Comments Welcome

Firm Characteristics, Covariances, and Portfolio Optimization*

Pengqin Gao†

First Draft: April, 2007
This Draft: November, 2009

Abstract

The covariance pattern of S&P 500 stocks, without being associated with any common factors, is explicitly linked to their characteristics such as size, momentum, and accounting-based fundamentals. I propose a new covariance estimator based on this observation. In comparison to factor-based covariance models, a characteristic-based covariance model brings substantial diversification benefits and utility gains for a risk-averse investor seeking a global minimum variance portfolio and an optimal tangency portfolio, respectively.

JEL Classification: C21, G11

Key Words: Characteristics; Covariance; Random Field; Spatial Model; Portfolio Optimization

*I am grateful to Alan Bester, Tim Conley, George Constantinides, Paul Gao, Chris Hansen, Ravi Jagannathan, Jeff Russell, Ruey Tsay, Keqi Yan, and seminar participants at the University of Chicago, the UBS O'Connor Global Asset Management (Quantitative Strategies) for helpful discussions and comments. I thank Ken French for sharing data. I am responsible for all remaining errors.

†The University of Chicago Booth School of Business. E-mail: pgao@chicagobooth.edu; Tel: (312) 504-8030; Web: http://home.uchicago.edu/~pgao

1 Introduction

Modeling stock return comovement is an essential task in finance. A factor-based model is perhaps the most popular method in quantitative portfolio management.[1] In this paper, I explore how, if at all, the covariance pattern of S&P 500 stocks, without being associated with any common factors, is explicitly linked to firm characteristics such as size, momentum, and accounting-based fundamentals. From the portfolio optimization perspective, I propose a new covariance estimator that exploits the direct and close relationship between firm characteristics and covariance structure.

The data generating process assumes that firms are located in an Euclidean space and their characteristics are coordinates. The locations (firm characteristics) and outcomes (stock returns) are both random in the space. This data generating process is called a random field. The fundamental assumption is that firm characteristics contain important information about dependence structure. In particular, the covariance is a decreasing function of economic distance (a measure of firm similarity in characteristics) and stocks tend to comove stronger (weaker) when their characteristics are more (less) similar. This relationship between return comovement and firm characteristics is what I call a *spatial covariance pattern* throughout this paper.

I find the abnormal returns of S&P 500 stocks display a strong spatial covariance pattern. Specifically, returns will covary more strongly next month when firms are closer this month in terms of economic distance, and returns will eventually become uncorrelated when firms are farther away. To be robust to model misspecification and return outliers, and more importantly, to let the data speak, I nonparametrically estimate spatial covariances based on a local average method. The findings of strong spatial covariance patterns are robust to different sample periods and to different risk-adjustment methods of estimating abnormal returns.

Is the spatial covariance pattern caused by common shocks to the market and to some industries? To show that it is not, I construct a bootstrapped acceptance region of common shocks by simulating an aggregate shock plus industry-specific shocks. Under a variety of industry classifications, I reject the hypothesis that the characteristic-based return covariations among S&P 500 stocks are

[1] The literature is too enormous to list. Earlier work includes the single-index model (Sharpe, 1963) and the constant correlation model (Elton and Gruber, 1973). Ledoit and Wolf (2003) propose a shrinkage estimator. Engle et al. (1990) impose a factor-ARCH covariance structure to model T-bills. Chan et al. (1998, 1999) examine the performance of factor models to forecast covariances. Jagannathan and Ma study the effects of portfolio constraints on factor models. Moskowitz (2003) investigates the link between characteristic-based factors and multivariate GARCH covariances.

attributed to market-wide shocks or industry common shocks. My findings suggest that a spatial covariance pattern cannot be produced by using additional factors, and in this sense a characteristic-based spatial covariance model is not compatible with a factor-based covariance model. No matter how many additional industry factors are included, a factor-based model can only partially describe the covariance structure that is directly associated with characteristic similarity.

Which characteristics are most useful to describe the characteristic space for S&P 500 stocks and therefore predict their future return comovement? To answer this question, I examine a battery of price-related and accounting-based characteristics. In particular, I investigate the relative importance of fundamental-based accounting characteristics. Interestingly, the ex-ante explanatory power of price-scaled variables (book-to-market, dividend yield, earnings yield, etc.) is largely dominated by pure accounting variables. The prediction power of accounting characteristics is mainly from capital structure, accounting liquidity, and operating efficiency, whereas accounting profitability has no contribution. In addition, the characteristics of market equity (size) and past one-year return (momentum) also have prediction power, and their effects are not minimized by the existence of accounting variables.

Modeling return covariance as explicitly related to firm characteristics is economically important. First, I calculate the contribution of covariances to the aggregate volatility of abnormal returns in my sample. The magnitude is substantial. On average, I find that the characteristic-based covariance terms collectively explain 69.5 (65.3) percent of the aggregate variation when the ratio of covariances to variances is measured on an equal-weighted (value-weighted) basis. Therefore, among S&P 500 stocks in this study, characteristic similarity seems to be associated with a systematic comovement that cannot be easily diversified.

Second, I find that an investor who seeks a global minimum variance portfolio will have substantial diversification benefits if she exploits a characteristic-based spatial covariance structure. Under no short sale constraint, for example, an investor on average will choose 397 stocks to hold under the spatial covariance model. In sharp contrast, she will only choose 50-60 stocks to hold under the factor-based covariance model.[2] Additionally, an investor will also experience much less extreme weights and a 2-3 times smaller portfolio turnover under the former case than the latter.

[2] I provide the details of data and sample selection later. On average, my sample contains approximately 401 non-financial stocks in the S&P 500 index each month.

Finally, I find that an investor who seeks an optimal tangency portfolio will have substantial utility gains if she exploits a characteristic-based spatial covariance structure. Under a Regulation T constraint, for instance, a risk-averse investor with quadratic utility would be willing to pay 2.8 to 17.9 percent of her wealth per year to switch from the Fama-French three-factor covariance model to a K-Bessel spatial covariance model. Without imposing any portfolio constraints, she would even be willing to pay 63.4 to 89.5 percent.

The random field approach presented in this paper has two advantages to modeling covariance. First, it places fewer restrictions on comovement structure. In a finite-factor setting (K-factor model), the covariance matrix of stock returns can have the rank of at most K. But in a spatial setting, the covariance matrix of a similar set of stock returns can have full rank. Second, the random field is parsimonious. Only a small number of parameters (typically 2 to 4) need be fitted even for a large number of stocks. As a result, less estimation error is induced and the estimated covariances are more stable. For S&P 500 stocks in this study, the condition number for the spatial-based covariance matrix is 40 on average, whereas it ranges from 865 to 1179 for the factor-based covariance matrix.[3]

To the best of my knowledge, this is the first paper to suggest a random field model to study return comovement across individual stocks. A similar concept is used in bond literature on modeling term structure (Kennedy, 1994; Goldstein, 2000).[4] However, the random field in term structure is rather different from that in stock returns. Bond maturity (time) is the natural index in the former case, whereas there is no such natural index in the latter. This paper also contributes to the literature by suggesting a viable approach to estimating covariances for a large cross section of stocks.

The rest of the paper proceeds as follows. Section 2 discusses the issues of modeling covariance structure and the related literature. Section 3 presents the random field model, estimation and hypothesis test. Section 4 describes the data, selection of characteristics, and summary statistics. Section 5 provides the results. Section 6 performs robustness checks. Finally, section 7 concludes. The appendix provides details in constructing characteristic variables and economic distance.

[3] The theory of matrix algebra has different definitions of matrix condition number. I use the ratio between the largest and the smallest eigenvalue. The larger the ratio is, the more ill-behaved the matrix is. I provide more details later.

[4] Bester (2004) compares the properties between random field models and factor models (affine models) in the term structure context. See the detailed references therein.

