Firm Commonality and Inference in Corporate Finance

Jiyuan Huang*

Job Market Paper

November 2, 2023

Abstract

In this paper, I explore latent connections among firms and their implications on 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 errors in the cross-section. I show that firm-clustered standard errors underestimate the true standard errors up to five-fold. Finally, Finally, I provide a bootstrap solution of standard errors to tackle 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 but are linked to each other through many types of relationships. Empirical literature has found that firms are interconnected across various characteristics, ranging from operational performance such as investment choices (Bustamante and Frésard (2021), Dougal, Parsons, and Titman (2015)) to corporate governance decisions such as executive compensation policies (Aggarwal and Samwick (1999)).

The interconnections between firms indicate that firm behaviors must be understood within the context of a broader interconnected corporate landscape, where decisions ripple through networks of firms, emphasizing the importance of recognizing these interrelationships in corporate finance. To better understand where the interconnections come from, I present at least four channels via which one firm's characteristics are linked with other firms.

Firstly, competition. Firms do not operate in a vacuum, instead, they have to compete with rivals, and thus their decisions take into account the characteristics of other firms. For example, in a Cournot competition model, one firm's output depends not only on its own characteristics but also on the actions of its rivals. Empirically, Chevalier (1995) and Campello (2003) find that firms react to shocks in external financing by adjusting their investment in market share, but the outcome from such actions depends on the financial structures of their industry rivals.

Secondly, peer effects. A firm's decisions are rarely autonomous, on the opposite, they are often influenced by the strategies and actions of peer entities within their business ecosystem. Such influences can arise from aspects such as social interactions (Gao, Ng, and Wang (2011)), agglomeration economies (Dougal, Parsons, and Titman (2015)), or learning (Leary and Roberts (2014)) and the consequences of these effects range from firm financing policies to investment choices.

Thirdly, supply chain effects. Analogous to peer effects, firms that are economically related suppliers and customers naturally affect each other's operations through the supply chain. Many empirical studies show that the customer–supplier links between firms provide large and robust predictability of stock returns (Cohen, Frazzini, and Malloy (2008), Menzly and Ozbas (2010)). Kelly, Lustig, and Van Nieuwerburgh (2013) estimates a structural model where shocks are transmitted from customers to suppliers in a production network and reveals the importance of a strong network effect without which the model cannot account for the large dispersion in firm volatilities.

Fourthly, common factors. The effect of interconnectedness does not necessarily come in the form of direct impact from other companies. It can come from the fact that firms face the same economic environment and have similar exposures to common factors. DeAngelo and Roll (2015) and Korajczyk and Levy (2003) show that factor structure explains a nontrivial part of the variation in firm capital structure. [Huang and Östberg (2023) illustrates]

While empirical literature has been looking into some aspects of the interconnections of firms, they are mostly under specific settings and focus on localized effects of firms in the immediate neighborhood. In this paper, I investigate a much broader corporate landscape with a focus on its impact on statistical inferences in corporate finance. The broad corporate network used in this paper has at least two advantages. On the one hand, instead of seeing the connections between firms as a binary relation, I take the intensity of the connections into account. On the other hand, the connections between firms are no longer sparse which matches the real-world data better.

We can imagine the corporate landscape as a virtual space where each firm lives inside and is connected with other firms nearer or further. One can think of this as a geographic space where many lands live inside and each land has different distances from other lands. All together, these firms make up the corporate landscape just as the lands make up the geographic landscape.

[There are mainly two ways in the literature that address the interconnectedness stemming from geographic proximity.]

[references about spatial regressions, preferably chronically]

[The contribution of this paper: How are we different than previous work? We study a broader landscape and the comovements of firm characteristics and its consequent impact on statistical inference.]

[motivate spatial regression]

[how to distinguish from peer effect non-sparse flexible intensity not for causal statement, but general impact on regression Regression technique differently? especially when using IV? linear-in-mean model]

2 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 follow Hoberg and Phillips (2010, 2016) because their method is sufficient as well as simple and their dataset is publicly accessible.

