping is to obtain the distribution under the null hypothesis that β is equal to 0. Then, I calculate a new y_{it} based on $\tilde{y}_{it} = \hat{y}_{it} + \nu_{it}\hat{\varepsilon}_{it}$ where $\hat{y}_{it}$ is the fitted value of the regression, $\hat{\varepsilon}_{it}$ is the residuals and ν_{it} is the random weights to residuals. I regress the $\tilde{y}_{it}$ back on X_{it} so that I obtain the bootstrapped distribution of β and consequently p-value under the null hypothesis.

I use the two-way resampling method to obtain random weights ν_{it}. To be specific, I get ν_{it} is defined by $\nu_{it} = \xi_i \zeta_t$ where ξ_i and ζ_t independently follow the Rademacher distribution. ξ_i and ζ_t independently follows the Rademacher distribution:

$$\xi_i \text{ or } \zeta_t = \begin{cases} -1, & \text{with probability } 0.5 \\ 1, & \text{with probability } 0.5 \end{cases}$$

The two-way resample method preserves both spatial dependence as well as time-series structure. We can see in table 11, that the simulated rejection rate of the bootstrap estimator is roughly 5%. Therefore, I propose using this two-way bootstrap rejection rate when firm commonality is present.

[5]

Table 11: Rejection rates of Panel Simulation using Bootstrapping

This table presents rejection rates of the coefficient β in the panel regression $Y_{it} = \beta X_{it} + FEs + \varepsilon_{it}$. Both ε_{it} and X_{it} follow the data-generating process described in equation 7. The true value of β is 0. The null hypothesis is $H_0 : \beta = 0$ and the alternative hypothesis is $H_1 : \beta \neq 0$. All simulations are run 1000 times. Each bootstrapped t-value is calculated through resampling 399 times. The significance level is 5%.

Parameters	Dataset	Double resample
$\phi = 0.8,\ \lambda = 0.8$	Benchmark	0.035
$\phi = 0.9,\ \lambda = 0.9$	Benchmark	0.037
$\phi = 0.8,\ \lambda = 0.8$	Short Panel	0.039
$\phi = 0.9,\ \lambda = 0.9$	Short Panel	0.035
$\phi = 0.8,\ \lambda = 0.8$	Unbalanced Panel	0.038
$\phi = 0.9,\ \lambda = 0.9$	Unbalanced Panel	0.038
$\phi = 0.8,\ \lambda = 0.8$	Time-varying Correlation	0.040
$\phi = 0.9,\ \lambda = 0.9$	Time-varying Correlation	0.037

7.3 Spatial dependence principal components

Besides the bootstrapping method, two recent papers (Müller and Watson (2022a,b)) propose another method that is able to account for spatial dependences. They show that their estimator SCPC (spatial dependence principal components) is correct under various settings with the presence of spatial dependence. This will be left to future

[5]The two-way clustering is also known as the pigeon hold bootstrap Owen (2007). Menzel (2021) discusses in detail how to bootstrap with multiple cluster dependence.

research. Further examination of this method will be left for future studies.

8 Conclusion

Firms do not operate in a vacuum and numerous studies have found that others' characteristics play an important role in firms' decisions. This paper emphasizes the spatial structure of the corporate landscape and its impact on statistical inferences. I show that spatial structure significantly changes the result of commonly used variables, and even in the presence of mild spatial dependence, the statistical significance of regressions is highly inflated.

This paper offers some simple guidelines to improve the credibility of empirical results. First, we should be aware of the presence of spatial dependences and run spatial diagnostics before regression. For instance, we should check Moran'I statistics. Second, we should take care of the spillover effect from the spatial structure. For example, run a robustness check with industry-mean or run a spatial regression aside from the main regression. Third, with the presence of correlations, we should estimate standard errors with the bootstrap method since clustering does not work properly. In sum, my findings suggest the empirical results in corporate finance should be treated with caution, and we should accommodate the firm commonality in our research framework.

Appendix

A Variable Defination

In this section, I list the definitions of variables used in this article.

Table A.1: Defination of Variables

This table presents the definition of variables. Variable are defined by Compustat items.

