SFB 649 Discussion Paper 2009-025

Measuring the effects of geographical distance on stock market correlation

Stefanie Eckel*
Gunter Löffler*
Alina Maurer*
Volker Schmidt*

* Universität Ulm, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Measuring the effects of geographical distance on stock market correlation

Stefanie Eckel* Gunter Löffler† Alina Maurer‡
Volker Schmidt§

This Version: Oktober 2008

Abstract: Recent studies suggest that the correlation of stock returns increases with decreasing geographical distance. However, there is some debate on the appropriate methodology for measuring the effects of distance on correlation. We modify a regression approach suggested in the literature and complement it with an approach from spatial statistics, the mark correlation function. For the stocks contained in the S&P 500 that we examine, both approaches lead to similar results: correlation increases with decreasing distance. Contrary to previous studies, however, we find that differences in distance do not matter much once the firms' headquarters are more than 40 miles apart, or separated through a federal border. Finally, we show through simulations that distance can significantly affect portfolio risk. Investors wishing to exploit local information should be aware that local portfolios are relatively risky.

Keywords: stock returns, residual correlation, mark correlation function, geographical comovement, portfolio analysis

JEL Classification: R12, G11, G14

Acknowledgements: This research was supported by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 649 "Economic Risk". S. E. is supported by a grant of the graduate college 1100 at Ulm University.

* Ulm University, Institute of Stochastics, Helmholtzstr. 18, 89069 Ulm, Germany

† Ulm University, Institute of Finance, Helmholtzstr. 18, 89069 Ulm, Germany

‡ Corresponding author: Alina Maurer, Institute of Finance, University of Ulm, 89069 Ulm, Germany. Phone: 0049-731-5023591, email: alina.maurer@uni-ulm.de

§ Ulm University, Institute of Stochastics, Helmholtzstr. 18, 89069 Ulm, Germany

1 Introduction

Several studies have documented that investment decisions are affected by geographical location within a country. Both institutional investors (e.g. Coval and Moskovitz, 2001) as well as retail investors (e.g. Ivkovic and Weisbrenner, 2005) allocate a disproportionately large fraction of their portfolios to firms that are close to their offices or homes. Possible reasons for such investment patterns are informational advantages and behavioral preferences for familiarity. Provided that investor sentiment or information arrival is locally correlated, geographical distance could affect return correlations across stocks. Pirinsky and Wang (2006) and Barker and Loughran (2007) test this conjecture and conclude that the correlation of stock returns increases with decreasing distance.

However, there is some debate on the appropriate methodology for measuring the effect of distance on correlation. Barker and Loughran (2007), for example, question the approach of Pirinsky and Wang (2006). One contribution of our paper is therefore methodological. We modify the regression analysis suggested by Barker and Loughran (2007) and complement it with an approach from spatial statistics, the mark correlation function. For the stocks contained in the Standard and Poor's 500 index (S&P 500) that we examine, both approaches lead to similar results: correlation increases with decreasing distance. Contrary to previous studies, however, we obtain that differences in distance do not matter much once the firms' headquarters are more than 40 miles apart. Also, proximity only leads to a higher correlation if two firms are located in the same federal state. This finding is not only relevant for correlation modelling. It also suggests a new route for uncovering the drivers of local correlation effects. Federal borders might well dampen investor interaction and information flow, which means that our results are consistent with the explanations favored in the literature. However, the strength of the finding suggests that one should also investigate the role of policy–driven differences between states. Differences in taxes, industrial policy or infrastructure could lead to fundamental proximity effects which do not require local correlation of investor behavior.

Using a simulation study, we then show that distance can significantly affect portfolio risk. If the number of stocks in a portfolio is fixed, portfolio risk increases considerably if the portfolio is composed of nearby firms. The increase in risk is large enough to neutralize the return advantage on local investments that has been documented by Ivkovic and Weisbrenner (2005).

The remainder of the paper is organized as follows. Section 2 describes the data, Section 3 the methodology. The empirical results on the relationship between distance and correlation are summarized in Section 4. Section 5 shows how distance affects portfolio risk. Section 6 concludes.

2 Data

Our analysis is based on firms contained in the S&P 500. Monthly stock returns are obtained from the Center for Research in Security Prices (CRSP). Address information (state, city and five digit zip–code) for the location of headquarters along with the Standard Industrial Classification (SIC) code are from the annual COMPUSTAT Database. Information on population numbers for each location is taken from http://www.city-data.com, which derive from the census 2000 carried out by the U.S. Census Bureau.

We restrict our sample to firms with headquarters located in the USA, which reduces the number of firms to 465 from a total of 500 firms in the S&P 500. We consider monthly stock returns for the years 2000–2004 resulting in a total of 27900 firm–months, 60 months for each firm. The final sample is constructed by generating a cross–product of the set of considered firms. Due to symmetry reasons each pair of firms only needs to be considered once. Hence the final sample compasses 107880 = 465(465-1)/2 pairwise observations.

For each pair of firms, we compute the distance between headquarter locations based on the geographical coordinates of the five digit zip–code. Applying the correction for the curvature of the earth, the distance $d(i,j)$ between two firms i and j is given by

$$d(i,j) = \rho \arccos(\cos(lat_i)\cos(lat_j)\cos(long_i - long_j) + \sin(lat_i)\sin(lat_j)),$$

where ρ is the earth circumference ($\rho = 3959.871$ miles) and $(lat_i, long_i)$ are the geographical coordinates, i.e. the latitude and longitude in radian (Zwillinger, 1995), of firm i. Geographical coordinates for US zip–codes can be obtained from http://www.census.gov.

Figure 1 shows the distribution of the population numbers of locations as well as of the distances between firm pairs. 36% of firms are located in cities with less than 100000 inhabitants, which we will refer to as small cities. The mean population number of headquarter locations is 1.1 million. A map cutout of the firm locations for three federal states can be found in Figure 2. Headquarter cities exceeding a population of 100000 are marked with $+$, otherwise with $-$. Note that two cities can be close, but still located in different federal states which will also be subject to later analysis.

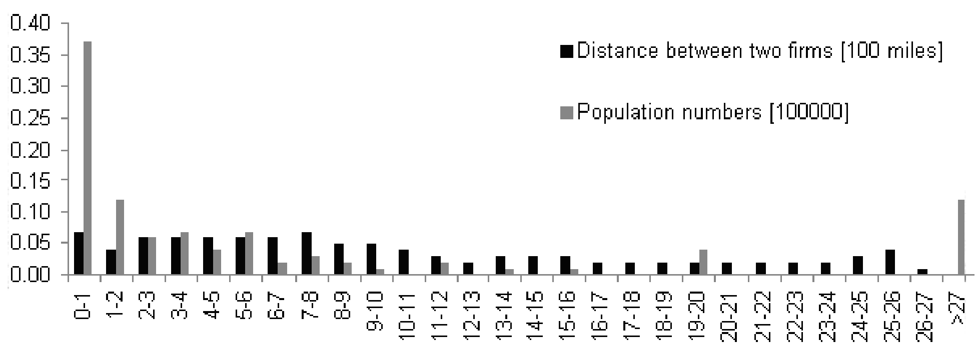


Figure 1: Sample characteristics: Proportion of population numbers measured in one hundred thousands for headquarter locations and proportion of distances between firm pairs measured in hundred miles.

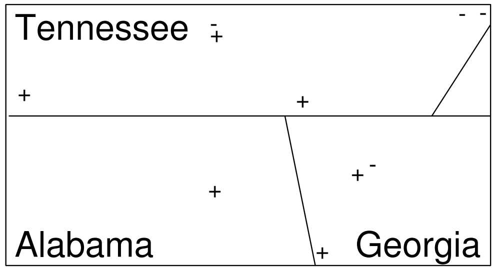


Figure 2: Map cutout of headquarter locations: Small (−) and big (+) cities.