2.1 Data of Firm Characteristics

In this section, I provide details about the firm characteristic variables. Firm fundamentals are from Compustat from 1962 to 2022. The definition of firm characteristics can be found in [Appendix]. [Table] shows the summary statistics.

Later in section **??**, I extract a balanced panel subsample from the whole dataset. During this process, xxx firms are dropped.

2.2 Measuring Commonality

As mentioned above, the measurement of firm commonality follows Hoberg and Phillips (2010, 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[1]. 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.

[1]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.

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

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

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

[why this method is valid] The formula assigns each firm a virtual location based on its business description. Each firm has a unique location and its own neighbors in the corporate landscape based on commonality score. [Higher commonality score implies higher Comovement]

The commonality between two firms can come from different channels. Firstly, [First, competitors. Second, learning. Third, supply chain. Fourth, common factor loadings. Literature reviews.]

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[2]. The simplification is appropriate since $w_{ij,t}$ do not vary a lot across t. Table 1 shows the descriptive statistics of commonality scores and within-firm-pair variance contributes to only 15% to 25%.

Table 1: 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.

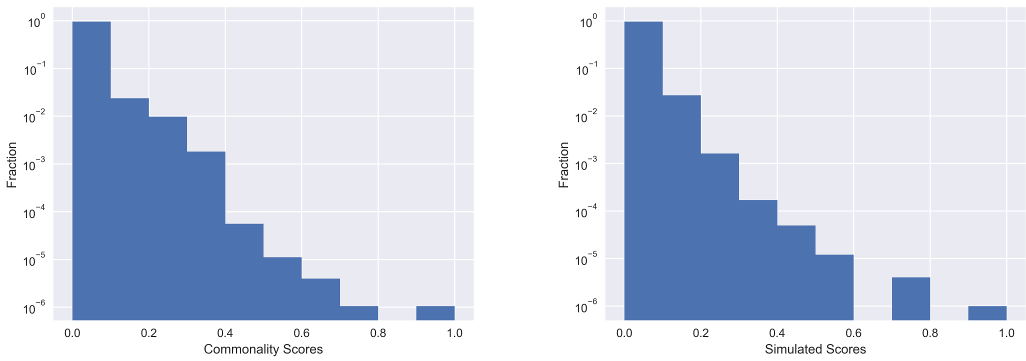
Sample range	Observations	Firms	Mean	Total Var.	Btw. Var.	Between %
Compustat	983,570,310	18633	0.0178	0.0014	0.0012	86.19%
Russell 3000	79,009,590	2766	0.0178	0.0013	0.0011	84.39%
S&P 1500	66,746,114	1929	0.0167	0.0010	0.0008	77.48%

Figure 1 displays the distribution of pair-wise commonality scores of Russell 3000

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

constituents. 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.

[Simulation gives a similar results. 2000 Firms, 1000 length dictionary, firm randomly choose 1-30 words]



(a) Real Commonality Scores

(b) Simulated Commonality Scores

Figure 1: Histogram of Commonality Scores

A natural question is how the commonality score relates to the SIC code, which also reflects partly how firms are related. 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.[reference quote]

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. Finance, Insurance, Real Estate are deeply linked industries. (In fact, they are so tightly linked that we even have an abbreviation to describe them – FIRE emoji emoji)

[Last but not least, the heatmap shows that the two-digit SIC code is not a reliable measurement of firm commonality.]