Variable	Definition
Capital expenditure	capxy/l.atq
Cash to asset ratio	cheq/atq
Cash flow to asset	(oibdpq-xintq-txtq-dvy)/l.atq
Market-to-book ratio	(atq-ceqq+prccq*cshoq)/atq
Book Leverage	(dlttq+dlcq)/atq
Size	ln(atq)
Net working capital	((atq-cheq)-(atq-dlcq-dlttq-mibq-pstkq-ceqq))/l.atq
R&D expenditure	xrdq/l.saleq
Financing cash flow	fincfy/l.atq
Dividend payment	dvy/l.atq
Acquisition expense	aqcy/l.atq
ΔCash	(cheq-l.cheq)/l.atq
Net equity issuance	(sstky-prstkcy)/l.atq
Net bond issuance	(dltisy-dltry+dlcchy)/l.atq
Other cash	(sppey+sivy)/l.atq
Tobin's Q	(atq-me-ceqq)/l.atq

B Consistency of the OLS estimator with presence of spatial dependence in panel data

In this section, I extend the idea from Rüttenauer (2022) and Pace and LeSage (2010) to the panel data framework. I suppose the $N \times N$ commonality score matrix $\boldsymbol{W}$ is exogenously determined, observed, and time-invariant. For simplicity of notation, I assume the dataset is already demeaned on the corresponding fixed effects. The data-generating process of $\{\boldsymbol{Y_i}\}$ follows

$$\begin{aligned}
&\boldsymbol{Y_i} = \rho \sum_j W_{ij} \boldsymbol{Y_j} + \boldsymbol{X_i}\beta + \sum_j W_{ij} \boldsymbol{X_i}\theta + \boldsymbol{\varepsilon_i} \\
&\boldsymbol{\varepsilon_i} = \lambda \sum_j W_{ij} \boldsymbol{\varepsilon_j} + \boldsymbol{u_i} \\
&\boldsymbol{u_i} | \boldsymbol{X_i} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_i})
\end{aligned} \tag{B.1}$$

where $\boldsymbol{\Sigma_i}$ is a $T \times T$ covariance matrix of error terms u_{it} on time dimension.

Define a $NT \times 1$ vector $\boldsymbol{Y} = \text{vec}(\boldsymbol{Y_i})$ in which the first T elements $(Y_{0*N+1}, \ldots, Y_{0*N+T})$ denote the values of Firm 1 from period 1 to T, the next T elements $(Y_{1*N+1}, \ldots, Y_{1*N+T})$ denote the values of Firm 2 from period 1 to T, so on. Then equation B.1 can be rewritten as

$$\boldsymbol{Y} = \left(\boldsymbol{I_{NT}} - \rho \widetilde{\boldsymbol{W}}\right)^{-1} \left(\boldsymbol{X}\beta + \widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{I_{NT}} - \lambda \widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u}\right) \tag{B.2}$$

where $\boldsymbol{X} = \text{vec}(\boldsymbol{X_i})$, $\boldsymbol{u} = \text{vec}(\boldsymbol{u_i})$, and $\widetilde{\boldsymbol{W}} = \boldsymbol{I_T} \otimes \boldsymbol{W}$.

Now we look at the OLS estimator. Since the data is already demeaned, the OLS estimator here is equivalent to the TWFE estimator of the undemeaned dataset.

$$\begin{aligned}
\hat{\beta}^{\text{OLS}} &= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\
&= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \left(\boldsymbol{I_N} - \rho \widetilde{\boldsymbol{W}}\right)^{-1} \left(\boldsymbol{X}\beta + \widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{I_N} - \lambda \widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u}\right) \\
&= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \widetilde{\boldsymbol{R}}_\rho \boldsymbol{X}\beta + \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \widetilde{\boldsymbol{R}}_\rho \widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \widetilde{\boldsymbol{R}}_\rho \widetilde{\boldsymbol{R}}_\lambda \boldsymbol{u}
\end{aligned} \tag{B.3}$$

where $\widetilde{\boldsymbol{R}}_\rho = \left(\boldsymbol{I_{NT}} - \rho \widetilde{\boldsymbol{W}}\right)^{-1}$ and $\widetilde{\boldsymbol{R}}_\lambda = \left(\boldsymbol{I_{NT}} - \lambda \widetilde{\boldsymbol{W}}\right)^{-1}$.