2 Covariance Structure and Related Literature

The factor-based model is perhaps the most popular approach to modeling covariance structure in current literature. Connor and Korajczyk (1995, 2007) present excellent surveys on multifactor models of asset returns. The common factors, either statistically extracted or empirically observed, considerably reduce the dimensionality of describing return comovement. In this section, I discuss an improvement to modeling covariance structure under a random field model.

Briner and Connor (2007) refer to a misspecification issue on factor structure. The factor model has to focus on the strongest sources of covariation because the number of factors need be much smaller than the number of stocks. As a result, any comovement is assumed to be completely driven by shocks to common factors, and magnitudes and directions depend on factor loadings. Other sources of comovement will be ignored. For example, the tendency of a stronger covariation among regional banks due to local proximity and exposure (e.g., residents leave because of increased crimes) will not be captured by a shock to a banking industry factor. Needless to say, adding a regional banking factor provides a solution in such a case. However, this specification issue exists in general as the number of factors is restricted to be small. Daniel and Titman (1997) question whether the covariation among value firms results from a common factor (Fama-French HML factor), rather than similar characteristics and properties.[5]

Previous papers also pay attention to the estimation error in factor loadings. For a factor-based covariance model, small estimation error in factor loadings can accumulate into large estimation error in the variance-covariance matrix as the number of assets increases. Chan et al. (1999) find that the low correlation between predicted and ex-post realized covariances is mainly attributed to enormous estimation error among individual stocks (250 randomly selected stocks in their study). For commonly used economic factors, estimation mostly relies on time-series regressions. In practice, we have a large number of stocks in the cross section, but a small number of observations in the time series. DeMiguel et al. (2007) show that existing portfolio strategies cannot consistently outperform the naive equal-weighted portfolio strategy because of estimation error. Their simulation shows that the estimation window needs be much longer, even for a small number of assets,

[5] In particular, Daniel and Titman (1997, p.3) state: "[... the covariances of high book-to-market stocks ...] *reflect the fact that* [they] *tend to have similar properties; e.g., they might be in related lines of businesses, in the same industries, or from the same regions.*"

in order to reduce estimation error in sample-based mean-variance models (e.g., 6000 months for a portfolio of 50 assets).[6]

A spatial covariance model in a random field is designed to overcome these issues. It accommodates the comovement associated with shocks to common factors, and meanwhile keeps the model parsimonious. I qualitatively foreshadow two properties of the random field approach, and in the next section I describe the model in detail.

First, the key element of a random field is in using locations to characterize comovement. For the previous bank example, two local banks are geographically close. But more importantly, they are close in terms of having clients from the same region. This similarity will directly play a role in modeling their comovement. Any shock to one bank, either locally (from the business region) or globally (from the banking industry), will have an impact on the other through economic distance. The more (less) similar they are, the stronger (weaker) covariation they have. Therefore, an important task is to find a sensible set of firm characteristics to measure similarity.

Second, a characteristic-based spatial covariance model only requires to fit a small number of parameters (typically 2 to 4), even for high dimensional asset returns. The estimation of spatial covariances implicitly applies a shrinkage approach. Another advantage is that the parametric specification of a covariance function automatically ensures a positive definite variance-covariance matrix. Moreover, the estimation makes full use of both cross-sectional and serial replicates of covariance patterns. A large number of observations in the cross section indeed turns the curse of using individual stocks into a blessing. The use of individual stock-level data also avoids the loss of information when stocks are sorted into portfolios and cross-sectional correlations are reported as averages in prior studies.[7]

The literature agrees on the principle that it is desirable to impose a small number of parameters when modeling covariance structure. This principle significantly avoids the overfitting problem and reduces estimation error. In a seminar work, Engle (2002) proposes a dynamic conditional

[6] Partially because of estimation error for individual assets, the literature often considers well-diversified portfolios as representative assets to study covariance patterns. This approach is subject to the data-snooping bias, inherent from portfolio-sorted assets (Lo and MacKinlay, 1990). Ideally, the use of individual stocks fully characterizes true cross-sectional covariation. However, Moskowitz (2003) finds that idiosyncratic error of individual stock returns makes it very difficult, if not impossible, to identify covariation patterns.

[7] An example is for newly listed stocks. They have a small number of serial observations (less than a year), which makes the estimates of factor loadings imprecise. These stocks combine to provide a large number of cross-sectional observations. By means of firm similarity and replicates of covariance patterns across all pairs of these stocks, we are more likely to obtain precise estimates of their covariances under a spatial structure.

correlation model (DCC) to describe the pairwise correlations of standardized returns by the same parametrization across stocks. The original DCC model is not readily applied to a large number of assets because of the extremely cumbersome computation of the maximum likelihood estimator. Recently, Engle and Kelly (2008) propose a dynamic equicorrelation model (DECO) to dramatically simplify the computational difficulties of the DCC model by assuming all pairs of returns have the same correlation at a given time. A characteristic-based spatial covariance model is different from a DCC-type model in terms of using past information to estimate conditional covariances because it directly exploits time-varying firm characteristics. In contrast, the DCC model utilizes statistical information about moments of past (standardized) returns.

In terms of using the information on cross-sectional characteristics of assets, this paper is also related to the literature studying optimal portfolio weights without explicitly modeling return moments. Brandt et al. (2007) argue that stocks with similar characteristics should have similar portfolio weights. Their novel approach circumvents the specification of the underlying covariance structure by maximizing investor's expected utility. The authors also directly model portfolio weights as a function of size, book-to-market, and momentum. Under the assumptions of a random field model, stocks with similar characteristics tend to move strongly and hence the inverse of the covariance matrix will implicitly generate similar weights.

3 Random Field Model

3.1 Data Generating Process

Firm i is located in a k-dimensional Euclidean space, $i \in \mathbb{R}^k$, where characteristics are abstracted as coordinates (indexes). Both locations and outcomes are random variables in this space (a random field). An investor observes realizations of firm locations (characteristics) C_i, and outcomes (returns) R_i.

At the end of each month t, I assume that the monthly return of stock i (in excess of the risk-free return), R^e_{it}, has two components:

$$R^e_{it} = \mu_{it} + \epsilon_{it},\ i = 1, ..., N \tag{1}$$

where μ_{it} is the time-varying benchmark return and ϵ_{it} is the risk-adjusted abnormal return. ϵ_{it} is not predictable with respect to the information set I_{t-1}, including firm characteristics $C_{i,t-1}$ and common factors F_{t-1}, i.e., $E_{t-1}\epsilon_{it} = 0$.

The data generating process (DGP) of a random field assumes that firm indexes provide credible information about their dependence. In particular, similar firms (close in distance) comove strongly, dissimilar firms (far in distance) comove weakly, and firms become eventually uncorrelated as distance increases farther away. Statistically, the random field model assumes that covariance is a function of firm characteristics (in particular, a function of economic distance):

$$Cov_{t-1}(\epsilon_{it}, \epsilon_{jt}) = f(||C_{i,t-1} - C_{j,t-1}||) \text{ for } i \neq j, \tag{2}$$

where $\|\cdot\|$ is the Euclidean norm that measures the distance between firms i and j.[8] Equation (2), which I call a characteristic-based spatial covariance function, will be used throughout this paper. For the purpose of describing abnormal return comovement, the spatial covariance at distance zero, $f(0)$, is not particularly interesting. And later I will discuss the issue of conditional variance in the context of portfolio optimization, where I need specify the variance-covariance matrix of stock excess returns.