[Descprition of heatmap]

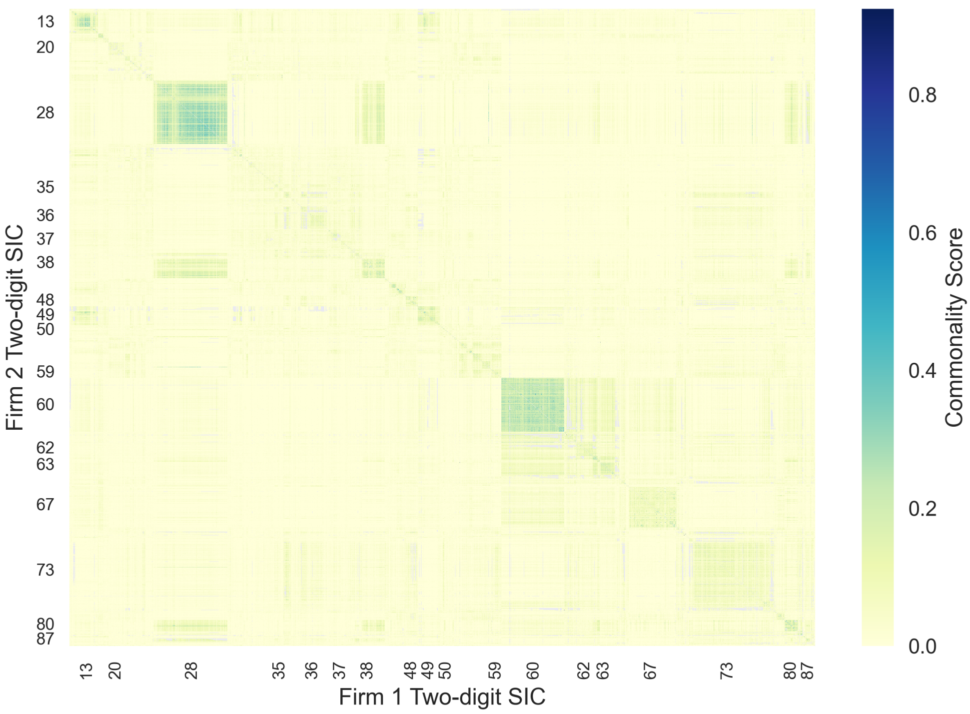


Figure 2: Heatmap of Commonality Scores

3 Firm commonality reflects comovement of firm characteristics

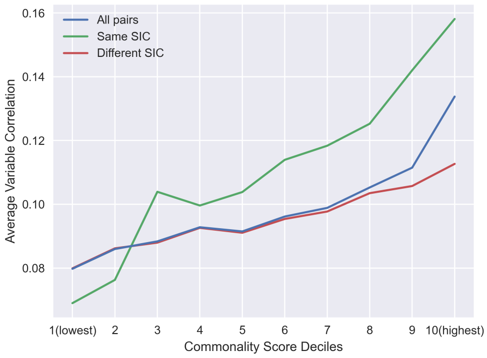
3.1 Pair-wise correlation of firm characteristics

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.

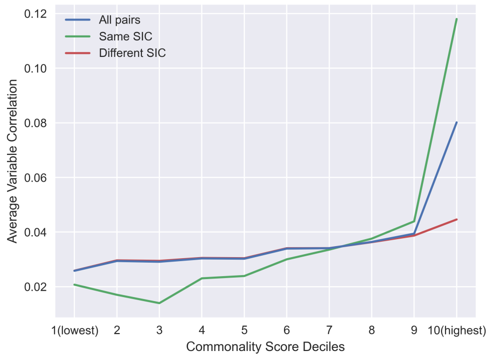
Empirically, I divide firm pairs into 10 decile groups based on their commonality scores. Then I calculate the average pair-wise correlation of firm characteristics in each decile group. Figure 3 plots the average correlation of different firm characteristics for each commonality score group.

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.