Because $\boldsymbol{u_i}|\boldsymbol{X_i}$ and $\boldsymbol{u_j}|\boldsymbol{X_j}$ are independently distributed and $\mathbb{E}\left[\boldsymbol{u_i}|\boldsymbol{X_i}\right] = 0$, we can find that

$$\lim_{NT \to \infty} \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \widetilde{\boldsymbol{R}}_\rho \widetilde{\boldsymbol{R}}_\lambda \boldsymbol{u} = 0$$

B.1 Spatial-independent explanatory variables

Assume $\boldsymbol{X_i}$ identically and independently follow standard normal distribution, according to Barry and Pace (1999) and Girard (1989), we can obtain the expectation and the variance of $\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}$ for any real symmetric matrix $\boldsymbol{A}$:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{\mathrm{tr}(\boldsymbol{A})}{NT} \equiv \mu_A \\
\sigma^2\left[\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{2\mathrm{tr}(\boldsymbol{A}^2)}{(NT)^2} \\
\mathbb{E}\left[\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \mu_A \\
\sigma^2\left[\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{2\mathrm{Var}(\lambda_A)}{NT+2} \equiv d_A^2
\end{aligned} \tag{B.4}$$

where $\mathrm{Var}(\lambda_A)$ is the "population variance" of eigenvalues of matrix $\boldsymbol{A}$, i.e. $\mathrm{Var}(\lambda_A) = \sum_i\sum_t(\lambda_{it}-\mu_A)^2/NT$ and λ_i are eigenvalues of matrix $\boldsymbol{A}$.

Recall $\widetilde{\boldsymbol{W}} = \boldsymbol{I_T}\otimes\boldsymbol{W}$. Therefore

$$\begin{aligned}
\mathrm{tr}\left(\widetilde{\boldsymbol{R}}_\rho\right) =& \mathrm{tr}\left(\boldsymbol{I_T}\otimes(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-1}\right) \\
=& T\cdot\mathrm{tr}\left(\boldsymbol{I_T}\otimes(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-1}\right) \equiv T\cdot\mathrm{tr}\left(\boldsymbol{R}_\rho\right) \\
\mathrm{Var}\left(\lambda_{\widetilde{W}}\right) =& \frac{\sum_i\sum_t(\lambda_{it}-\mu_{\widetilde{W}})^2}{NT} \\
=& \frac{\sum_i(\lambda_i-\mu_W)^2}{N} = \mathrm{Var}(\lambda_W)
\end{aligned} \tag{B.5}$$

When N is fixed but T goes to infinite, we can immediately find that $\lim_{T\to\infty} d_A^2 \to 0$ for any real symmetric matrix $\boldsymbol{A}$. Therefore,

$$\lim_{T\to\infty}\hat{\beta}^{\mathrm{OLS}} \xrightarrow{p} \frac{\mathrm{tr}(\boldsymbol{R_\rho})}{N}\beta + \frac{\mathrm{tr}(\boldsymbol{R_\rho W})}{N}\theta \tag{B.6}$$

When N goes to infinite, $\sigma^2[\boldsymbol{X_i}^\top\boldsymbol{X_i}] \to 0$ because $\boldsymbol{X_i}$ are independently and identically distributed. According to Slutsky's theorem,

$$\lim_{N\to\infty}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widetilde{\boldsymbol{R}}_\rho\boldsymbol{X} = \frac{\mathrm{tr}(\boldsymbol{R_\rho})}{N}$$

Then

$$\lim_{N\to\infty} \hat{\beta}^{\text{OLS}} \xrightarrow{p} \frac{\operatorname{tr}(\boldsymbol{R_\rho})}{N}\beta + \frac{\operatorname{tr}(\boldsymbol{R_\rho W})}{N}\theta \tag{B.7}$$

B.2 Spatial-dependent explanatory variables

If explanatory variables $\boldsymbol{X_i}$ are spatially correlated, the OLS estimator is no longer consistent in estimating the direct effects. When $\boldsymbol{X_i}|\boldsymbol{W}$ and $\boldsymbol{\varepsilon}|\boldsymbol{W}$ are not independent, there exists omitted variable bias which leads to the inconsistency of the OLS estimator.

Specifically, I assume $\boldsymbol{X_i}$ follows the data-generating process:

$$\boldsymbol{X_i} = \phi \sum_j W_{ij}\boldsymbol{X_j} + \boldsymbol{Z_i} \tag{B.8}$$

where $\boldsymbol{Z_i}$ is i.i.d. normal distribution. Similarly, we stack $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ and denote $\boldsymbol{X} = \operatorname{vec}(X_i)$ and $\boldsymbol{Z} = \operatorname{vec}(Z_i)$. We can rewrite it as $\boldsymbol{X} = \widetilde{\boldsymbol{R}}_\phi \boldsymbol{Z}$ where $\widetilde{\boldsymbol{R}}_\phi = \left(\boldsymbol{I_{NT}} - \phi\widetilde{\boldsymbol{W}}\right)^{-1}$.