A summary of the sample with respect to federal states is presented in Table 1. For each federal state the number of firms, the number of firms in large cities and the average distance between all firms is tabulated.

State	Firms	Large cities	d	State	Firms	Large cities	d
Alabama	5	4	230.8	Minnesota	14	9	41.9
Arizona	4	4	4.7	Missouri	8	6	111.2
Arkansas	5	2	127.0	Nebraska	2	2	12.5
California	68	43	196.8	Nevada	2	2	333.7
Colorado	7	4	11.4	New Hampshire	1	0	-
Connecticut	13	6	243.0	New Jersey	17	1	22.5
Delaware	3	0	0.0	New York	58	45	106.5
District of Columbia	2	2	3.9	North Carolina	13	10	58.8
Florida	11	8	135.3	Ohio	25	21	111.4
Georgia	13	12	34.4.	Oklahoma	3	3	65.4
Idaho	2	2	8.4	Oregon	2	0	15.7
Illinois	32	14	40.8	Pennsylvania	20	15	143.8
Indiana	6	3	82.3	Rhode Island	4	1	18.9
Iowa	2	1	32.7	Tennessee	10	7	192.1
Kentucky	5	4	55.9	Texas	41	37	135.7
Louisiana	2	1	216.5	Utah	1	1	-
Maryland	8	2	25.6	Virginia	10	4	59.6
Massachusetts	18	4	14.0	Washington	9	6	10.2
Michigan	11	4	76.7	Wisconsin	8	5	15.4

Table 1: State distribution: Number of headquarters of firms, number of headquarters of firms in large cities and average distance d [miles] between firms for each federal state.

For the analysis of pairwise comovements of stock returns, we consider (a) raw stock returns and (b) residual stock returns from a factor analysis. Residual returns are analyzed in order to control for the possibility that geographical clustering of firms with similar characteristics leads to higher correlations of nearby firms. The analysis of raw stock returns is in particular interesting from the perspective of undiversified investors with a tilt towards local stocks. Since those investors are undiversified with respect to distance as well as with respect to industry and other factors, their portfolio risk will be affected by correlation effects of distance as such but also by geographical clustering of similar firms. Neglecting the latter in the assessment of diversification losses would lead to an underestimation of such losses.

Residual returns are obtained from a linear regression with the excess return $R - RF$ as dependent variable, where R is the monthly stock return and RF denotes the risk–free return. The set of independent variables includes the following five factors[1]

- Fama and French (1993) show that common variation in stock returns can be explained to a large part by three factors: market excess return $RM - RF$, the return of small minus big stocks (SMB) and the return of high book–to–market minus low book–to–market stocks (HML) .

- Following Carhart (1997) we add a momentum factor (MOM).

- Barker and Loughran (2007) find that correlation between returns of firms operating in the same industry is 2.5 times larger than that of two average firms in different industries. In order to control for industrial clustering, we include the difference between mean industry return and market return $RI - RM$ as an additional factor. We consider the 48 industry classes defined in Fama and French (1997), which are based on the four digit SIC code. The corresponding monthly mean industry returns RI are not based on our sample of 465 firms but are taken from Kenneth French's data library, who computes the mean by industry classes over all stocks traded on the AMEX, NYSE and NASDAQ.

Residuals are obtained by first running a five–factor regression of the form

$$\begin{aligned} R_{j,t} - RF_t = a_j + b_j(RM_t - RF_t) + s_j SMB_t + h_j HML_t \\ + m_j MOM_t + c_j(RI_{j,t} - RM_t) + \varepsilon_{j,t} \end{aligned} \tag{1}$$

separately for each firm j in the sample. Here, $RI_{j,t}$ denotes the mean industry return at time t for industry class of firm j. The corresponding estimated residuals are then given by

$$\begin{aligned} \widehat{\varepsilon}_{j,t} = (R_{j,t} - RF_t) - \widehat{a}_j - \widehat{b}_j(RM_t - RF_t) - \widehat{s}_j SMB_t \\ - \widehat{h}_j HML_t - \widehat{m}_j MOM_t - \widehat{c}_j(RI_{j,t} - RM_t)\,, \end{aligned} \tag{2}$$

where $(\widehat{a}_j, \widehat{b}_j, \widehat{s}_j, \widehat{h}_j, \widehat{m}_j, \widehat{c}_j)$ denotes the least squares estimator of $(a_j, b_j, s_j, h_j, m_j, c_j)$.

Note that the mean correlation of raw returns over all firm pairs considered is 0.191 whereas the correlation for estimated residual stock returns is 0.008 on average.

[1] We use the data provided by Kenneth French; see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Barker and Loughran (2007) have chosen a different approach to account for factors that might affect correlations. They regress pairwise correlations of raw stock returns on distance as well as a set of control variables. To capture industry effects, for example, Barker and Loughran (2007) include the mean industry correlation as an explanatory variable; to capture differences in systematic risk, they include the differences of the Fama–French 3–factor betas. However, we favor the factor model approach as it better captures firm characteristics. Consider three firms from the same industry, two of them exhibiting average exposure to industry risk, one a lower exposure (e.g. because some operations belong to another industry). In the Barker–Loughran model one would implicitly assume that the correlation of these three firms is the same. In our model, differences will be captured through the industry coefficient in the factor regression. Similarly, consider two pairs of firms: the betas of the pair {firm 1, firm 2} are $\beta_1 = 0.5$ and $\beta_2 = 1$, the betas of {firm 3, firm 4} are $\beta_3 = 1$ and $\beta_4 = 1.5$. The beta difference used as an explanatory variable by Barker and Loughran (2007) is the same for each pair. Ceteris paribus, firms 3 and 4 should have a higher correlation than firms 1 and 2, though.

3 Methodology

We introduce two approaches for the analysis of spatial correlations, namely a linear regression model and the mark correlation function. The main difference between these two approaches relates to the order in which the data is analyzed. Our data consists of a number of measurement locations (the 465 locations of firms' headquarters) and a number of measurement times (60 months in our sample). This data is decomposed in two different ways: in the regression approach, we first average over time by computing the time series correlation for each pair (i, j) of the 465 headquarter locations individually, then we average over space by analyzing the drivers of these correlations. On the other hand, computing the mark correlation function, we get a functional correlation estimate for each month which is then averaged over time.

3.1 Regression Model

The effect of distances between headquarter locations on the correlation of stock returns is first analyzed by means of an ordinary least squares (OLS) regression. Pairwise correlation is regressed on a set of dummy variables that capture the distance between two headquarter locations. For this purpose we define distance classes and set the respective dummy to one if the distance between two firms belongs to a certain distance class, i.e. for some $m < m'$ we put

$$D_{i,j}^{m,m'} = \begin{cases} 1 \text{ if } m \text{ miles} \leq \text{distance between firm } i \text{ and firm } j < m' \text{ miles}, \\ 0 \text{ else.} \end{cases} \tag{3}$$

The empirical correlation $CORR_{i,j}$ of firms i and j is computed based on 60 monthly observations. When we use raw returns the correlation is estimated by

$$CORR_{i,j} = \frac{\sum_{t=1}^{60}(R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)}{\sqrt{\sum_{t=1}^{60}(R_{i,t} - \bar{R}_i)^2 \sum_{t=1}^{60}(R_{j,t} - \bar{R}_j)^2}}, \tag{4}$$

where $R_{i,t}$ denotes the return of firm i at time t and $\bar{R}_i$ is the mean return of firm i. When we analyze correlations based on residual stock returns we replace $R_{i,t}$ and $\bar{R}_i$ in equation (4) by the respective values for residual returns obtained from (2).

The general regression model for empirical correlations has the form

$$CORR_{i,j} = \alpha + \beta^\top D_{i,j} + u_{i,j}, \tag{5}$$

where u denotes the error term, D is a vector of dummy distance variables considered, $\beta^\top$ is the transposed vector of respective coefficients, and α denotes the regression constant.