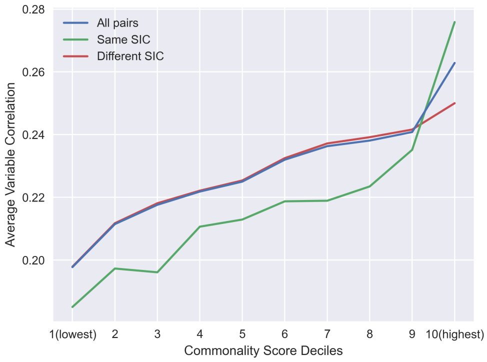
For each firm pair, I categorize it into the "same SIC" group and the "different SIC" group and then I divide them into commonality score decile groups and perform the same exercise as described in the last paragraph. This analysis yields two findings. On one hand, the SIC code is not a reliable measurement 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.



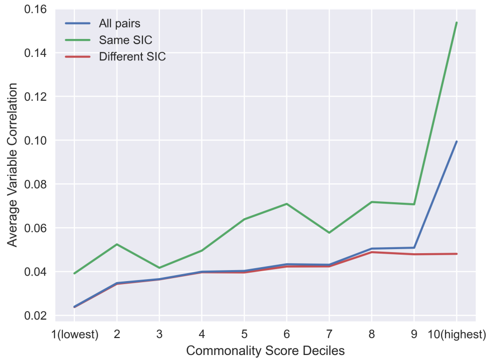
(a) Book-to-Market Ratio



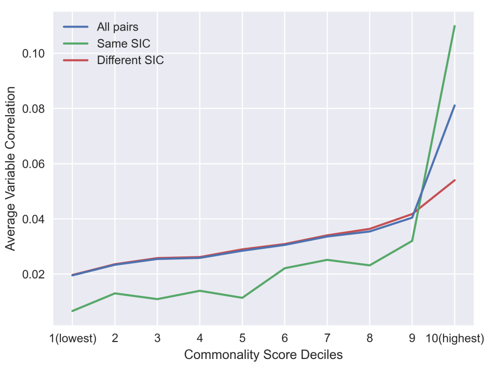
(b) Cash Holding



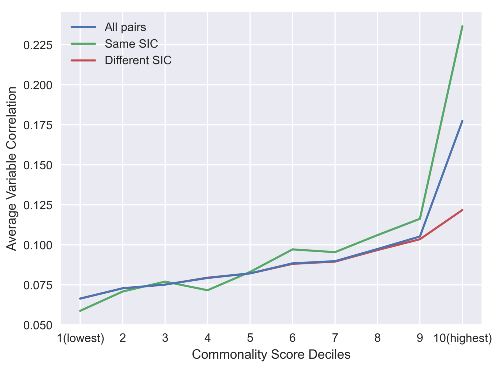
(c) Capital Expenditure



(d) Book Leverage



(e) Return of Asset



(f) Tobin's Q

Figure 3: Firm Characteristics Correlation Increases with Commonality Score

3.2 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 two methods, the standardized spatial autoregressive model, and the spatial autoregressive model in the corporate landscape.

The standardized spatial autoregressive model assumes variable X_i follows a stationary Gaussian process. Specifically, the vector of variable $\boldsymbol{X}$ follows the data generating process:

$$\begin{aligned} \boldsymbol{X} &\sim \mathcal{N}(0, \boldsymbol{\Sigma}) \\ \Sigma_{ij} &= \mathrm{e}^{-\delta|s_i - s_j|} \\ s_i &\overset{\text{iid}}{\sim} \mathcal{U}(0,1) \end{aligned} \tag{3}$$

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

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 $K_h = \mathrm{cov}(y_t, y_{t-h}) = \psi^h \sigma_y^2$ decreases exponentially with the temporal distance h.