Following Pace and LeSage (2010), we can find that for any real symmetric matrix $\boldsymbol{A}$

$$\begin{aligned}
&\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\\
=&\left(\boldsymbol{Z}^\top\widetilde{\boldsymbol{R}}_\phi^\top\widetilde{\boldsymbol{R}}_\phi\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\widetilde{\boldsymbol{R}}_\phi^\top\boldsymbol{A}\widetilde{\boldsymbol{R}}_\phi\boldsymbol{Z}\\
=&\frac{\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\widetilde{\boldsymbol{R}}_\phi^\top\boldsymbol{A}\widetilde{\boldsymbol{R}}_\phi\boldsymbol{Z}}{\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\widetilde{\boldsymbol{R}}_\phi^\top\widetilde{\boldsymbol{R}}_\phi\boldsymbol{Z}}
\end{aligned}$$

When N is fixed but T goes to infinity, equation B.4 and B.5 guarantee that both the variance of the nominator and denominator converges to zero. Using Slutsky's theorem, we have

$$\lim_{T\to\infty}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X} = \frac{\operatorname{tr}(\widetilde{\boldsymbol{R}}_\phi^\top\boldsymbol{A}\widetilde{\boldsymbol{R}}_\phi)}{\operatorname{tr}(\widetilde{\boldsymbol{R}}_\phi^\top\widetilde{\boldsymbol{R}}_\phi)} = \frac{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{A}\boldsymbol{R}_\phi)}{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\phi)}$$

which implies

$$\lim_{T\to\infty}\hat{\beta}^{\text{OLS}} = \frac{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\rho\boldsymbol{R}_\phi)}{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\phi)}\beta + \frac{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\rho W\boldsymbol{R}_\phi)}{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\phi)}\theta \tag{B.9}$$

When N goes to infinity, according to equation B.4

$$\mathrm{Var}\left(\boldsymbol{Z}^\top \widetilde{\boldsymbol{R}}_\phi^\top \widetilde{\boldsymbol{R}}_\phi \boldsymbol{Z}\right) = \frac{2\mathrm{tr}\left(\widetilde{\boldsymbol{R}}_\phi^\top \widetilde{\boldsymbol{R}}_\phi \widetilde{\boldsymbol{R}}_\phi^\top \widetilde{\boldsymbol{R}}_\phi\right)}{(NT)^2} = \frac{2\mathrm{tr}\left(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi \boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi\right)}{N^2} \longrightarrow 0$$

Therefore we can apply Slutsky's theorem and obtain that

$$\lim_{N\to\infty} \hat{\beta}^{\mathrm{OLS}} = \frac{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\rho \boldsymbol{R}_\phi)}{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi)}\beta + \frac{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\rho W \boldsymbol{R}_\phi)}{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi)}\theta \tag{B.10}$$

B.3 Firm-clustered standard error

In this subsection, I calculate the firm-clustered standard error when error terms are spatially correlated. I assume the following data-generating process,

$$\begin{aligned} \boldsymbol{Y_i} &= \boldsymbol{X_i}\beta + \boldsymbol{\varepsilon_i} \\ \boldsymbol{\varepsilon_i} &= \lambda \sum_j W_{ij}\boldsymbol{\varepsilon_j} + \boldsymbol{u_i} \\ \boldsymbol{u_i}|\boldsymbol{X_i} &\overset{\mathrm{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_i}) \end{aligned} \tag{B.11}$$

The OLS estimator here is an unbiased and consistent estimator of β.

$$\hat{\beta}^{\mathrm{OLS}} = \beta + \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon}$$

Then we rearrange the formula as

$$\hat{\beta}^{\mathrm{OLS}} - \beta = \left(\sum_i \boldsymbol{X_i}^\top \boldsymbol{X_i}\right)^{-1} \sum_i \boldsymbol{X_i}^\top \boldsymbol{\varepsilon_i}$$

Define $\boldsymbol{\varepsilon} = \mathrm{vec}(\boldsymbol{\varepsilon_i})$, $\boldsymbol{u} = \mathrm{vec}(\boldsymbol{u_i})$, and $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma_i})$,

$$\boldsymbol{\varepsilon} = \left(\boldsymbol{I}_{NT} - \lambda\widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u} = \widetilde{\boldsymbol{R}}_\lambda \boldsymbol{u} \tag{B.12}$$

where $\widetilde{\boldsymbol{R}}_\lambda = \left(\boldsymbol{I_{NT}} - \lambda\widetilde{\boldsymbol{W}}\right)^{-1}$.