By construction the observations in our sample are not independent. Each firm contributes to multiple observations because we consider pairwise correlations. Consequently, the OLS standard errors are biased downward, which leads to inflated t–statistics for α and β. Following Barker and Loughran (2007) we address this problem by estimating the standard errors through a bootstrap simulation. More precisely, we generate 1000 bootstrap samples by randomly drawing 465 firms with replacement from the set of all firms and constructing all possible pairwise combinations (pairwise combinations for the same firm are omitted). We then run an OLS regression for each bootstrap sample to get the 1000 bootstrap coefficient estimates for α and β. Finally, the standard errors are set to the standard deviations of the respective 1000 bootstrap coefficient estimates. Note that in our data analysis (see Section 4) the means of the bootstrap coefficient estimates are similar to the values obtained from an OLS regression for the original sample.

In a first approach to the problem in question, we consider various distance classes for the specification of the regression model given in (5). We begin with 15 distance classes each capturing a distance of 100 miles: $0-100, 100-200, \ldots, 1400-1500$ miles and a reference class for all distances larger than 1500 miles. This first analysis reveals (see Section 4 below) that correlations of stock returns are significantly higher if firm headquarters are located not further apart than 100 miles; beyond that, distance effects are relatively small or not significant. As a consequence we focus on distances less than 100 miles and analyze whether a more detailed statement can be made on the structure of spatial correlations within this range. For this purpose we let the reference class include all distances larger than 100 miles and consider five further distance classes, each for a distance of 20 miles (i.e. $0-20, \ldots, 80-100$ miles). We will refer to this specification as Model 1.

In our sample approximately 60% of all pairs of firm headquarters located within a distance of 100 miles are located in the same federal state. Here the question arises as to what extent correlations of returns are due to geographical proximity or to the simple fact that two firms are located in the same federal state. In order to capture the border effect, i.e. two firms are located in different federal states, we define a binary border variable B as

$$B = \begin{cases} 1 \text{ if two firms separated by a federal border,} \\ 0 \text{ else.} \end{cases}$$

We then expand Model 1 by including five additional covariates, each being the product of the border variable and a distance class. The resulting linear regression model is referred to as Model 2.

A third model (Model 3) analyzes the hypothesis that the correlation of returns can be attributed to the fact that the respective firms are located in the same city. In analogy to Model 2, we enlarge Model 1 by including the covariate $C(1-B)D^{0,20}$, where C is a dummy variable set to one if two firms are located in different cities and zero otherwise. The dummy $C(1-B)D^{0,20}$ captures the cases in which two firms with a distance of at most 20 miles are located in the same federal state, but in different cities. We only include the product of the city and distance class dummies for distances less than 20 miles since for firm pairs further apart only 0.005% are located in the same city.

3.2 Mark Correlation Function

Besides the regression approach explained in Section 3.1 we use a method from spatial statistics, which is based on the so–called mark correlation function for marked point processes (Illian et al., 2008). In order to analyze the spatial correlations of stock returns, we then consider the locations of firm headquarters as points X_i on the (spherical) surface of the earth and their raw stock returns and residual stock returns, respectively, as marks R_i of these points. The sequence $(X_1, R_1), (X_2, R_2), \ldots$ of all firm locations together with their stock returns is considered as a marked point process on the sphere $S_\rho \subset \mathbb{R}^3$, where S_ρ has its midpoint at the origin and circumference $\rho = 3959.871$ miles.

More precisely, since the headquarters of S&P 500 firms considered in the present paper are located within the USA, we can consider the sequence $(X_1, R_1), (X_2, R_2), \ldots$ as the restriction of a (more comprehensive) marked point process on S_ρ, which is restricted to the territory of the USA. The latter point process on the whole sphere S_ρ is assumed to be isotropic, which means that its distribution is invariant with respect to arbitrary rotations of the (spherical) coordinate system. Furthermore, we will use the notion of independent marking, where the synonymic notion of independent labelling is also used by some authors. The marked point process $(X_1, R_1), (X_2, R_2), \ldots$ is called independently marked if the marks $R_1, R_2, \ldots$ are independent and identically distributed random variables, which are independent of the sequence $X_1, X_2, \ldots$.

The mark correlation function $\kappa(r)$ of the marked point process $(X_1, R_1), (X_2, R_2), \ldots$ gives information on how the values of the marks of points that are located a given distance $r > 0$ apart are stochastically correlated. Heuristically speaking, positive values of $\kappa(r)$ indicate that pairs of points with distance r have similar marks, while negative values of $\kappa(r)$ indicate that pairs of points with distance r tend to have rather different marks. In case of independent marking, it can be shown that $\kappa(r) \equiv 0$ holds for any $r > 0$. The mark correlation function of $(X_1, R_1), (X_2, R_2), \ldots$ can therefore be interpreted as a quantitative characteristic of the spatial interaction between the marks R_i of the points X_i.

A more formal definition of the mark correlation function $\kappa(r)$ can be found, e.g. in Illian et al. (2008), where numerous applications of this point process characteristic are discussed. Further examples of statistical correlation analysis for spatial marked point patterns are investigated, e.g. in Eckel et al. (2008), where the temporal trend of the geographical correlations of the purchasing power in Baden–Württemberg, Germany, is analysed by means of the mark correlation function, and in Mattfeldt et al. (2008), which presents a spatial correlation analysis of labelling patterns for mammary carcinoma cell nuclei.

For any $r \in (0, r_{\max})$, where $r_{\max}$ is a suitably chosen maximum distance, a statistical estimator $\widehat{\kappa}(r)$ for $\kappa(r)$ can be given by

$$\widehat{\kappa}(r) = \frac{\displaystyle\sum_{X_i, X_j \in W, i \neq j} k_h(r - |X_i - X_j|)(R_i - \widehat{\mu})(R_j - \widehat{\mu})}{\displaystyle\sum_{X_i, X_j \in W, i \neq j} k_h(r - |X_i - X_j|)} \Big/ \widehat{\sigma^2}, \tag{6}$$

where $|X_i - X_j|$ is the spherical distance of X_i and X_j, while k_h is the Epanechnikov kernel with bandwidth $h = 20$ miles, and W denotes the sampling window (in our case, the territory of the USA). Furthermore,

$$\widehat{\mu} = \frac{1}{\#\{n : X_i \in W\}} \sum_{X_i \in W} R_i$$

and

$$\widehat{\sigma^2} = \frac{1}{\#\{i : X_i \in W\} - 1} \sum_{X_i \in W} (R_i - \widehat{\mu})^2$$

are estimators for the mean and variance of the marks, respectively. Note that an edge–correction of the estimator $\widehat{\kappa}(r)$ proposed in (6) is not straightforward for the spherical case. However, it can be omitted in this study, since we consider only short–range correlations.

We also remark that for the definition and estimation of the mark correlation function it is convenient to consider so–called simple point patterns only, i.e., there is at most one mark R_i at any location X_i, which means that $R_i = R_j$ if $X_i = X_j$. Thus, we aggregate the (raw or residual) returns of firms with the same zip code to one (raw or residual) return, where we use the mean (raw or residual) return of the affected firms as joint mark. This leaves us with the 356 locations, i.e. 356 points X_i in the point pattern. The estimator $\widehat{\kappa}(r)$ given in (6) for the mark correlation function $\kappa(r)$ has been implemented using the Java–based GeoStoch library, which has been developed during the last 10 years at Ulm University (Mayer et al., 2004).

3.3 Confidence Intervals

Besides computing estimates for the model characteristics β and $\kappa(r)$ introduced in Sections 3.1 and 3.2, respectively, we determine confidence intervals for these characteristics.