Analogous to the time-series autoregressive model, I assume 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 4(4a displays the relationship between the strength of spatial correlation 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 $\rho = \mathrm{e}^{-\delta/(T-1)} = 0.82$

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:

$$\begin{aligned} X_i &= \sum_j \rho W_{ij} X_j + Z_i \\ Z_i &\overset{\text{iid}}{\sim} \mathcal{N}(0,1) \end{aligned} \tag{4}$$

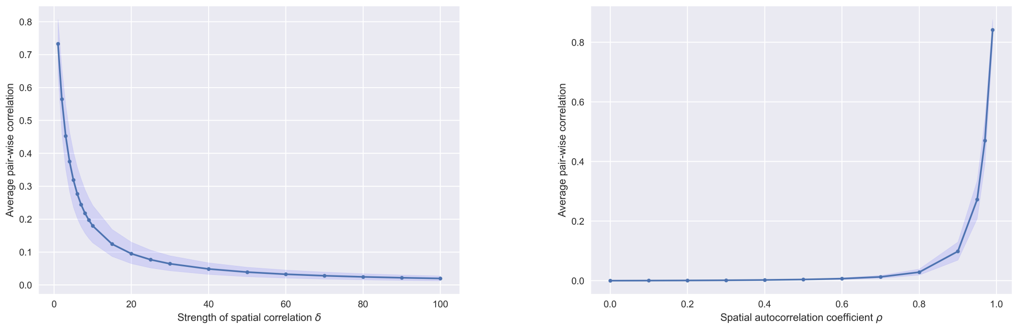
where ρ is the spatial autoregression coefficient with respect to commonality metrics W. Equation 4 can be rewritten in matrix form,

$$\boldsymbol{X} = \rho \boldsymbol{W} \boldsymbol{X} + \boldsymbol{Z}$$
$$\boldsymbol{X} = (\boldsymbol{I_n} - \rho \boldsymbol{W})^{-1} \boldsymbol{Z}$$

Let $\boldsymbol{R_\rho} = (\boldsymbol{I_n} - \rho \boldsymbol{W})^{-1}$. If $\boldsymbol{R_\rho} \boldsymbol{R_\rho}^\top = \boldsymbol{\Sigma}$, two models are equivalent. In practice, due to commonality scores do not work as the inverse of a distance measure, these two models are generally not equivalent.

rho = 0.8, average pairwise corr = 0.028; delta = 70, average pairwise corr = 0.028; T = 100, time series autocorrelation 0.49, average pairwise corr = 0.28

[Some explanation here]



(a) Standardized model (b) Corporate landscape

Figure 4: Average pair-wise correlation versus the strength of autoregression

3.3 Over-rejection

3.4 Moran's I statistics

[Now that we have an intuitive understanding of spatial autoregression, we examine the coefficient ρ formally using Moran's I statistics[citation]. The null hypothesis is $\rho = 0$. Conduct permutation procedure to obtain the distribution of Moran's I statistics under the null hypothesis.]

[Reprahse the following paragraph: "Global indices of spatial autocorrelation have been used to evaluate the degree to which similar observations tend to occur near each other [1–4]. Spatial autocorrelation among disease counts or incidence proportions may

Table 2: 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 1000 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

reflect real association between cases due to infection, or perceived association based on a spatial aggregation of similar values. Moran's I [5] is a widely used global index that measures the similarity for values in neighboring places from an overall mean value and reflects a spatially weighted form of Pearson's correlation coefficient."]

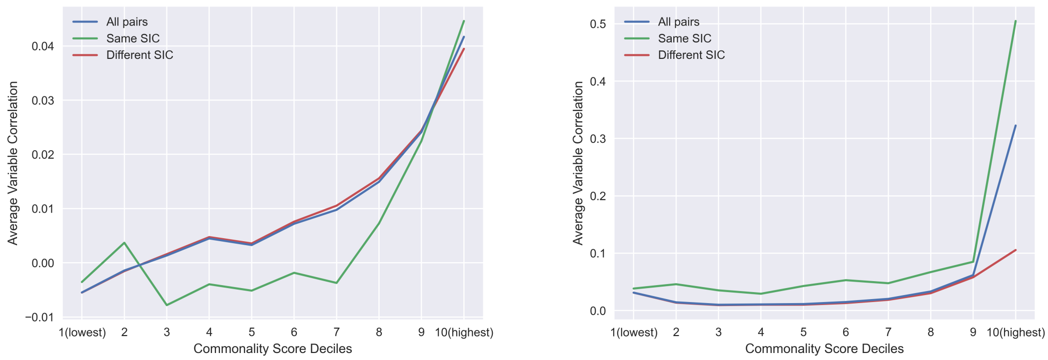
[So far, we have established the correlation between firm characteristics and commonality scores. Now we show that neglecting this relationship in the regression leads to biased results.]