To obtain the benchmark return μ_{it} in equation (1), I employ the characteristic-matched portfolio adjustment method in Daniel et al. (1997). In their study of mutual fund performance, they compare the returns of equities held by a fund to the returns of portfolios of stocks with similar characteristics matched on size, book-to-market equity, and momentum. Statistically, this is equivalent to a simple nonparametric regression:

$$\begin{aligned} \widehat{\mu}_{i,t} &= \frac{1}{N}\sum_{j=1}^{N} w_j(C_{i,t-1}, C_{j,t-1}) \cdot R^e_{jt} \qquad (3) \\ w_j(\cdot) &= \begin{cases} 1 & \text{if } i, j \in \text{char. port.} \\ 0 & \text{otherwise} \end{cases}, \end{aligned}$$

where the characteristic-sorted portfolios are formed at time $t-1$, and the equal-weighted portfolio

[8] In geostatistics (Cressie, 1993), if a spatial covariance function depends on location, it is called to be nonstationary. If a spatial covariance function depends on direction, it is called to be anisotropic. In the context of stock returns, I assume a stationary and isotropic covariance function that only depends on distance and magnitude.

returns are used here for expositional simplicity. The weighting function in equation (3) captures the principle that similar firms should have similar expected returns, and that the cross-sectional average return across a large set of stocks within a characteristic-matched portfolio provides a benchmark return for individual firms.

To better capture time-varying information, I form 125 passive portfolios each month based on conditional triple sorts on market equity (size), book-to-market equity (BM), and past one-year cumulative returns (momentum). Specifically, at the beginning of month t, I first sort all common stocks into size quintiles. Then, within each size quintile, I sort stocks into additional quintiles based on their BM at the end of month $t-6$. Finally, within each size-BM quintile, I sort stocks into additional quintiles based on their momentum (cumulative returns from $t-12$ to $t-2$). All quintile breakpoints are determined using only NYSE stocks, although all stocks traded on NYSE/AMEX/NASDAQ are included to form portfolios. In addition, one firm's BM is industry-adjusted by subtracting the long-term industry average BM using the Fama-French 49-industry classifications (Cohen and Polk, 1995). The value-weighted monthly returns of these 125 portfolios are used as benchmark returns.

As a robustness check, I also use a conditional factor model to estimate benchmark returns. In particular, I specify the conditional beta as a function of firm characteristics (Avramov and Chordia, 2006):

$$R_{it}^{e} = E_{t-1}R_{it}^{e} + \beta_{i,t-1}' f_t + \epsilon_{it}, \tag{4}$$

where f_t is k-dimensional unanticipated factor shocks, $E_{t-1}R_{it}^{e} = \beta_{i,t-1}'\lambda$ is based on an exact factor-pricing equation of factor loadings ($\beta_{i,t-1}$) and risk premium (λ), and $\beta_{i,t-1}$ is a function of firm i's characteristics $C_{i,t-1}$. Specifically, I estimate the following time-series regression for each stock:

$$\begin{aligned} R_{it}^{e} \quad = \quad & \theta_{i0}MKTRF_t + (\theta_{i1} + \theta_{i2}ME_{i,t-1})\,SMB_t + (\theta_{i3} + \theta_{i4}BM_{i,t-1})\,HML_t + \\ & (\theta_{i5} + \theta_{i6}MOM_{i,t-1})\,UMD_t + \epsilon_{it}, \end{aligned} \tag{5}$$

where ME, BM, and MOM are size, book-to-market equity, and momentum, respectively, and $MKTRF$, SMB, HML, and UMD, are the Fama-French three factors and Carhart's momentum factor, respectively. Due to the estimation error in factor loadings, the estimates of abnormal returns are more noisy using the benchmarks from prespecified factor portfolios, than those from characteristic-matched portfolios.[9]

3.2 Nonparametric Estimation

To estimate the covariance function in equation (2), I follow a nonparametric approach to use data to identify a comovement pattern (Conley and Dupor, 2003). Since one objective of this paper is to investigate whether the covariation of abnormal returns is consistent with the prediction of a random field model, it is important for an estimator to be robust to both model misspecification and return outliers.

I denote the pairwise distance between firms i and j at time $t-1$ as $d_{ij,t-1} = ||C_{i,t-1} - C_{j,t-1}||$. The local average estimate of spatial covariance at a given input of economic distance h is:

$$\widehat{f}(h) = \sum_{t=1}^{T} \sum_{i \neq j}^{N} W_T(h - d_{ij,t-1}) \cdot \widehat{\epsilon}_{it}\widehat{\epsilon}_{jt} \text{ for } i \neq j \text{ and } h > 0, \tag{6}$$

where $\widehat{\epsilon}$ is the estimate of abnormal return from equation (1) and $W_T(\cdot)$ is the weighting function that sums to one and concentrates its mass at zero as $T \to \infty$. Essentially, equation (6) behaves like a sample moment estimate. At each input distance h, the pairs of cross products of abnormal returns at time t, $\widehat{\epsilon}_{it}\widehat{\epsilon}_{jt}$, have more weight put on them as their distances at time $t-1$ get closer to h, and are effectively excluded as their distances get farther away from h. If abnormal returns display the spatial covariance pattern, i.e., the comovement weakens when distance increases (firms become dissimilar), the estimator in equation (6) will precisely capture it. An easy way to visualize a spatial covariance pattern is to plot estimated covariances against characteristic-based economic distances.

For the empirical results presented later, I choose a smoothed Gaussian kernel for $W_T(\cdot)$. To select a bandwidth, I use the principle of nearest neighbors as in Fan and Gijbels (1996). That is, given an input h, I set its bandwidth equal to the fifth percentile of all differences between this input and all pairwise distances $d_{ij,t-1}$. Connor and Linton (2006) apply the same method in their

[9] In an unreported study, I also consider the risk adjustment method by running five-year rolling-sample regressions as in Brennan et al. (1998). The results (available upon request) are hardly changed in terms of the spatial covariance patterns of the abnormal returns of S&P 500 stocks.

semiparametric characteristic-based factor model. This procedure guarantees that the bandwidth is narrower (wider) where the data is locally denser (sparser).[10]

To investigate the relative importance of firm characteristics in describing comovement, I further estimate spatial covariance as a function of two types of distance (h_1, h_2) simultaneously:

$$\widehat{f}(h_1, h_2) = \sum_{t=1}^{T} \sum_{i \neq j}^{N} W_T^*(h_1 - d_{ij,t-1}^1, h_2 - d_{ij,t-1}^2) \cdot \widehat{\epsilon}_{it}\widehat{\epsilon}_{jt} \text{ for } i \neq j \text{ and } h > 0, \tag{7}$$

where d_{ij}^1 is the first type of distance (e.g., based on accounting-based fundamentals), d_{ij}^2 is the second type of distance (e.g., based on fundamental-to-price ratio attributes), and the weighting function $W_T^*(\cdot)$ is a bivariate Gaussian kernel that sums to one and concentrates its mass at zero as $T \to \infty$. In particular, I select two bandwidths separately for inputs h_1 and h_2 (use the principle of nearest neighbors), get the univariate Gaussian density for each distance, and multiply together to get the bivariate Gaussian kernel.

It is worth noting the several statistical features of the estimator in equation (6). First, it uses the replicates of return cross products at similar distances across firms. This principle of using cross-sectional replicates of covariance patterns, in fact, makes a full use of individual stock-level data. To see this, note that each individual $\widehat{\epsilon}_{it}$ has a fair amount of idiosyncratic error. But as we make full use of more and more replicates in the cross section, the noises are likely to be averaged out and the remaining component of covariation (if any) shows up.

Second, the estimator in equation (6) provides indirect evidence on whether comovement patterns and magnitudes are similar among the firms within different characteristic categories (e.g., small firms vs. big firms, growth firms vs. value firms, etc.). For example, the prior literature documents that both small and big firms tend to move together but they do not move in tandem. But the literature provides little evidence (if any) on whether the pattern and the extent of covariation are similar between small and big firms.[11]

[10] I also perform a robustness check by setting the bandwidth equal to the first percentile. The spatial covariance patterns are very similar to those shown in the paper.