We first consider the regression model introduced in Section 3.1 and construct approximative 95% confidence intervals for the components $\beta_1, \ldots, \beta_\ell$ of the vector $\beta^\top = (\beta_1, \ldots, \beta_\ell)$ of regression coefficients, which are based on the OLS estimator $\widehat{\beta}^\top = (\widehat{\beta}_1, \cdots, \widehat{\beta}_\ell)$ for β and on the bootstrap estimates $\widehat{\beta}^{(i)\top} = (\widehat{\beta}_1^{(i)}, \ldots, \widehat{\beta}_\ell^{(i)})$, where ℓ denotes the number of distance classes considered in the respective regression model; $i = 1, \ldots, 1000$. For each $k = 1, \ldots, \ell$, we compute the mean value $\overline{\beta}_k$ and the respective standard error $\mathrm{SE}(\beta_k)$, where

$$\overline{\beta}_k = \big(\widehat{\beta}_k^{(1)} + \ldots + \widehat{\beta}_k^{(1000)}\big)/1000 \quad \text{and}$$
$$\mathrm{SE}(\beta_k) = \sqrt{\frac{1}{999}\sum_{n=1}^{1000}\big(\widehat{\beta}_k^{(n)} - \overline{\beta}_k\big)^2}\,. \tag{7}$$

Then, assuming that the quotient $(\widehat{\beta_k} - \beta_k)/\mathrm{SE}(\beta_k)$ is drawn from a distribution which is close to the standard normal distribution, an approximative 95% confidence interval for the k-th component β_k of β is given by $\big(\widehat{\beta_k} - z_{0.975}\mathrm{SE}(\beta_k),\, \widehat{\beta_k} + z_{0.975}\mathrm{SE}(\beta_k)\big)$, where $z_{0.975}$ is the 0.975 quantile of the standard normal distribution. Note that the use of formula (7) for the standard error $\mathrm{SE}(\beta_k)$ is justified by the application of bootstrapping.

In order to construct (pointwise) 95% confidence intervals for the values $\kappa(r)$ of the mark correlation function we can proceed in a similar way. For any $r \in (0, r_{\max})$ and for each of the 60 months we can compute the estimates $\widehat{\kappa}^{(1)}(r), \ldots, \widehat{\kappa}^{(60)}(r)$ using the formula for $\widehat{\kappa}(r)$ given in (6). Then, we compute the mean value $\overline{\kappa}(r)$ and the standard error $\mathrm{SE}(\kappa(r))$, where

$$\overline{\kappa}(r) = \big(\widehat{\kappa}^{(1)}(r) + \ldots + \widehat{\kappa}^{(60)}(r)\big)/60 \quad \text{and}$$
$$\mathrm{SE}(\kappa(r)) = \sqrt{\frac{1}{59}\sum_{n=1}^{60}\big(\widehat{\kappa}^{(n)}(r) - \overline{\kappa}(r)\big)^2}\,. \tag{8}$$

Assuming again that the quotient $\sqrt{60}(\overline{\kappa}(r) - \kappa(r))/\mathrm{SE}(\kappa(r))$ is drawn from a distribution which is close to the standard normal distribution, an approximative 95% confidence interval for $\kappa(r)$ is given by

$$\Big(\overline{\kappa}(r) - \frac{1}{\sqrt{60}} z_{0.975}\mathrm{SE}(\kappa(r)),\, \overline{\kappa}(r) + \frac{1}{\sqrt{60}} z_{0.975}\mathrm{SE}(\kappa(r))\Big).$$

The formula for the standard error $\mathrm{SE}(\kappa(r))$ is based on the assumption that the estimated values $\widehat{\kappa}^{(1)}(r), \ldots, \widehat{\kappa}^{(60)}(r)$ are independently sampled. This independence property can be tested by standard statistical procedures for time series, e.g. Fisher's g–test (Brockwell and Davis, 1991).

Note that it is also possible to compute confidence intervals for the values $\kappa(r)$ of the mark correlation function based on bootstrap estimates of $\kappa(r)$, where a method suggested in Mattfeldt et al. (2006) can be used. Here, to obtain bootstrap confidence intervals for $\kappa(r)$, we create 1000 bootstrap samples $S_1, \ldots, S_{1000}$ with 60 items from the original sample $\{\widehat{\kappa}^{(1)}(r), \ldots, \widehat{\kappa}^{(60)}(r)\}$ for each r separately. The sampling is independent and with replacement. For all 1000 bootstrap samples their corresponding mean values $\bar{\kappa}_i(r) = \frac{1}{60}\sum_{j=1}^{60} \widehat{\kappa}_i^{(j)}(r)$, $i = 1, \ldots, 1000$ are computed, where $\widehat{\kappa}_i^{(j)}(r)$ denotes the j-th item in S_i. A 95% confidence interval of $\kappa(r)$ is then given by $(\bar{\kappa}^*_{26}(r), \bar{\kappa}^*_{975}(r))$, where the sequence $\bar{\kappa}^*_1(r), \ldots, \bar{\kappa}^*_{1000}(r)$ is the values $\bar{\kappa}_1(r), \ldots, \bar{\kappa}_{1000}(r)$ sorted by size.

4 Empirical Results

In the following we present the results of our spatial correlation analysis of stock returns which we obtained by means of the regression approach introduced in Section 3.1 as well as the results of our analysis using the mark correlation function explained in Section 3.2.

First, the results for residual stock returns are presented in Section 4.1, whereas in Section 4.2 we compare some of these results with corresponding results for raw stock returns.

4.1 Analysis of Residual Stock Returns

4.1.1 A First Look at the Effects of Distance on Correlation

We first consider 15 distance classes, each for a distance of 100 miles with the reference class capturing all distances larger than 1500 miles, i.e.,

$$CORR_{i,j} = \alpha + \sum_{k=1}^{15} \beta_k D_{i,j}^{100(k-1),100k} + u_{i,j}. \tag{9}$$

Coefficient estimates are presented in the first column of Table 2. We see a significant increase in correlation with respect to the reference class only for distances less than 100 miles. We find this observation confirmed in the results presented in columns 2 to 4. Here the specification (9) has been modified with respect to the number of included dummy variables. Columns 2 and 3 show regression results when including dummy variables for distances of 100 miles up to a distance of 1000 and 500 miles, respectively. In column 4 the specification also takes into account dummies capturing distances between 500 and 1000 as well as between 1000 and 1500 miles.

The results presented in the last column of Table 2 show coefficient estimates for a model including $D^{0,100}$, $D^{100,500}$, $D^{500,1000}$ and $D^{1000,1500}$ as explanatory variables. We again observe a significant increase in correlation of residual stock returns for distances less than 100 miles. Additionally, we find a statistically significant increase for distances between 100 and 500 miles. Yet, the increase is rather moderate, only 0.008. The increase for the first distance class is twice as large.

4.1.2 Detailed Analysis of Short–range Correlations

In the following we report results for a detailed analysis of short–range correlations, i.e. Models 1 to 3 introduced in Section 3.1. Note that the effects captured by Models 2 and 3 cannot be investigated by the mark correlation function introduced in Section 3.2.

	1	2	3	4	5
$D^{0,100}$	0.016	0.015	0.014	0.016	0.016
	(2.745)	(2.595)	(2.482)	(2.757)	(2.745)
$D^{100,200}$	0.006	0.005	0.004	0.006	
	(1.084)	(0.891)	(0.772)	(1.063)	
$D^{200,300}$	0.009	0.007	0.007	0.009	
	(1.795)	(1.652)	(1.563)	(1.783)	
$D^{300,400}$	0.008	0.007	0.006	0.008	
	(1.776)	(1.675)	(1.602)	(1.845)	
$D^{400,500}$	0.007	0.006	0.005	0.007	
	(1.302)	(1.145)	(1.06)	(1.303)	
$D^{500,600}$	0.004	0.003			
	(0.944)	(0.74)			
$D^{600,700}$	0.004	0.003			
	(0.893)	(0.638)			
$D^{700,800}$	0.003	0.002			
	(0.751)	(0.51)			
$D^{800,900}$	0.002	0.001			
	(0.313)	(0.121)			
$D^{900,1000}$	0.001	-0.001			
	(0.114)	(-0.121)			
$D^{1000,1100}$	0.005				
	(0.986)				
$D^{1100,1200}$	0.001				
	(0.18)				
$D^{1200,1300}$	-0.003				
	(-0.41)				
$D^{1300,1400}$	0.001				
	(0.192)				
$D^{1400,1500}$	0.007				
	(1.158)				
$D^{100,500}$					0.008
					(2.134)
$D^{500,1000}$				0.003	0.003
				(0.901)	(0.933)
$D^{1000,1500}$				0.003	0.003
				(0.867)	(0.877)
Const	0.004	0.005	0.006	0.004	0.004
	(1.631)	(2.631)	(3.444)	(1.704)	(1.631)
Adjusted R^2	0.0012	0.0010	0.0008	0.0009	0.0008
Firm pairs	107186	107184	107184	107184	107186

Table 2: Regression results for residual stock returns: Dependent variable is the pairwise correlation of residual stock returns. Independent variables are distance dummy variables. t–statistics are in parentheses. Adjusted R^2 and firm pairs are averages from 1000 bootstrap simulations.