4 TWFE Residual Correlates with Firm Commonality

In this section, I provide evidence that a two-way fixed effect estimator without accounting for firm commonality leads to over-rejection.

First, 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.

However, the residual shows a strong positive correlation with the commonality score. 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.



(a) Cash Holding

(b) Capital Expenditure

Figure 5: Residuals of TWFE Estimator Correlates with Commonality Score

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, since we only have a few decades of data, this is normally not feasible.[reference]

In the later section, Table **??** shows the over-rejection rate in detail. We will see that neither way of standard error clustering can mitigate the problem.

[If we fail to take the graph into account, then we have a bias in our estimator.]

5 Empirical Example on Capital Expenditure

6 Potential solution

6.1 Spatial regression models

6.2 Bootstrapping standard errors

7 Consistency of the OLS estimator in Panel Setting

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

$$
\begin{aligned}
\boldsymbol{Y_i} &= \rho \sum_j W_{ij} \boldsymbol{Y_j} + \boldsymbol{X_i}\beta + \sum_j W_{ij} \boldsymbol{X_i}\theta + \boldsymbol{\varepsilon_i} \\
\boldsymbol{\varepsilon_i} &= \lambda \sum_j W_{ij} \boldsymbol{\varepsilon_j} + \boldsymbol{u_i} \\
\boldsymbol{u_i}|\boldsymbol{X_i} &\overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_i})
\end{aligned}
\tag{5}
$$

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 15 can be rewritten as

$$
\boldsymbol{Y} = \left(\boldsymbol{I_{NT}} - \rho\widetilde{\boldsymbol{W}}\right)^{-1} \left(\boldsymbol{X}\beta + \widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{I_{NT}} - \lambda\widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u}\right) \tag{6}
$$

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.

$$
\begin{aligned}
\hat{\beta}^{\text{OLS}} &= \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{Y} \\
&= \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top \left(\boldsymbol{I_N} - \rho\widetilde{\boldsymbol{W}}\right)^{-1} \left(\boldsymbol{X}\beta + \widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{I_N} - \lambda\widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u}\right) \\
&= \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widetilde{\boldsymbol{R}}_\rho\boldsymbol{X}\beta + \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widetilde{\boldsymbol{R}}_\rho\widetilde{\boldsymbol{W}}\boldsymbol{X}\theta + \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widetilde{\boldsymbol{R}}_\rho\widetilde{\boldsymbol{R}}_\lambda\boldsymbol{u}
\end{aligned}
\tag{7}
$$

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

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

7.1 Independent variables are spatial independent

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 $\boldsymbol{A}$:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{\operatorname{tr}(\boldsymbol{A})}{NT} \equiv \mu_A \\
\sigma^2\left[\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{2\operatorname{tr}(\boldsymbol{A}^2)}{(NT)^2} \\
\mathbb{E}\left[\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \mu_A \\
\sigma^2\left[\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}\right] &= \frac{2\operatorname{Var}(\lambda_A)}{NT+2} \equiv d_A^2
\end{aligned} \tag{8}$$