[11] In geostatistic theory, this observation is related to the issue of covariance global stationarity vs. covariance local stationarity. For example, if the return covariation among small firms is different from that among big firms, this covariation doesn't exhibit covariance global stationary. As a result, estimating spatial covariance by using distances across all stocks will certainly fail to reveal any spatial covariation pattern. In such case, we need to impose covariance local stationarity and estimate separately. On the other side, if covariance global stationarity holds, we want to use more replicates of distances across all stocks, since this approach can lead to a more precise estimate.

Third, the estimator in equation (6) is asymptotically consistent (Hall and Patil, 1994). Unfortunately, due to sampling error, it cannot always generate a positive definite covariance matrix. This is not an issue when our goal is to describe return comovement pattern.

3.3 Hypothesis Test

One implication from the characteristic-based spatial covariance model is that abnormal returns tend to be uncorrelated as economic distance increases. To test this hypothesis, I employ a bootstrap method, since the usual parametric test statistic for a nonparametric estimator has inferior finite sample properties.

Specifically, conditional on each firm's location (characteristics) in the space, $C_{i,t-1}$, I simulate a bootstrap sample $\{\widehat{\epsilon}_{it}^{boot}\}$ from a distribution with the same stationary and marginal distribution of the abnormal returns. To do this, I draw random samples with replacement from each firm's empirical distribution of $\widehat{\epsilon}_{it}$. These independently and identically distributed (IID) samples are generated by preserving cross-sectional heterogeneity across firms and time-series homogeneity within each firm. After obtaining a series of bootstrap samples, I use the same local average estimator specified in equation (6) to calculate spatial covariances:

$$\widehat{f}^{boot}(h) = \sum_{t=1}^{T} \sum_{i \neq j}^{N} w_T(h - d_{ij,t-1}) \cdot \widehat{\epsilon}_{it}^{boot} \widehat{\epsilon}_{jt}^{boot} \text{ for } h > 0, \tag{8}$$

where the input distance h, the pairwise distance $d_{ij,t-1}$, and the weight $W_T(\cdot)$ are the same values used in estimating spatial covariances from the real data. By design, the bootstrapped spatial covariance $\widehat{f}^{boot}$ is equal to zero at all distances. To construct a 95 percent acceptance region of spatial independence at each distance, I repeat this bootstrap experiment many times, and in the end, drop the lower and upper 2.5 percentile of all bootstrapped covariances. In summary, if $\widehat{f}(h)$ estimated in equation (6) is contained in the acceptance region at a given distance h, we cannot reject the null hypothesis that its spatial covariance is not statistically different from zero.

Another important question is whether the characteristic-based spatial comovement pattern can be characterized by a few pervasive factors. For example, it is reasonable to believe that firms within a same industry have similar exposures to industry common shocks. Recall the previous discussion about the sources of covariation captured by a factor model, which is not necessarily compatible with a random field model. The bootstrap procedure involves the following two steps.

First, I decompose the abnormal return into three components in order to simulate both common shocks and idiosyncratic shocks:

- To generate a proxy for the distribution of the market-wide common shock, I use the empirical distribution of the cross-sectional average of $\widehat{\epsilon}_{it}$ at each point in time:

$$\overline{\epsilon}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \widehat{\epsilon}_{it}.$$

- To generate the proxies for the distributions of industry-specific common shocks, I calculate deviations of abnormal returns from the aggregate component, $v_{it} = \widehat{\epsilon}_{it} - \overline{\epsilon}_t$, and use the empirical distributions of the cross-sectional average of firm deviations within industries at each point in time (e.g., IND is an industry, M_{IND} is the number of firms within this industry):

$$\overline{v}_{IND,t} = \frac{1}{M_{IND,t}} \sum_{i \in IND}^{M_{IND,t}} v_{it}.$$

- To generate the proxies for the distributions of firm-specific idiosyncratic shocks, I use the empirical distributions of the remaining part of firm deviations from their industry components, $v_{i,t} - \overline{v}_{IND,t}$, for $i \in IND$.

Second, I draw samples in the order of decomposition above, and add them together to get the final bootstrap samples:

- Market-wide shocks are independently drawn from $\{\overline{\epsilon}_t\}$ over time.

- Industry-specific common shocks are independently drawn from $\{\overline{v}_{IND,t}\}$ over time. Block sample at each point of time t is used to allow arbitrary correlations across industries.

- Firm-specific idiosyncratic shocks are independently drawn from $\{v_{i,t} - \overline{v}_{IND,t}\}$ over time.

After obtaining a series of bootstrap samples in one experiment, I use the same local average estimator specified in equation (6) to calculate spatial covariances. After repeating the experiment many times, I follow the same procedure in testing spatial independence to construct a pointwise

95 percent acceptance region of common shocks. I examine a variety of industry shocks, and for the estimated spatial covariances outside the acceptance region, we can reject the hypothesis that the comovement is caused by an aggregate shock, plus industry-specific common shocks.

3.4 Portfolio Optimization

According to the mean-variance portfolio theory (Markowitz, 1952), a myopic investor holds a optimal portfolio by solving the following problem at time $t-1$:

$$\begin{aligned}\min_{w} \sigma_p^2 &= w' \left[Cov_{t-1}\left(R_t^e, R_t^{e\prime}\right)\right] w \\ s.t. \quad w' E_{t-1} R_t^e &= r_p \\ w' I &= 1,\end{aligned}$$

where σ_p^2 is the optimal portfolio variance, w is an $N \times 1$ vector of optimal portfolio weights, R_t^e is an $N \times 1$ vector of stock excess returns at time t, r_p is the target return (in excess of a risk-free rate), $E_{t-1}R_t^e$ is the conditional expected returns, and $Cov_{t-1}\left(R_t^e, R_t^{e\prime}\right)$ is the conditional variance-covariance matrix of excess returns.

To explore the economic value of characteristic-based spatial covariance structure in portfolio management, I assume the following conditional mean representation of monthly excess return R_{it}^e:

$$R_{it}^e = E_{t-1}R_{it}^e + \eta_{it}, \tag{9}$$

where $E_{t-1}R_{it}^e$ is the conditional expected return and η_{it} is the unexpected return that is not predictable, $E_{t-1}\eta_{it} = 0$. To model conditional variance of stock i, I further assume that η_{it} can be decomposed into two components:

$$\begin{aligned}\eta_{it} &= z_{it} + u_{it}, \qquad (10) \\ E_{t-1}z_{it} &= E_{t-1}u_{it} = 0.\end{aligned}$$

The first component, z_{it}, is idiosyncratic. It is uncorrelated in both the time series and the cross section: $Cov_{t-1}(z_{it}, z_{jt}) = 0$ for $i \neq j$, with a finite time-varying variance σ_{it}^2. The second component,

u_{it}, is orthogonal to the first one: $E_{t-1}(z_{it}u_{it}) = 0$. It is serially uncorrelated but cross-sectionally correlated: $Cov_{t-1}(u_{it}, u_{jt}) \neq 0$ for $i \neq j$, with a finite homoscedastic variance $Var_{t-1}u_{it} = \sigma^2$.

The random field assumption about u_{it} is similar to that made in equation (2): $Cov_{t-1}(u_{it}, u_{jt}) = g(||C_{i,t-1} - C_{j,t-1}||)$ for $i \neq j$. As a result, the conditional variance-covariance matrix of stock excess returns:

$$\begin{aligned} Cov_{t-1}\left(R^e_{it}, R^e_{jt}\right) &= g(||C_{i,t-1} - C_{j,t-1}||) \text{ for } i \neq j \\ Var_{t-1}\left(R^e_{it}\right) &= \sigma^2_{it} + \sigma^2 = \sigma^2_{it} + g(0) \end{aligned} \tag{11}$$

To estimate the conditional mean, I use characteristic-matched portfolio returns up to $t-1$ and calculate their sample mean returns over the past 60 months.