Model 1: The first model considers 5 distance classes, each capturing a distance of 20 miles, and a reference class for all distances exceeding 100 miles, i.e.,

$$CORR_{i,j} = \alpha + \sum_{k=1}^{5} \beta_k D_{i,j}^{20(k-1),20k} + u_{i,j}. \quad (10)$$

Regression results for residual stock returns are shown in the first column of Table 3. Detailed results for the mark correlation function are omitted in the following for lack of space. The respective estimated spatial correlations (with pointwise 95% confidence intervals) are shown in Figure 3, where Figure 3(a) presents results for the regression approach and Figure 3(b) for the mark correlation function. For the regression approach we graph estimated coefficients for distance dummies included in Model 1.[2]

Both approaches used for the analysis of spatial correlations indicate similar results: correlations of residual stock returns are clearly recognizable for distances less than 40 miles. Yet, the correlation declines with distance and for firms further apart than 40 miles, we find no significant correlation of returns. A closer comparison of the results of the two methods shows that the principle form of the curves is identical. Moreover, the magnitude of the spatial correlations is similar. However, using the mark correlation function the detected significance of spatial correlations appears more clearly. Confidence intervals in Figure 3(b), which have been computed for the mark correlation function, are shorter than those computed in Figure 3(a). To examine whether this is due to the aggregation of locations for the mark correlation function as mentioned at the end of Section 3.2, we reran the regression analysis on the aggregated data. However, we found that this does not narrow the confidence intervals shown in Figure 3(a). We also computed bootstrap confidence intervals for the mark correlation function using the method mentioned at the end of Section 3.3, where we obtained results which are practically identical to those shown in Figure 3(b). Note that Fisher's g–test has been applied in order to test whether the estimates $\widehat{\kappa}^{(1)}(r), \cdots, \widehat{\kappa}^{(60)}(r)$ of the mark correlation function considered in Figure 3(b) are independently sampled. The null hypothesis of independence was not rejected.

Model 2: In the second model considered, we add variables capturing the border effect for distance classes considered in Model 1, resulting in the following regression specification:

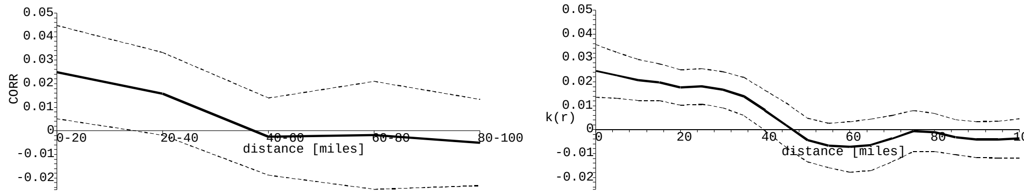
$$CORR_{i,j} = \alpha + \sum_{k=1}^{5} \Big(\beta_k D_{i,j}^{20(k-1),20k} + \gamma_k B D_{i,j}^{20(k-1),20k} \Big) + u_{i,j}, \quad (11)$$

with binary variable B set to one if two firms are located in different federal states. Analyzing the regression results reported in the second column of Table 3 we find that coefficient estimates for distance dummies $D^{0,20}$ and $D^{20,40}$ in Model 2 increase both in absolute value and in significance as compared to Model 1. In other words, we observe

[2] Graphics are provided for a better comparison. The illustration of results from the regression analysis derives from a linear interpolation of the estimated coefficients of five distance classes, whereas the mark correlation function provides "quasi-continuous" values.

	All observations		
	Model 1	Model 2	Model 3
$D^{0,20}$	0.025	0.030	0.038
	(2.471)	(2.743)	(2.285)
$D^{20,40}$	0.016	0.028	0.015
	(1.755)	(2.223)	(1.749)
$D^{40,60}$	-0.002	-0.005	-0.002
	(-0.292)	(-0.382)	(-0.279)
$D^{60,80}$	-0.002	0.013	-0.002
	(-0.163)	(0.541)	(-0.199)
$D^{80,100}$	-0.005	-0.008	-0.005
	(-0.534)	(-0.341)	(-0.555)
$BD^{0,20}$		-0.043	-0.051
		(-2.033)	(-2.184)
$BD^{20,40}$		-0.033	
		(-1.75)	
$BD^{40,60}$		0.004	
		(0.246)	
$BD^{60,80}$		-0.020	
		(-0.75)	
$BD^{80,100}$		0.003	
		(0.128)	
$C(1-B)D^{0,20}$			-0.018
			(-0.924)
Constant	0.007	0.007	0.007
	(4.3805)	(4.414)	(4.507)
Adjusted R^2	0.0010	0.0015	0.0014
Firm pairs	107187	107196	107186

Table 3: Regression results for residual stock returns: Dependent variable is the pairwise correlation of residual stock returns. t–statistics are in parentheses. Model 1 includes dummy variables for five distance classes. Model 2 additionally captures the interaction of distance classes and federal borders. Model 3 controls for city borders in the closest distance class. Adjusted R^2 and firm pairs are averages from 1000 bootstrap simulations.



(a) Regression analysis

(b) Mark correlation function

Figure 3: Estimated spatial correlations for residual stock returns of all S&P 500 firms using (left) regression analysis and (right) the mark correlation function. Dashed lines are pointwise 95% confidence intervals.

a stronger increase in correlation relative to the reference class if two nearby firms are located in the same federal state. We also observe that for each distance class coefficient estimates of the distance dummy variable D and the respective variable capturing the interaction between distance and federal border BD are opposite in sign and close in magnitude, e.g. for the first distance class the estimate for the distance dummy coefficient is $\widehat{\beta}_1 = 0.030$ with the coefficient estimate of $BD^{0,20}$ being $\widehat{\gamma}_1 = -0.043$. In other words, if two firms are located in the same federal state and the distance between their headquarters is at most 20 miles we find an increase in correlation of residual returns by 0.03. Yet, if two firms are located in distinct federal states we would expect no increase in residual correlation $(0.030 - 0.043 = -0.013)$ despite their geographical proximity. A test of the null hypothesis $H_0 : \beta_i + \gamma_i = 0$ for each distance class $i = 1, \ldots, 5$ confirms the observed relation. t–statistics deduced from 1000 bootstrap estimates range from -0.703 to -0.043. Hence, we cannot reject the null hypothesis at a reasonable significance level for any of the distance classes and conclude that correlation of stock returns due to geographical proximity is only nonzero if firms are located in the same federal state and the distance between headquarter locations is less than 40 miles.