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

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

$$\begin{aligned}
\operatorname{tr}\left(\widetilde{\boldsymbol{R}}_\rho\right) =& \operatorname{tr}\left(\boldsymbol{I_T}\otimes(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-1}\right) \\
=& T\cdot\operatorname{tr}\left(\boldsymbol{I_T}\otimes(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-1}\right) \equiv T\cdot\operatorname{tr}\left(\boldsymbol{R}_\rho\right) \\
\operatorname{Var}\left(\lambda_{\widetilde{W}}\right) =& \frac{\sum_i\sum_t(\lambda_{it}-\mu_{\widetilde{W}})^2}{NT} \\
=& \frac{\sum_i(\lambda_i-\mu_W)^2}{N} = \operatorname{Var}(\lambda_W)
\end{aligned} \tag{9}$$

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,

$$\lim_{T\to\infty}\hat{\beta}^{\text{OLS}} \xrightarrow{p} \frac{\operatorname{tr}(\boldsymbol{R_\rho})}{N}\beta + \frac{\operatorname{tr}(\boldsymbol{R_\rho}\boldsymbol{W})}{N}\theta \tag{10}$$

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,

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

Then

$$\lim_{N\to\infty} \hat{\beta}^{\text{OLS}} \xrightarrow{p} \frac{\operatorname{tr}(\boldsymbol{R}_{\boldsymbol{\rho}})}{N}\beta + \frac{\operatorname{tr}(\boldsymbol{R}_{\boldsymbol{\rho}}\boldsymbol{W})}{N}\theta \tag{11}$$

7.2 Independent variables are spatial dependent

If independent 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:

$$\boldsymbol{X_i} = \phi \sum_j W_{ij}\boldsymbol{X_j} + \boldsymbol{Z_i} \tag{12}$$

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 8 and 9 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

$$\lim_{T\to\infty}\hat{\beta}^{\text{OLS}} = \frac{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\rho\boldsymbol{R}_\phi)}{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\phi)}\beta + \frac{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\rho W\boldsymbol{R}_\phi)}{\operatorname{tr}(\boldsymbol{R}_\phi^\top\boldsymbol{R}_\phi)}\theta \tag{13}$$

When N goes to infinity, according to equation 8

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

Therefore we can apply Slutsky's theorem and obtain that

$$\lim_{N\to\infty} \hat{\beta}^{\mathrm{OLS}} = \frac{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\rho \boldsymbol{R}_\phi)}{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi)}\beta + \frac{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\rho W \boldsymbol{R}_\phi)}{\mathrm{tr}(\boldsymbol{R}_\phi^\top \boldsymbol{R}_\phi)}\theta \tag{14}$$

7.3 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,

$$\begin{aligned} \boldsymbol{Y_i} &= \boldsymbol{X_i}\beta + \boldsymbol{\varepsilon_i} \\ \boldsymbol{\varepsilon_i} &= \lambda \sum_j W_{ij}\boldsymbol{\varepsilon_j} + \boldsymbol{u_i} \\ \boldsymbol{u_i}|\boldsymbol{X_i} &\overset{\mathrm{iid}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_i}) \end{aligned} \tag{15}$$

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 $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma_i})$,

$$\boldsymbol{\varepsilon} = \left(\boldsymbol{I}_{NT} - \lambda\widetilde{\boldsymbol{W}}\right)^{-1} \boldsymbol{u} = \widetilde{\boldsymbol{R}}_\lambda \boldsymbol{u} \tag{16}$$

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 $\hat{\beta}^{\mathrm{OLS}}$ by

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

However, the estimated firm-clustered standard error will be

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

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

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

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

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

References

Aggarwal, Rajesh K and Andrew A Samwick. 1999. "The other side of the trade-off: The impact of risk on executive compensation." *Journal of political economy* 107 (1):65–105.

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.

Bustamante, M Cecilia and Laurent Frésard. 2021. "Does firm investment respond to peers' investment?" *Management Science* 67 (8):4703–4724.

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

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, Andrea Frazzini, and Christopher Malloy. 2008. "The small world of investing: Board connections and mutual fund returns." *Journal of Political Economy* 116 (5):951–979.

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

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

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

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

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).

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.

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

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.

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