To estimate the conditional covariance, I assume a parametric form for the spatial covariance function in equation (11).[12] Specifically, I employ a general form of the K-Bessel (Matérn) model:

$$g(h) = \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{a}\right)^{\nu} K_{\nu}\left(\frac{h}{a}\right) \text{ for } a > 0,\ \nu \geqq 0, \text{ and } h > 0, \tag{12}$$

where h is the distance, σ^2 is the variance, ν is the parameter of smoothness, a is the parameter of scale, $\Gamma(\cdot)$ is the Gamma function, and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind of order ν. This model is flexible enough to specify a variety of smoothness and decay rates of a spatial covariance function.

The particular type of K-Bessel model when $\nu = \frac{3}{2}$ yields a simpler form:

$$g(h;\theta) = \sigma^2(1 + \frac{h}{a})\exp(-\frac{h}{a}) \text{ for } a > 0 \text{ and } h > 0, \tag{13}$$

where $\theta \equiv (a, \sigma^2)$ stands for two parameters. Using a small number of parameters to predict the covariances across a large number of stocks avoids the overfitting problem and consequently reduces the estimation error substantially. I estimate θ by a nonlinear least square (NLS) method.

[12] Cressie (1993) presents various parametric functions used in geostatistic applications (e.g., Gaussian, Exponential, and Gamma).

Particularly, I minimize the following residual sum squares (RSS):

$$\min_{\theta} \; RSS = \sum_{t=1}^{T} \sum_{i \neq j}^{N} \left\{ \widehat{\eta}_{it} \widehat{\eta}_{jt} - g(d_{ij,t-1}; \theta) \right\}^2, \tag{14}$$

where $\widehat{\eta}_{it}$ is estimated from equation (9) and $g(\cdot)$ is the function in equation (13).

To estimate the conditional variance, I calculate the rolling-sample variance of stock i, using its monthly excess returns in the prior 60 months. In practice, there is no advantage to estimate σ_{it}^2 and σ^2 separately.

The parametric estimator in equation (14) has several advantages. First, fitting a nonlinear regression is computationally efficient given a large number of pairwise distances. The estimate is also asymptotically consistent. Second, the least square approach implicitly applies weights on $\widehat{\eta}_{it}\widehat{\eta}_{jt}$ in our case, because we tend to have less observations as distances increase. Third, the overfitting problem is unlikely to happen here, since we only estimate two parameters for all covariance terms. Finally, the estimator in equation (14) produces a covariance matrix that is guaranteed to be positive definite.

4 Data, Characteristics Selection, and Summary Statistics

My sample consists of non-financial stocks in the S&P 500 index during December 1975 to October 2006. I identify these stocks from the CRSP monthly S&P 500 constituent files and extract all accounting information from the Compustat quarterly and annual industrial files. I exclude the financial firms in the S&P 500 index because of the different meaning of their high leverages (Fama and French, 1992) and accounting standards. On average, I have about 401 non-financial S&P 500 stocks every month. In this paper, I only include the firms when they are in the index. As a result, my panel data is unbalanced.

I choose three broad categories of characteristics to measure firm similarity. The first one is price-related attributes, including size (ME), momentum (RET_12), and past average return (AVGRET_12). The second one is fundamental-to-price ratio attributes, including book-to-market equity (BM), sales-to-price (S/P), cash flow to firm value (CF/FV), dividend yield (D/P), and earnings yield (E/P). The last one is accounting-based fundamental attributes from the perspectives of capital structure, profitability, and operating efficiency. Specifically, the accounting variables include total assets (TA), book equity (BE), leverage ratio (LEV), retained earnings to total assets (RE/TA), return on assets (ROA), return on equity (ROE), assets growth rate (AGR), sales growth rate (SGR), current ratio (LIQ), current gross margin ratio (MARG), and accounting turnover ratio (TO). The appendix provides the details about constructing these aforementioned characteristic variables.

To ensure the availability of accounting information, I require that there be a three-month lag between when these variables are used and the last fiscal quarter ending date. In addition, I combine various past levels of firm characteristics in order to maximally measure similarity in a multidimensional space. For example, in terms of momentum, I further decompose past one-year cumulative returns into past two-month, three-month, and six-month. In terms of fundamentals, I compute the average level of each accounting variable over the past four quarters. To deal with the issue of different measurement scales, I standardize each variable by subtracting its cross-sectional mean and dividing by the standard deviation, before calculating all pairwise distances. The appendix provides the details about calculating economic distances.

My selection of price-related and fundamental-to-price characteristics is motivated by prior studies on market anomalies. Although the literature has an ongoing debate about the source of market anomalies, both the risk-based and behavioral-based interpretation agree that firms within these characteristic groups comove strongly. As a result, it is natural to investigate the explicit relationship between return comovement and price-related and fundamental-to-price characteristics.

My selection of accounting-based fundamental characteristics is motivated by previous literature on historical financial performance and future stock returns (Ou and Penman, 1989; Piotroski, 2001). Firm-level accounting variables based on financial statements are direct measures of past financial performance. Fama and French (1995) study the economic content of common factors in terms of sales and earnings, and find that firms tend to behave similarly within the same size or book-to-market group in terms of profitability. Although the accounting information updates relatively slowly, the accounting-based fundamentals largely reflect firms' real activity and are less subject to market mispricing. As a result, the comovement pattern (if any) explicitly related to these accounting characteristics in the past is more likely to persist in the future.

Table 1 reports the summary statistics of selected characteristics over the subsamples (Panels

A-C) and the full sample (Panel D). Not surprisingly, S&P 500 firms are big caps in the market. In terms of book-to-market, they are slightly tilted towards growth stocks. However, these stocks display strong variability in characteristics. For example, in terms of past one-year return, the average momentum over all periods is 9.7 percent, but the standard deviation reaches 31.2 percent; in terms of assets and liabilities, the current ratio has standard deviation (3.04) that is much higher than its mean (1.86). In addition, both the level and the variability of characteristics change over time.

Figure 1 shows a classical multidimensional scaling (CMDS) representation of economic distances, calculating on accounting-based fundamentals (12 characteristics as coordinates) at the end of four selected months.[13] By means of computing eigenvectors for the distance matrix, CMDS essentially finds a configuration of low-dimensional points whose interpoint distances approximate those in the original space of high dimension. It is a convenient tool with which to visualize a high-dimensional object into a two-dimensional plane or a three-dimensional surface.[14] These plots clearly show that firm relationship (similarity in characteristics) varies substantially over time. For example, the locations of S&P 500 firms are more spread out more in 1981 than in 2005. As I shall present later, time-varying characteristics capture time-varying covariance structure and changing investment opportunity.

5 Main Results

5.1 Spatial Covariance Pattern

For S&P 500 stocks from December 1975 to November 2006, I first estimate their monthly abnormal returns by the characteristic-matched portfolio adjustment method (see equation(3)). Then, I estimate their covariances as a function of distances constructed from size, past return, capital structure, accounting liquidity, and operating efficiency (see equation(6)). I discuss the choice of these variables shortly (see Figure 2 for the complete list).

Figure 2 plots the spatial covariances, along with the bootstrap acceptance region of spatial

[13] These 12 characteristics are based on capital structure, accounting liquidity, and operating efficiency (TA, BE, LEV, LIQ, MARG, TO, TA_12, BE_12, LEV_12, MARG_12, TO_12). I discuss the choice of these variables in Section 5.1.

[14] For a complete description of CMDS and detailed procedures, see Mardia et al. (1979).

independence (top panel), and the histogram of characteristic-based economic distances (bottom panel).[15] The abnormal returns of the S&P 500 stocks display a strong spatial covariance pattern and they covary strongly in month t when characteristics are similar in month $t-1$. The decreasing comovement pattern is consistent with the prediction from a random field model. In particular, as distance increases from 1 to 5, the covariance decreases almost monotonically, and these abnormal returns become uncorrelated as distance increases over 5.8. In the end, covariances are not statistically different from zero among 25 percent of distances in the sample.