Model 3: A similar conclusion is to be drawn for Model 3, where we distinguish between firm pairs located in the same or in distinct cities. The question of interest here is whether for two nearby firms the fact that these are located in different cities does intensify the border effect seen in Model 2. We analyze this question by including the dummy variable $C(1-B)D^{0,20}$ in the regression equation. This variable equals one only if two firms with a distance of less than 20 miles are located in the same federal state, but in different cities. Formally, Model 3 is specified by

$$CORR_{i,j} = \alpha + \sum_{k=1}^{5} \beta_k D_{i,j}^{20(k-1),20k} + \delta_1 C(1-B) D_{i,j}^{0,20} + u_{i,j}. \tag{12}$$

Regression results are presented in the third column of Table 3. For nearby firms located in the same city, we find an increase in residual correlation of 0.038, significant at the 5% significance level. On the other hand, if two nearby firms are separated by a federal border (and hence are located in different cities) we find the results from Model 2 confirmed. The respective coefficients add up to -0.013 and we again reject the null hypothesis $H_0 : \beta_1 + \gamma_1 = 0$ based on a t–statistic of -0.7713. For firms located in the same federal state but in different cities, we observe a less pronounced decline to $0.038 - 0.018 = 0.02$. The remaining increase in correlation is marginally significant, with the t–statistic for the null hypothesis $H_0 : \beta_1 + \delta_1 = 0$ being 1.7936. Overall, we see a noticeably stronger effect of a federal border as compared to a city border on the correlation of stock returns for nearby firms.

Summarizing, we can conclude that after controlling for other factors which might influence the comovements of stock returns (e.g. industry), residual stock returns show a larger correlation if the respective firms are located in the same city or in the same federal state and if they are not further apart than 40 miles.

4.1.3 Correlation Analysis for Different City Sizes

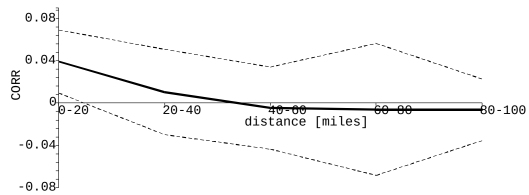
In addition to a city or state border we consider the size (i.e. population number) of headquarter cities as a further potential factor to influence the relation of distance and stock return comovements. Previous findings concerning the city size are controversial. Pirinsky and Wang (2006) report that correlation of stock returns is higher in larger cities. Barker and Loughran (2007) draw the opposite conclusion from the observation that the difference in the populations of headquarter cities does not explain correlations. Note that the approach chosen by Barker and Loughran (2007) does not distinguish between pairs of small cities and large cities. Consider two large and two small firms, both pairs equally sized. Consequently, the difference in populations for the large and for the small pair is zero and does not control for the difference between the pairs. We therefore choose to control for differences in population numbers by separating the sample with respect to population. Specifically, we classify cities as large if the population number exceeds 100000 and as small otherwise. This results in a total of 43365 firm pairs located in large cities and 14365 firm pairs located in small cities. Regression results as well as results from the analysis using the mark correlation function are presented in Table 4 and Figure 4 for each subsample. As for the regression results we present coefficient estimates for Model 1.

From the regression analysis on the entire sample discussed in Section 4.1.2, we have seen that a federal border has an impact on the stock return correlation of a closely located firm pair. For this reason we introduce border variables for the first two distance classes in addition to the five distance dummies in the regression specification for the subsamples.

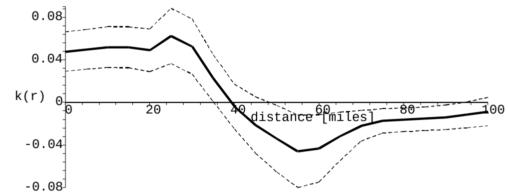
When controlling for size of headquarter cities, we find that the results reported in Table 3 and Figure 3 are mainly driven by large cities (see Figures 4(a) and (b)). If firms are located in small cities, we find no significant correlation of residual stock returns (see Figures 4(c) and (d)). Furthermore, using the mark correlation function, we unexpectedly detect slight, but significant negative correlations for firms located about $50 - 80$ miles apart (see Figure 4(b)).

	Large cities		Small cities	
	Model 1	Model 2	Model 1	Model 2
$D^{0,20}$	0.039	0.041	-0.003	0.004
	(2.583)	(2.587)	(-0.229)	(0.274)
$D^{20,40}$	0.010	0.022	0.009	0.008
	(0.51)	(0.846)	(0.558)	(0.438)
$D^{40,60}$	-0.005	-0.004	-0.002	-0.003
	(-0.239)	(-0.228)	(-0.139)	(-0.169)
$D^{60,80}$	-0.006	-0.006	-0.008	-0.007
	(-0.187)	(-0.213)	(-0.396)	(-0.376)
$D^{80,100}$	-0.006	-0.006	0.018	0.020
	(-0.428)	(-0.446)	(0.802)	(0.853)
$BD^{0,20}$		-0.022		-0.063
		(-0.336)		(-1.637)
$BD^{20,40}$		-0.031		-0.004
		(-0.537)		(-0.123)
Constant	0.007	0.007	0.008	0.008
	(3.326)	(3.268)	(2.422)	(2.518)
Adjusted R^2	0.0027	0.0029	0.0009	0.0016
Firm pairs	43311	43165	14144	14224

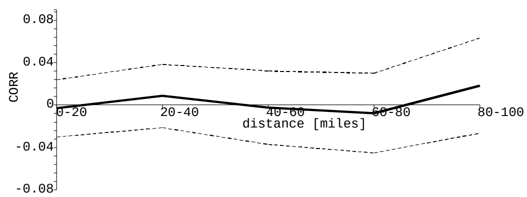
Table 4: Regression results for residual stock returns (subsamples): Dependent variable is the pairwise correlation of residual stock returns. Model 1 includes dummy variables for five distance classes. Model 2 additionally captures the interaction of distance classes and federal borders for the first two distance classes. t–statistics are in parentheses. Subsamples analyzed are firm pairs located in cities with more than or less than 100000 inhabitants. Adjusted R^2 and firm pairs are averages from 1000 bootstrap simulations.



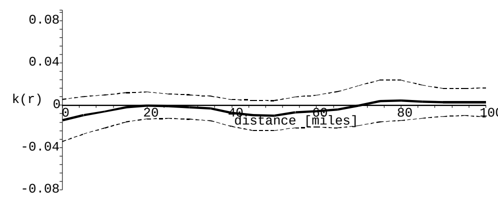
(a) Large cities - regression analysis



(b) Large cities - mark correlation function



(c) Small cities - regression analysis



(d) Small cities - mark correlation function

Figure 4: Estimated spatial correlations for residual stock returns for S&P 500 firms in large and small cities, respectively, using (left) regression analysis and (right) the mark correlation function. Dashed lines are pointwise 95% confidence intervals.

4.2 Analysis of Raw Stock Returns

For a comparison of the results obtained in Section 4.1 for residual stock returns with those for raw stock returns we also perform a short–range correlation analysis for raw stock returns of all S&P 500 firms with headquarters located in the USA. For the regression approach we again use Models 1, 2 and 3.

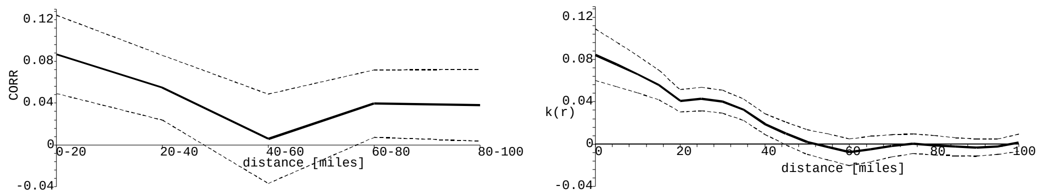
It can be seen from Table 5 and Figure 5 that we detect significant spatial correlations up to a distance of 40 miles for all three regression models as well as for the mark correlation function. A comparison of these results with those obtained in Section 4.1 from the analysis of residual stock returns shows that the distance up to which we detect significant spatial correlations is similar, but the magnitude has changed. The geographical comovement of residual stock returns is less strong than in the case of raw stock returns. A possible explanation is that some (but not all) of the geographical effects are already captured by the five–factor regression analysis discussed in Section 2.