Since $\boldsymbol{u_i}$ and $\boldsymbol{u_j}$ are independent and $\boldsymbol{u_i}$ follows a normal distribution with $\boldsymbol{\Sigma_i}$ as the covariance matrix, we can obtain the variance of $\hat{\beta}^{\text{OLS}}$ by

$$\text{Var}(\hat{\beta}^{\text{OLS}}) = \mathbb{E}\left[\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widetilde{\boldsymbol{R}}_\lambda^\top\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\widetilde{\boldsymbol{R}}_\lambda\boldsymbol{X}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\right]$$

However, the estimated firm-clustered standard error will be

$$\widehat{\text{Var}}^{\text{Firm}}(\hat{\beta}^{\text{OLS}}) = \frac{N-1}{N}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\sum_i\left(\boldsymbol{X}_{\boldsymbol{i}}^\top\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}^\top\boldsymbol{X}_i\right)\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}$$

In empirical research, the autocorrelation of error terms is most likely positive but smaller than 1. I assume that all elements in $\boldsymbol{W}$ are non-negative, the maximum and principle eigenvalue of $\boldsymbol{W}$ equals 1 (if not, we normalize $\boldsymbol{W}$ by its principle eigenvalue) and $\lambda \in (0,1)$. This condition implies

$$\widetilde{\boldsymbol{R}}_\lambda^\top\boldsymbol{\Sigma}\widetilde{\boldsymbol{R}}_\lambda - \text{diag}\left(\mathbb{E}\left[\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}^\top\right]\right) \text{ is a positive definite matrix}$$

where $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} = \boldsymbol{R}_{\lambda,ii}\boldsymbol{u}_{\boldsymbol{i}}$.

Every elements on the main diagonal of $\text{Var}(\hat{\beta}^{\text{OLS}})$ is larger than $\widehat{\text{Var}}^{\text{Firm}}(\hat{\beta}^{\text{OLS}})$. It suggests that firm-clustering standard error will underestimate the variance if spatially correlated error terms exist.

C Standardized Model

Another type of commonly used spatial autoregressive model in the econometrics literature is the standardized spatial autoregressive model. Th model assumes variable X_i follows a stationary Gaussian process. Specifically, the vector of variable $\boldsymbol{X}$ follows the data generating process:

$$\begin{aligned} \boldsymbol{X} &\sim \mathcal{N}(0, \boldsymbol{\Sigma}) \\ \Sigma_{ij} &= \mathrm{e}^{-\delta|s_i - s_j|} \\ s_i &\overset{\text{iid}}{\sim} \mathcal{U}(0,1) \end{aligned} \tag{C.1}$$

where $\delta > 0$ measures the strength of spatial dependence and higher δ indicates a lower spatial dependence.

A standardized spatial autoregressive model is a spatial equivalent of a time-series autoregressive model. Recall time-series AR(1) process follows $y_t = (1-\psi)\mu + \psi y_{t-1} + \varepsilon_t$, and its autocovariance $K_h = \text{cov}(y_t, y_{t-h}) = \psi^h \sigma_y^2$ decreases exponentially with the temporal distance h.

Analogous to the time-series autoregressive model, the standardized model assumes the covariance matrix of the standardized spatial autoregressive model also follows an exponential function whose base is $\mathrm{e}^{-\delta}$ and whose index is the spatial distance $|s_i - s_j|$. Figure C.1 displays the relationship between the strength of spatial dependence δ and average pair-wise correlation. A spatial strength $\delta = 20$ implies an average pair-wise correlation of 0.1, which is equivalent to an average autocovariance of 0.1 in a time-series AR(1) process with $T = 100$ periods and $\rho = \mathrm{e}^{-\delta/(T-1)} = 0.82$.