Figure 3 plots the bootstrap acceptance region of common shocks. Specifically, I simulate an aggregate shock plus 10 Fama-French industry shocks. The block sample preserves the correlations across industries. Only for distances between 1.9 and 3.5, I observe point estimates inside the acceptance region. However, the range of distances at which the statistically significant non-zero covariances lie outside the region constitutes approximately 45 percent of distances in the sample. This broad range of distances strongly rejects the hypothesis that the spatial comovement of risk-adjusted returns can be characterized by a few common factors.

To look closely at the relative importance of characteristics, I estimate covariances as a function of two types of distance simultaneously. Figure 4 plots the three-dimensional surfaces of spatial covariance on combinations of distances constructed from fundamental-to-price ratios (“BM”), accounting-based fundamentals (“ACCT”), capital structure (“C”), profitability (“P”), accounting liquidity and operating efficiency (“E”), and price-related attributes (“SM”). Figure 4 provides the details of the variables in calculating each type of distance.

Panel A shows a strong spatial comovement along “ACCT” distance, conditional on a “BM” distance. Given an “ACCT” distance, however, I find no discernible spatial covariance pattern in “BM” distance (the surface is relatively flat). This finding indicates that the characteristic of fundamental-to-price ratio plays no incremental role in describing covariation of abnormal returns, once I take into account the characteristic of accounting-based fundamental.[16]

Panel B shows that the surface in “P” distance tends to be flat, given either “C” or “E”

[15] To achieve a balance between revealing covariance pattern and computing time, I select 100 equally-spaced points and 500 bootstrap draws. It takes about an hour for a single run of the simulation over the full sample.

[16] In an unreported study, I find a weak book-to-market effect in the S&P 500 stocks. If a characteristic has no power to predict future return comovement, it is less likely to explain cross-sectional average returns. My finding is consistent with the traditional risk-based interpretation. For a large set common stocks in the market, Gao (2009) studies in detail the power of characteristic-based covariances in explaining cross-sectional expected returns.

distance. Therefore, the predicting power on future comovement from past similarity in profitability is completely dominated by that in capital structure, accounting liquidity, and operating efficiency. In contrast, both "C" and "E" distance are important in describing spatial comovement.

Panel C shows a strong spatial covariance pattern on "CE" distance, given a "SM" distance, and vice versa. Neither one of these two types of distance dominates the other one. Price-related attributes and accounting-based fundamentals are equally important in measuring firm similarity.

In summary, Figure 4 shows that the spatial covariance pattern of S&P 500 stocks is closely associated with price-related attributes (size and past return) and accounting-based fundamentals (capital structure, accounting liquidity, and operating efficiency). Throughout this paper, I estimate spatial covariance as a function of "SMCE" distance.

5.2 Economic Value of Spatial Covariance

So far, without assuming the existence of any common factors, I provide evidence supporting an explicit link between characteristic similarity and comovement of abnormal returns. An important question is how important it is for an investor, in terms of economic value, to exploit this characteristic-based spatial covariance structure. In this subsection, I calculate the contribution of covariances to the aggregate volatility of abnormal returns, and examine the out-of-sample performances of global minimum variance portfolio and optimal tangency portfolio (see Section 3.4 for portfolio optimization).

5.2.1 Covariance Contribution

To report the contribution of spatial covariance terms in collectively explaining the aggregate variation of risk-adjusted abnormal returns, I calculate the following equal-weighted and value-weighted measures at the end of each month t:

$$x_t^{ew} = \frac{1'\Omega_t\left(C_{ij,t-1};\theta\right)1}{1'V_t 1}, \tag{15}$$

and

$$x_t^{vw} = \frac{w_{t-1}'\Omega_t\left(C_{ij,t-1};\theta\right)w_{t-1}}{w_{t-1}'V_t w_{t-1}}, \tag{16}$$

where $\Omega_t(C_{ij,t-1};\theta)$ is the conditional covariance matrix ($i \neq j$), V_t is the diagonal matrix of conditional variance, 1 is a column vector of ones, and w_{t-1} is a column vector of weights at $t-1$ (based on market cap).

To estimate the conditional covariances, I use the NLS method for the K-Bessel function specified in equation (13). To estimate the conditional variances, I calculate rolling-sample variances using monthly returns in the prior 60 months. The measures in equations (15) and (16) can be interpreted as the contributions of covariance risk to portfolio volatility. A higher ratio implies a larger role for firm similarity in generating aggregate volatility in abnormal returns. Over the full sample, the time-series mean (median) of x_t^{ew} is 69.5 percent (66.9%) and the mean (median) of x_t^{vw} is 65.3 percent (60.7%). Therefore, the spatial comovement pattern is not only statistically present, but also economically meaningful.[17]

5.2.2 Global Minimum Variance Portfolio

The global minimum variance (GMV) portfolio avoids the issue of estimating mean returns. Therefore, I can directly check the out-of-sample performances of different covariance estimators. In this paper, the characteristic-based factor models are the one-factor model (Sharpe, 1963), the three-factor model (Fama and French, 1993), and the four-factor model (Carhart, 1997); the characteristic-based random field model is the K-Bessel spatial model in equation (13).

The covariance estimates are based on a rolling-sample approach: at the end of month t, I use data in the prior 60 months to construct the GMV portfolio. In addition, I examine both the unconstraint case and the no short sale constraint case under the following two scenarios: (i) the portfolio is monthly formed from December 1981 to October 2006 and is held for one month; and (ii) the portfolio is annually formed at the end of each year from 1981 to 2005 and is held for one year. In the end, I obtain a time-series sequence of out-of-sample GMV portfolio monthly returns. For the benchmarks, I also calculate equal-weighted and value-weighted portfolio returns during the same holding period.

To evaluate out-of-sample performance, I compute the following quantities: mean and variance of return, Sharpe ratio, average number of stocks to long and to short, average of the lowest and

[17] These ratios are robust to a different specification of the parametric covariance model. For instance, under the Gamma function, the mean of x_t^{ew} (x_t^{vw}) is 66.7 percent (62.3%). In addition, the ratios are all around 60 to 70 percent in various subsamples. These results are available upon request.

the highest portfolio weight, average of portfolio short interest, and average of portfolio turnover. The portfolio turnover (PTO), as in DeMiguel et al. (2007), is defined to be the sum of the absolute value of the trades across N assets:

$$\begin{aligned} PTO &= \sum_{i=1}^{N} \left| w_{i,t+1} - w_{i,t^+} \right| \\ w_{i,t^+} &= \frac{w_{i,t}(1 + r_{i,t+1})}{\sum_j w_{j,t}(1 + r_{j,t+1})}, \end{aligned} \tag{17}$$

where $w_{i,t+1}$is the calculated portfolio weight of asset i at $t+1$, based on information available up to time t, w_{i,t^+} is the portfolio weight before rebalancing at $t+1$, and $r_{i,t+1}$ is the return at $t+1$.

Table 2 shows the performance of GMV portfolios. The results are striking. In terms of diversification, holding the GMV portfolio, where the forecasted covariances are based on the K-Bessel spatial model, is like holding a "market" portfolio. For example, in the case of annually forming portfolios with no short sale constraint (Panel D), an investor holds 397 stocks on average (recall the average number of all available stocks is 401). Yet the annualized volatility of this GMV portfolio (12.4%) is lower than that of the simple value-weighted portfolio (14.2%). In sharp contrast, the average number of stocks to hold under factor-based models is only about $50-60$ and the annualized portfolio volatility (about $15-15.6\%$) is higher.