5 Implication of Geographical Proximity on Portfolio Performance

In the previous section we have seen that after controlling for other factors, stock returns of firms located closely to each other (up to 40 miles) exhibit a higher correlation than firms further apart from each other. In order to analyze the implication of this result on portfolio performance we simulate portfolios that differ in the average distance between portfolio constituents.

	All observations		
	Model 1	Model 2	Model 3
$D^{0,20}$	0.087	0.097	0.093
	(4.5304)	(4.6483)	(4.403)
$D^{20,40}$	0.055	0.077	0.055
	(3.4603)	(3.9664)	(3.458)
$D^{40,60}$	0.006	-0.026	0.006
	(0.2696)	(-0.6963)	(0.269)
$D^{60,80}$	0.040	0.060	0.040
	(2.3974)	(1.7336)	(2.400)
$D^{80,100}$	0.038	-0.004	0.038
	(2.1712)	(-0.1105)	(2.173)
$BD^{0,20}$		-0.067	-0.085
		(-2.1558)	(-2.589)
$BD^{20,40}$		-0.056	
		(-2.0493)	
$BD^{40,60}$		0.059	
		(1.3719)	
$BD^{60,80}$		-0.028	
		(-0.6906)	
$BD^{80,100}$		0.048	
		(1.2602)	
C(1-B)$\mathrm{D}^{0,20}$			0.009
			(0.283)
Constant	0.186	0.186	0.186
	(26.9944)	(26.9944)	(27.002)
Adjusted R^2	0.0079	0.0094	0.0089
Firm pairs	107184	107184	107186

Table 5: Regression results for raw stock returns: Dependent variable is the pairwise correlation of raw stock returns. t–statistics are denoted in parentheses. Model 1 includes dummy variables for five distance classes. Model 2 additionally captures the interaction of distance classes and federal borders. Model 3 controls for city borders in the closest distance class. Adjusted R^2 and firm pairs are averages from 1000 bootstrap simulations.



(a) Regression analysis (b) Mark correlation function

Figure 5: Estimated spatial correlations for raw stock returns of all S&P 500 firms using (left) regression analysis of Model 1 and (right) the mark correlation function. Dashed lines are pointwise 95% confidence intervals.

We construct three types of equally weighted portfolios based on the following rules:

1. **Close portfolio:** Given a randomly chosen firm sort all remaining firms with respect to distance and select the $n-1$ closest firms for a portfolio size of n.

2. **Far portfolio:** Given a randomly chosen firm, select $n-1$ firms at random that are further away than a prespecified distance d. If less than $n-1$ firms fulfill this condition, allow for firms at a distance closer than d miles, but with headquarters in distinct federal states.

3. **Random portfolio:** Select n firms at random.

We set the portfolio size to $n = 100$ and $n = 30$, respectively, and simulate 1000 portfolios of each type. The minimum distance for far portfolios is set to $d = 50$ miles, ensuring that pairwise correlation between firms is low. For each of the simulated portfolios, we compute the portfolio return μ_P and volatility σ_P as

$$\mu_P = \frac{1}{n}\sum_{i=1}^{n}\mu_i \qquad \text{and} \qquad \sigma_P = \Big(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{i,j}\Big)^{1/2},$$

where μ_i denotes the mean monthly return[3] of asset i and $\sigma_{i,j}$ is the covariance of returns of assets i and j.

Furthermore, we consider the portfolio performance as measured by the coefficient a from the factor analysis (see equation (1)) as well as the residual variance σ^{ε}. The respective values for a portfolio are given by

$$\alpha_P = \frac{1}{n}\sum_{i=1}^{n}a_i \qquad \text{and} \qquad \sigma_P^{\varepsilon} = \Big(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{i,j}^{\varepsilon}\Big)^{1/2},$$

where a_i is the constant term of the factor analysis for firm i and $\sigma_{i,j}^{\varepsilon}$ is the covariance of residual stock returns of assets i and j.

[3] Both, the mean and volatility of returns are estimated using 60 monthly observations for the years 2000 to 2004.

Table 6 displays the mean characteristics for 1000 simulated portfolios. Panels A and B show values for raw stock returns and residual stock returns, respectively. Results are presented for all three portfolio types each of size $n = 100$ and $n = 30$, respectively.

Both raw and residual stock returns show the same patterns with respect to volatility. Volatility is smallest for portfolios composed of distant firms and largest for portfolios including only firms that are located closely together. For portfolios combining $n = 30$ stocks, return volatility of the far portfolio is more than one percentage point lower than that of the close portfolio; per annum (p.a.), the difference is a sizeable $(6.20\% - 4.95\%)\sqrt{12} = 4.33\%$. The p.a. difference in residual volatility between the far and the close portfolio is 0.8% and 1.21% for a portfolio size of $n = 100$ and $n = 30$, respectively.

A test of the null hypothesis for equality of portfolio volatilities is rejected for each pairwise comparison. The lowest t–statistics (in absolute value) observed are 4.983 and 10.643 (comparing the volatility of random and close portfolios of size $n = 100$ and $n = 30$, respectively).

The simulation results also suggest that close portfolios have a lower mean return than both far and random portfolios, further increasing the advantage of far portfolios. The patterns in the portfolio alphas differ. Here far portfolios have a lower value of α compared to the other two portfolios. Overall, we would not overinterpret the result because they are likely to be incidental. By chance or for some reason not captured in our factor analysis, returns and alphas may differ across regions. For example, the mean return of firms located in the north–east is 1.05%, lower than the overall average of 1.21%. This affects the mean returns of the far and the close portfolio because their average geographical distribution departs from the sample distribution. The far portfolios are geographically more dispersed than the firms in the sample, which leads to an underrepresentation of firms belonging to the low–return cluster in the north–east. The converse is true for the close portfolios.

When translating the risk differences into return differentials, we therefore opt to be conservative and assume that the mean return is the same for all portfolios. Consider a portfolio of size $n = 30$. Given a risk–free rate[4] of $RF = 0.22\%$ and a mean return of $\mu_P = 1.21\%$ then the lower volatility of the far portfolio can be translated into a monthly return advantage of 0.0684% $(= 0.0121 - (RF + \frac{4.95}{5.31}(0.0121 - RF)))$ and 0.2034% relative to the random and close portfolio, respectively. On an annual basis this amounts to 0.8% and 2.4%, which appears to be economically significant. For an equally weighted portfolio with a fixed number of stocks, investors can therefore benefit from selecting stocks based on distance. Put differently, investors who focus on local stocks because they hope to exploit informational advantages should be aware of the increased risk. Ivkovic and Weisbrenner (2005, Table V) find that investors' return advantage from selecting local S&P 500 stocks is 0.8% to 2.1%, depending on the regression specification. Based on our simulation results, the median investor, who owns just two stocks in the sample of Ivkovic and Weisbrenner (2005), should expect that such gains are neutralized by an increase in risk relative to more a geographically dispersed portfolio.

[4] We set the risk–free rate to the average one–month treasury bill rate, which over the period 2000–2004 was $RF = 0.22\%$.

Panel A: Portfolio characteristics — Raw returns

	Portfolio size $n = 100$			Portfolio size $n = 30$		
	Far	Random	Close	Far	Random	Close
Return μ_p	1.25%	1.21%	1.17%	1.27%	1.21%	1.14%
	(0.08%)	(0.11%)	(0.12%)	(0.17%)	(0.19%)	(0.21%)
Volatility σ_p	4.56%	5.06%	5.22%	4.95%	5.31%	6.20%
	(0.17%)	(0.32%)	(0.91%)	(0.46%)	(0.60%)	(2.55%)

Panel B: Portfolio characteristics — Residual returns

	Portfolio size $n = 100$			Portfolio size $n = 30$		
	Far	Random	Close	Far	Random	Close
α_P	0.45%	0.55%	0.49%	0.49%	0.55%	0.53%
	(0.09%)	(0.13%)	(0.29%)	(0.22%)	(0.24%)	(0.40%)
Residual volatility σ_p^ε	1.02%	1.16%	1.27%	1.71%	1.77%	2.06%
	(0.09%)	(0.15%)	(0.29%)	(0.22%)	(0.25%)	(0.83%)

Table 6: Portfolio simulation: Mean characteristics for 1000 simulated portfolios consisting only of firms far apart from each other, only of firms close by to each other or of firms with a random location. Panel A shows the mean return and the mean volatility of raw stock returns. Panel B tabulates mean values for α and mean residual volatility. Respective standard deviations are in parenthesis.