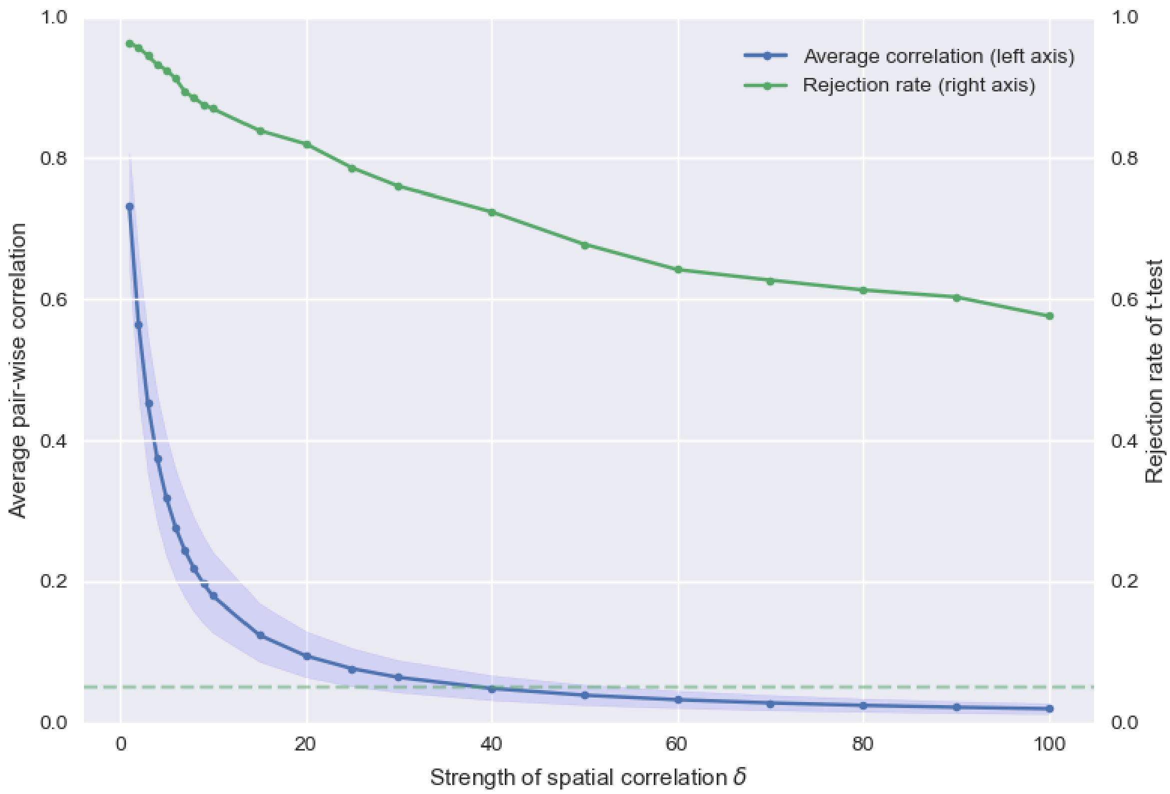


Figure C.1: Average pair-wise correlation in Standardized model

I illustrate the over-rejection problem with the standardized model as well. Similar to figure **??**, figure C.2 presents rejection rates of the coefficient β in the cross-sectional regression $Y_i = \beta X_i + \varepsilon_i$, where ε_i and X_i follow the data-generating process described in equation C.1 and the true value of β is 0. I test the null hypothesis $H_0 : \beta = 0$ against the alternative hypothesis $H_1 : \beta \neq 0$ and run the simulations 199 times. Standard errors are heteroskedasticity-robust and the significance level is 5%. we observe the same pattern as in figure ??, as the variables become more spatially dependent, the rejection rate increases rapidly to 100% and the over-rejection problem always exists as long as spatial dependence is present.

Table C.1: Rejection rates of t-test in the presence of spatial correlation

This table presents rejection rates of the student's t-test for the standardized model (data generating process described in equation C.1). The expectation of X_i is 0. The null hypothesis is $H_0 : \bar{X} = 0$ and the alternative hypothesis is $H_1 : \bar{X} \neq 0$ where $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$. All simulations are run 1000 times. The significance level is 5%.