More importantly, the GMV portfolio weights are much more reasonable and less extreme under the spatial model than under factor-based models. Although the spatial model requires holding a number of stocks almost 8 times larger than the factor model, it generates portfolio turnover that is in fact 2-3 times smaller. While factor-based covariance models generate lower portfolio volatility when the holding period is one month (Panels A and C), they generate substantially higher portfolio turnover and lead to portfolio short interest that is unreasonably large. When the holding period is one year (Panels B and D), the GMV portfolios using factor-based covariance models perform poorly out-of-sample and cannot beat the simple equal-weighted or value-weighted portfolio strategy.

In summary, the portfolio constructed using spatial covariance estimates rather than factor-based covariance estimates is more consistent with the definition of a GMV portfolio.

5.2.3 Optimal Tangency Portfolio

Comparison based on the performance of the GMV portfolio comes with a cost that the minimum variance portfolio, on average, tends to select stocks with low variances and covariances. In practice, since investors believe that higher risks are associated with higher returns, actual portfolios, on average, are likely to contain stocks with high variances and covariances. As a result, it is desirable to study the optimal tangency portfolio from an investment perspective.

In order to fully investigate the accuracy of tangency portfolios over a full range of risks and mitigate the difficult problem of forecasting mean returns, I simulate the distribution of expected returns of individual stocks (Engle and Sheppard, 2003; Elton et al., 2006). Specifically, at the end of each month from December 1981 to October 2006, I first simulate the expected return of each firm in the next month, assuming its realized returns in the past 60 months are generated from a Normal distribution. Then, based on the same ex-ante information set of expected returns, I construct the optimal tangency portfolio, hold it for one month, and reform it at the end of the next month. During one simulation experiment, I obtain a time-series sequence of out-of-sample tangency portfolio returns. I repeat the experiment 1000 times and report the summary statistics. The simulation helps investigate how sensitive the covariance structure is to estimation error in expected return, and the performance of the optimal tangency portfolio helps measure the economic value of the spatial covariance in such a context.

Since some investors are allowed to short stocks in practice, I consider two additional cases where short sales are restricted: the 120/20 enhanced active strategy constraint and the Regulation T constraint (Jacobs and Levy, 2007; Pástor and Stambaugh, 2000). In addition to the performance measures used in the GMV portfolio, I calculate the fee that a risk-averse investor would be willing to pay to switch from a characteristic-based factor model to a characteristic-based spatial model (Fleming et al. 2001). Assuming that an investor has realized quadratic utility and a constant relative risk aversion coefficient, I use average realized utility to consistently estimate the expected utility for a given level of initial wealth. The maximum performance fee that she would be willing to pay is the solution to the following equation:

$$\sum_{t=1}^{T}\left[(R_{p,t}^{cha}-\delta)-\frac{\gamma}{2(1+\gamma)}(R_{p,t}^{cha}-\delta)^2\right]=\sum_{t=1}^{T}\left[R_{p,t}^{fac}-\frac{\gamma}{2(1+\gamma)}(R_{p,t}^{fac})^2\right], \tag{18}$$

where δ is the fee that is expressed as a fraction of wealth invested, γ is the risk aversion coefficient, $R_{p,t}^{cha}$ is the gross return of the tangency portfolio under a characteristic-based spatial covariance structure, and $R_{p,t}^{fac}$ is the gross return of the tangency portfolio under a characteristic-based factor covariance structure. To calculate the fee after transaction costs, I set the proportional transaction cost equal to 50 basis points per transaction (DeMiguel et al., 2007).

Table 3 presents the results of the optimal tangency portfolios. There are considerable utility gains for an investor to exploit the spatial comovement pattern. First, the spatial model produces a higher Sharpe ratio under each investment strategy. In the case of no short sale constraint (Panel B), factor models produce larger Sharpe ratio in only $12-26$ trials out of total 1000 trials. In other cases, there is no single time that factor models outstand (pseudo p-values are all zero).

Second, the spatial model helps an investor realize a significant economic value under each investment strategy. In the unconstraint case (Panel A), she would pay on the order of 63.4 to 89.5 percent (depending on risk aversion) of her wealth per year to switch from the Fama-French three-factor model to the K-Bessel spatial model. With transaction costs, she would pay on the order of 67.3 to 84.8 percent of her wealth per year. In constraint cases (Panels B-D) where factor models lead to the improved Sharpe ratios, she would still like to pay on the order of 1.36 to 17.9 percent before transaction costs, and 1.49 to 15.8 percent after transaction costs. These utility gains are substantial from an investment perspective.

Third, the spatial model produces more reasonable portfolio weights. For example, in the case of the 120/20 active strategy constraint (Panel D), the K-Bessel spatial model chooses the number of stocks to long (short) that is 2.6 to 3.7 (2.4 to 3.8) times larger than the factor model. However, it generates the maximum (minimum) portfolio holding that is 2.9 to 3.8 (2.4 to 3.2) times smaller. In the unconstraint case (Panel A) where the more extreme portfolio weights are observed, the K-Bessel spatial model generates portfolio turnover that is 25 to 60 times smaller. The dramatic difference in turnover is crucial to portfolio transactions.

5.3 Conditioning of Covariance Matrix

What can explain the striking differences in the portfolio performances of the characteristic-based spatial model and the characteristic-based factor model? In previous sections, I argue that the specification of comovement structure is one reason. In this section, I show that the property of the estimated covariance matrix is another.

Ahn et al. (2006) discuss the conditioning of the covariance matrix in asset pricing. In the theory of matrix algebra, the second-norm condition number of a covariance matrix (the ratio between the largest and the smallest eigenvalue) gauges the sensitivity of covariances to measurement error and perturbation in the data. This is particularly relevant in mean-variance portfolio optimization, since the estimates of expected returns based on realized returns are noisy. The bigger the condition number of a covariance matrix, the more collinearity in the data and the more sensitive of portfolio weights to the error. As a result, the ill-behaved covariance matrix will certainly magnify the error embedded in the estimated mean returns. Ironically, portfolio optimization becomes an error magnification. The assets allocation based on the unreliable input of the covariance matrix will eventually lead to poor out-of-sample performance.

Table 4 presents the summary statistics of condition numbers. At the end of each month from December 1981 to October 2006, I estimate the covariance matrix of excess returns of the S&P 500 stocks in my sample, using the data in the prior 60 months. The covariance models are the same ones used in studying the GMV portfolio before. The time-series summary statistics show that the condition number of the covariance matrix produced by the spatial model is much smaller than that produced by the factor model. For example, the mean (median) under the K-Bessel spatial model is 40 (35), whereas it is 1094 (1058) under the Fama-French three-factor model. In fact, the maximum condition number under the K-Bessel spatial model (206) is substantially less than the minimum condition number under each of factor models (293, 377, and 395 for the single-index model, the Fama-French three-factor model, and the Carhart four-factor model, respectively). As we have seen in the optimal tangency portfolio context, the huge condition number of the covariance matrix causes portfolio weights to be highly sensitive to estimation error in expected returns, and consequently leads to a significant utility loss in portfolio investment.

6 Robustness Checks

In this section, I examine whether there is a spatial covariance pattern of abnormal returns under an alternative risk-adjustment method. As stated in equation (5), I run a conditional version of the Fama-French factor regression to estimate benchmark returns. Since I need to estimate factor loadings, this approach introduces more noise than the method of the characteristic-matched portfolio. On one hand, if the comovement of abnormal returns is indeed associated with firm similarity, I should still be able to detect spatial covariance patterns, since all firms' characteristics are unchanged under both risk-adjustment methods. On the other hand, noisy estimates of abnormal returns under a conditional factor model make it difficult for a local average method to reliably estimate spatial covariances.

Figure 5 plots the spatial covariances and the bootstrap acceptance region of spatial independence against the same characteristic-based distances used in Figure 2. The spatial comovement pattern strongly shows up. The abnormal returns, under factor risk adjustment, covary strongly in month t when firms are similar in month $t-1$. The magnitudes of estimated covariances in Figure 5, however, are almost twice the size of those in Figure 2. In addition, covariances are uncorrelated among 15 percent of distances in Figure 5 and 25 percent of distances in Figure 2. In other words, abnormal returns tend to comove stronger (weaker) when benchmark returns are based on common factors (characteristic-matched portfolios).