6 Conclusion

We have examined returns of stocks contained in the S&P 500 and found that the correlation is larger for nearby firms. This result, however, holds only for firms located in the same federal state and for firms that are less than 40 miles apart. Once the distance is larger than 50 miles, it matters little or nothing how large it is. The findings should help to detect the drivers of distance effects. Based on our results, it seems likely that it is local rather than regional effects which are at work, and that policy–based differences between states partly explain the observed patterns.

The findings differ from the ones reported by Pirinsky and Wang (2006) and Barker and Loughran (2007) showing that the choice of a research methodology is crucial for examining the effects of distance on cross–correlation. Barker and Loughran (2007) criticize the methodology by Pirinsky and Wang (2006) and obtain different results for large firms. We further modify the regression approach by Barker and Loughran (2007) and get results which are broadly consistent but differ in some aspects: The irrelevance of distance beyond 40 miles, the role of federal borders and city sizes. We also suggest an alternative approach from spatial statistics. The results are similar to the ones from our modified regression approach, strengthening confidence in the robustness of results.

An advantage of the mark correlation function is that it leads to shorter confidence intervals in contrast to those obtained from the regression analysis. Moreover, the mark correlation function is a quasi-continuous function, i.e. for each distance r of a dense grid of distances we obtain an estimate of the spatial correlation $\kappa(r)$. Thus, considering the mark correlation function, there is no need to aggregate the data into distance classes. On the other hand, an advantage of the regression approach is the higher flexibility of the model, e.g. it is straightforward to account for firms being located in different federal states.

Using simulations, we show that it matters for portfolio risk whether stock selection is based on distance. Portfolio risk can be significantly reduced if one avoids nearby firms. This result is also relevant for judging the value of local information. While return advantages appear to exist (Ivkovic and Weisbrenner, 2005), it seems doubtful whether the typical, undiversified investor benefits from those advantages in terms of an improved risk–return–ratio.

References

[1] Barker D, Loughran T. The geography of S&P 500 stock returns. *Journal of Behavioral Finance* 2007; **8**; 177–190.

[2] Brockwell PJ, Davis RA. *Time series: theory and methods* (2nd ed). Springer: New York; 1991.

[3] Carhart MM. On persistence in mutual fund performance. *Journal of Finance* 1997; **52**; 57–82.

[4] Coval JD, Moskowitz TJ. The geography of investment: informed trading and asset prices. *Journal of Political Economy* 2001; **109**; 811–841.

[5] Eckel S, Fleischer F, Grabarnik P, Schmidt V. An investigation of the spatial correlations for relative purchasing power in Baden–Württemberg. *Advances in Statistical Analysis* 2008; **92**; 135–152.

[6] Fama EF, French KR. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 1993; **33**; 3–56.

[7] Fama EF, French KR. Industry cost of equity. *Journal of Financial Economics* 1997; **43**; 153–193.

[8] Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical analysis and modelling of spatial point patterns.* J. Wiley & Sons: Chichester; 2008.

[9] Ivkovic Z, Weisbenner S. Local does as local is: information content of the geography of individual investors common stock investments. *Journal of Finance* 2005; **60**; 267–306.

[10] Mattfeldt T, Eckel S, Fleischer F, Schmidt V. Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. *Journal of Microscopy* 2006; **223**: 107–119.

[11] Mattfeldt T, Eckel S, Fleischer F, Schmidt V. Statistical analysis of labelling patterns of mammary carcinoma cell nuclei on histological sections. *Working Paper, University of Ulm* 2008.

[12] Mayer J, Schmidt V, Schweiggert F. A unified simulation framework for spatial stochastic models. *Simulation Modelling Practice and Theory* 2004; **12**; 307–326.

[13] Pirinsky CA, Wang Q. Does corporate headquarters location matter for stock returns? *Journal of Finance* 2006; **61**; 1991–2015.

[14] Zwillinger D (ed). *Spherical geometry and trigonometry.* CRC Press: Boca Raton; 1995.

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Implied Market Price of Weather Risk" by Wolfgang Härdle and Brenda López Cabrera, January 2009.
002 "On the Systemic Nature of Weather Risk" by Guenther Filler, Martin Odening, Ostap Okhrin and Wei Xu, January 2009.
003 "Localized Realized Volatility Modelling" by Ying Chen, Wolfgang Karl Härdle and Uta Pigorsch, January 2009.
004 "New recipes for estimating default intensities" by Alexander Baranovski, Carsten von Lieres and André Wilch, January 2009.
005 "Panel Cointegration Testing in the Presence of a Time Trend" by Bernd Droge and Deniz Dilan Karaman Örsal, January 2009.
006 "Regulatory Risk under Optimal Incentive Regulation" by Roland Strausz, January 2009.
007 "Combination of multivariate volatility forecasts" by Alessandra Amendola and Giuseppe Storti, January 2009.
008 "Mortality modeling: Lee-Carter and the macroeconomy" by Katja Hanewald, January 2009.
009 "Stochastic Population Forecast for Germany and its Consequence for the German Pension System" by Wolfgang Härdle and Alena Mysickova, February 2009.
010 "A Microeconomic Explanation of the EPK Paradox" by Wolfgang Härdle, Volker Krätschmer and Rouslan Moro, February 2009.
011 "Defending Against Speculative Attacks" by Tijmen Daniëls, Henk Jager and Franc Klaassen, February 2009.
012 "On the Existence of the Moments of the Asymptotic Trace Statistic" by Deniz Dilan Karaman Örsal and Bernd Droge, February 2009.
013 "CDO Pricing with Copulae" by Barbara Choros, Wolfgang Härdle and Ostap Okhrin, March 2009.
014 "Properties of Hierarchical Archimedean Copulas" by Ostap Okhrin, Yarema Okhrin and Wolfgang Schmid, March 2009.
015 "Stochastic Mortality, Macroeconomic Risks, and Life Insurer Solvency" by Katja Hanewald, Thomas Post and Helmut Gründl, March 2009.
016 "Men, Women, and the Ballot Woman Suffrage in the United States" by Sebastian Braun and Michael Kvasnicka, March 2009.
017 "The Importance of Two-Sided Heterogeneity for the Cyclicality of Labour Market Dynamics" by Ronald Bachmann and Peggy David, March 2009.
018 "Transparency through Financial Claims with Fingerprints – A Free Market Mechanism for Preventing Mortgage Securitization Induced Financial Crises" by Helmut Gründl and Thomas Post, March 2009.
019 "A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics" by Ji Cao, Wolfgang Härdle and Julius Mungo, March 2009.
020 "Putting Up a Good Fight: The Galí-Monacelli Model versus 'The Six Major Puzzles in International Macroeconomics'", by Stefan Ried, April 2009.
021 "Spectral estimation of the fractional order of a Lévy process" by Denis Belomestny, April 2009.
022 "Individual Welfare Gains from Deferred Life-Annuities under Stochastic Lee-Carter Mortality" by Thomas Post, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Pricing Bermudan options using regression: optimal rates of convergence for lower estimates" by Denis Belomestny, April 2009.

024 "Incorporating the Dynamics of Leverage into Default Prediction" by Gunter Löffler and Alina Maurer, April 2009.

025 "Measuring the effects of geographical distance on stock market correlation" by Stefanie Eckel, Gunter Löffler, Alina Maurer and Volker Schmidt, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".