Parameter δ	Rejection	Parameter δ	Rejection
1	0.963	15	0.839
2	0.956	20	0.820
3	0.945	30	0.760
4	0.933	40	0.724
5	0.925	50	0.678
6	0.913	60	0.642
7	0.894	70	0.627
8	0.885	80	0.613
9	0.876	90	0.603
10	0.870	100	0.576

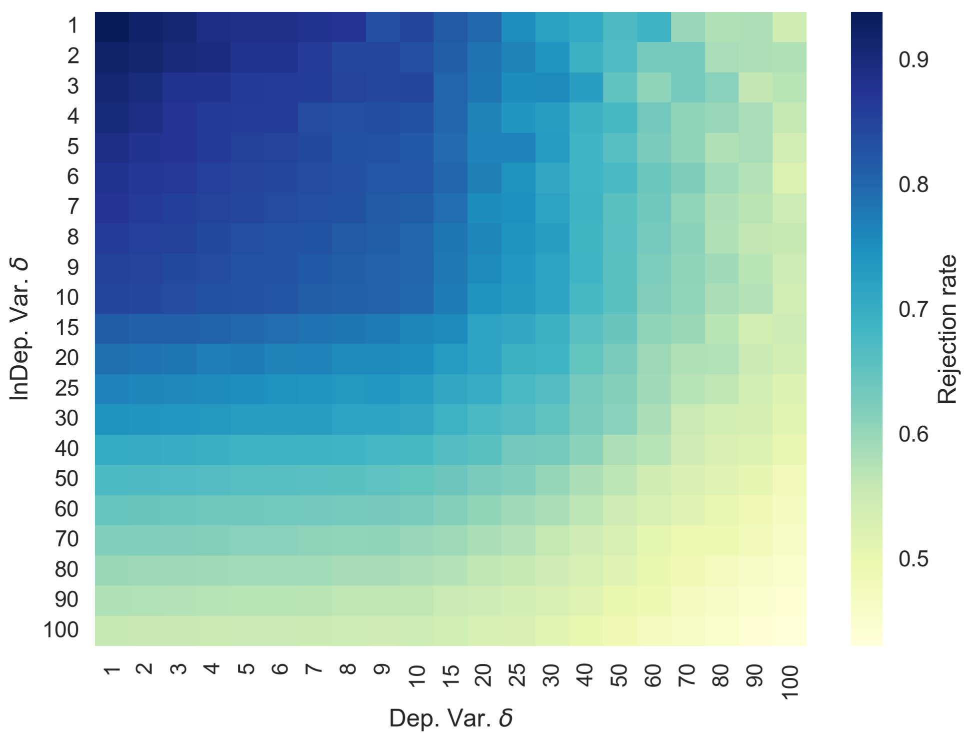


Figure C.2: Rejection rates of cross-sectional regressions in standardized model

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, Andrew C, David F Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics* 144 (2):370–395.

Barry, Ronald Paul and R Kelley Pace. 1999. "Monte Carlo estimates of the log determinant of large sparse matrices." *Linear Algebra and its applications* 289 (1-3):41–54.

Berg, Tobias, Markus Reisinger, and Daniel Streitz. 2021. "Spillover effects in empirical corporate finance." *Journal of Financial Economics* 142 (3):1109–1127.

Cameron, A Colin and Douglas L Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of human resources* 50 (2):317–372.

Campello, Murillo. 2003. "Capital structure and product markets interactions: evidence from business cycles." *Journal of financial economics* 68 (3):353–378.

Carter, Andrew V, Kevin T Schnepel, and Douglas G Steigerwald. 2017. "Asymptotic behavior of at-test robust to cluster heterogeneity." *Review of Economics and Statistics* 99 (4):698–709.

Chevalier, Judith A. 1995. "Do LBO supermarkets charge more? An empirical analysis of the effects of LBOs on supermarket pricing." *The Journal of Finance* 50 (4):1095–1112.

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

De Paula, Aureo, Imran Rasul, and Pedro Souza. 2018. "Recovering social networks from panel data: identification, simulations and an application." .

DeAngelo, Harry and Richard Roll. 2015. "How stable are corporate capital structures?" *The Journal of Finance* 70 (1):373–418.

Djogbenou, Antoine A, James G MacKinnon, and Morten Ørregaard Nielsen. 2019. "Asymptotic theory and wild bootstrap inference with clustered errors." *Journal of Econometrics* 212 (2):393–412.

Donald, Stephen G and Kevin Lang. 2007. "Inference with difference-in-differences and other panel data." *The review of Economics and Statistics* 89 (2):221–233.

Dougal, Casey, Christopher A Parsons, and Sheridan Titman. 2015. "Urban vibrancy and corporate growth." *The Journal of Finance* 70 (1):163–210.

Foucault, Thierry and Laurent Fresard. 2014. "Learning from peers' stock prices and corporate investment." *Journal of Financial Economics* 111 (3):554–577.

Gao, Wenlian, Lilian Ng, and Qinghai Wang. 2011. "Does corporate headquarters location matter for firm capital structure?" *Financial Management* 40 (1):113–138.

Gatchev, Vladimir A, Todd Pulvino, and Vefa Tarhan. 2010. "The interdependent and intertemporal nature of financial decisions: An application to cash flow sensitivities." *The Journal of Finance* 65 (2):725–763.

Girard, A. 1989. "A fast 'Monte-Carlo cross-validation'procedure for large least squares problems with noisy data." *Numerische Mathematik* 56:1–23.

Güner, A Burak, Ulrike Malmendier, and Geoffrey Tate. 2008. "Financial expertise of directors." *Journal of financial Economics* 88 (2):323–354.

Hoberg, Gerard and Gordon Phillips. 2010. "Product market synergies and competition in mergers and acquisitions: A text-based analysis." *The Review of Financial Studies* 23 (10):3773–3811.

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

Huang, Jiyuan and Per Östberg. 2023. "Difference-in-differences with Economic Factors and the Case of Housing Returns." *Swiss Finance Institute Research Paper* (23-55).

Kelly, Bryan T, Hanno N Lustig, and Stijn Van Nieuwerburgh. 2013. "Firm volatility in granular networks." *NBER Working paper* (w19466).

Kelly, Morgan. 2019. "The standard errors of persistence." .

Korajczyk, Robert A and Amnon Levy. 2003. "Capital structure choice: macroeconomic conditions and financial constraints." *Journal of financial economics* 68 (1):75–109.

Leary, Mark T and Michael R Roberts. 2014. “Do peer firms affect corporate financial policy?” *The Journal of Finance* 69 (1):139–178.

MacKinnon, James G, Morten Ørregaard Nielsen, and Matthew D Webb. 2021. “Wild bootstrap and asymptotic inference with multiway clustering.” *Journal of Business & Economic Statistics* 39 (2):505–519.

———. 2023. “Cluster-robust inference: A guide to empirical practice.” *Journal of Econometrics* 232 (2):272–299.

MacKinnon, James G and Matthew D Webb. 2017. “Wild bootstrap inference for wildly different cluster sizes.” *Journal of Applied Econometrics* 32 (2):233–254.

McLean, R David. 2011. “Share issuance and cash savings.” *Journal of Financial Economics* 99 (3):693–715.

Menzel, Konrad. 2021. “Bootstrap with cluster-dependence in two or more dimensions.” *Econometrica* 89 (5):2143–2188.

Menzly, Lior and Oguzhan Ozbas. 2010. “Market segmentation and cross-predictability of returns.” *The Journal of Finance* 65 (4):1555–1580.

Mitton, Todd. 2021. “Economic significance in corporate finance.” *Available at SSRN 3667830* .

———. 2022. “Methodological variation in empirical corporate finance.” *The Review of Financial Studies* 35 (2):527–575.

Moran, Patrick AP. 1950. “Notes on continuous stochastic phenomena.” *Biometrika* 37 (1/2):17–23.

Müller, Ulrich K and Mark W Watson. 2022a. “Spatial correlation robust inference.” *Econometrica* 90 (6):2901–2935.

———. 2022b. “Spatial Correlation Robust Inference in Linear Regression and Panel Models.” *Journal of Business & Economic Statistics* :1–15.

Owen, Art B. 2007. “The pigeonhole bootstrap.” .

Pace, R Kelley and James P LeSage. 2010. "Omitted variable biases of OLS and spatial lag models." *Progress in spatial analysis: Methods and applications* :17–28.

Petersen, Mitchell A. 2008. "Estimating standard errors in finance panel data sets: Comparing approaches." *The Review of financial studies* 22 (1):435–480.

Rüttenauer, Tobias. 2022. "Spatial regression models: a systematic comparison of different model specifications using Monte Carlo experiments." *Sociological Methods & Research* 51 (2):728–759.

Webb, Matthew D. 2023. "Reworking wild bootstrap-based inference for clustered errors." *Canadian Journal of Economics/Revue canadienne d'économique* 56 (3):839–858.

White, Halbert. 2014. *Asymptotic theory for econometricians.* Academic press.