My findings in the previous section indicate that the spatial covariation of abnormal returns cannot be associated with 10 Fama-French industry shocks, in addition to an aggregate shock. In this section, I examine whether this conclusion is robust to different industry specifications. Specifically, I consider the Fama-French 5- and 49-industry classifications based on four-digit SIC codes, and the Moskowitz-Grinblatt 20-industry classification based on two-digit SIC codes.

Figure 6 plots the bootstrap acceptance region of common shocks under different industry specifications. The acceptance regions move around, but none of them completely cover the range of distances at which spatial covariances are statistically different from zero (range from 1.02 to 7.23). Over this range of distances, the acceptance regions systematically exclude spatial covariances on the short distance (firms are more similar in characteristics) and the long distance (firms are less similar in characteristics). This observation implies that the characteristic-based spatial comovement pattern is not necessarily compatible with the factor-based covariance structure. No matter how many additional common shocks from industries are considered, the pervasive factors can only partially characterize the systematic comovement that is explicitly linked to characteristics.

I also check the spatial covariance patterns in various subsamples. In particular, I consider two half samples and six five-year samples. I use the local average method to estimate spatial covariances, and find the reliable spatial comovement pattern in each of these subsamples. Moreover, I find substantial time-varying patterns of these covariances. For example, the largest covariance at the small distance during December 1996 to November 2001 is about 0.038 percent, whereas it is only about 0.006 percent during December 1991 to November 1996. For brevity, the figures are not shown here but these results are available upon request.

7 Conclusion

The results in this paper are easy to summarize. First, I show that the abnormal returns of S&P 500 stocks display a strong spatial covariance pattern that is consistent with the prediction from a random field model. In particular, firm characteristics (size, past return, and accounting-based fundamentals) are directly associated with future return comovement. This type of spatial covariance structure cannot be fully characterized by a few pervasive factors, such as market and industry factors.

Second, I suggest a new covariance estimator in portfolio optimization. The parametric covariance function based on pairwise economic distance provides a parsimonious way to model a large number of stock returns. In comparison to factor-based covariance models, a spatial covariance model brings substantial diversification benefits and utility gains for an investor seeking the global minimum variance portfolio and the optimal tangency portfolio, respectively.

The random field approach presented in this paper can be used to study comovement beyond stock return. For example, we can investigate how firm characteristics are linked to liquidity commonality and construct a better measure of covariance with market liquidity. In addition, firm characteristics may play an important role as the link between liquidity comovement and return comovement. These interesting topics are left for the future research.

References

Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar, 2006, Basis Assets, *The Review of Financial Studies,* forthcoming.

Avramov, Doron, and Tarun Chordia, 2006, Asset Pricing Models and Financial Market Anomalies, *The Review of Financial Studies* 19, 1001-1040.

Bester, Alan C., 2004, Random Field and Affine Models for Interest Rates: An Empirical Comparison, working paper, University of Chicago.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2007, Parametric Portfolio Policies: Exploiting Characteristics in the Cross Section of Equity Returns, *The Review of Financial Studies*, forthcoming.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns, *Journal of Financial Economics* 49, 345-373.

Briner, Beat G., and Gregory Connor, 2007, How Much Structure is Best? A Comparison of Market Model, Factor Model, and Unstructured Equity Covariance Matrices, *Journal of Risk*, forthcoming.

Carhart, Mark, 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 1998, The Risk and Return from Factors, *Journal of Financial and Quantitative Analysis* 33, 159-188.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 1999, On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model, *The Review of Financial Studies* 12, 937-974.

Cohen, Randolph B., and Christopher K. Polk, 1998, The Impact of Industry Factors in Asset-Pricing Tests, working paper, Harvard Business School.

Conley, Timothy G., and Bill Dupor, 2003, A Spatial Analysis of Sectoral Complementarity, *Journal of Political Economy* 111, 311-352.

Connor, Gregory, and Robert Korajczyk, 1995, The Arbitrage Pricing Theory and Multifactor Models of Asset Returns, *Finance Handbooks in Operations Research and Management Science,* Volume 9, Chapter 4.

Connor, Gregory, and Robert Korajczyk, 2007, Factor Models of Asset Returns, *Encyclopedia of Quantitative Finance*, forthcoming.

Connor, Gregory, and Oliver Linton, 2006, Semiparametric Estimation of a Characteristic-based Factor Model of Common Stock Returns, *Journal of Empirical Finance*, forthcoming.

Cressie, Noel A., 1993, Statistics for Spatial Data, (Wiley Series in Probability and Statistics, London).

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russell Wermers, 1997, Measuring Mutual Fund Performance with Characteristic-based Benchmarks, *Journal of Finance* 52, 1035-1058.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the Characteristics of Cross Sectional Variation in Stock Returns, *Journal of Finance* 52, 1-34.

DeAngelo, Harry, Linda DeAngelo, and Rene M. Stulz, 2006, Dividend Policy and the Earned/Contributed Capital Mix: a Test of the Life-Cycle Theory, *Journal of Financial Economics* 81, 227-254.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, forthcoming.

Elton, Edwin J., and Martin J. Gruber, 1973, Estimating the Dependence Structure of Share Prices - Implications for Portfolio Selection, *Journal of Finance* 28, 1203-1232.

Elton, Edwin J., Martin J. Gruber, and Jonathan Spitzer, 2006, Improved Estimates of Correlation Coefficients and Their Impact on Optimum Portfolios, *European Financial Management* 12, 303-318.

Engle, Robert F., 2002, Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models, *Journal of Business and Economic Statistics* 20, 339-350.

Engle, Robert F., and Kevin Sheppard, 2003, Theoretical Properties of Dynamic Conditional Correlation Multivariate GARCH, working paper, New York University.

Engle, Robert F., and Bryan Kelly, 2008, Dynamic Equicorrelation, working paper, New York University.

Engle, Robert F., Victor K. NG, and Michael Rothschild, 1990, Asset Pricing with a Factor-GARCH Covariance Structure, *Journal of Econometrics* 45, 213-237.

Fama, Eugene F., and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1995, Size and Book-to-market factors in Earnings and Returns, *Journal of Finance* 50, 131-156.

Fama, Eugene F., and Kenneth R. French, 1997, Industry Costs of Equity, *Journal of Financial Economics* 43, 153-193.

Fama, Eugene F., and Kenneth R. French, 2001, Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay, *Journal of Financial Economics* 60, 3-43.

Fan, Jianqiang, and Irene I. Gijbels, 1996, Local Polynomial Modelling and Applications (Chapman and Hall, London).

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2001, The Economic Value of Volatility Timing, *Journal of Finance* 56, 329-352.

Gao, George P., 2009, Characteristic-Based Covariances and Cross-Sectional Expected Returns, working paper, University of Chicago.

Goldstein, Robert S., 2000, The Term Structure of Interest Rates as a Random Field, *The Review of Financial Studies* 13, 365-384.

Hall, Peter, and Prakash Patil, 1994, Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields, *Probability Theory and Related Fields* 99, 399-424.

Jacobs, Bruce I., and Kenneth N. Levy, 20 Myths about Enhanced Active 120-20 Strategies, *Financial Analysts Journal* 63, 19-26.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance* 58, 1651-1683.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91.

Kennedy, D.P., 1994 Characterizing Gaussian Models of the Term Structure of Interest Rates, *Mathematical Finance*, 4, 247-258.

Ledoit, Olivier, and Michael Wolf, 2003, Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection, *Journal of Empirical Finance* 10, 603-621.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-Snooping Biases in Tests of Financial Asset Pricing Models, *The Review of Financial Studies* 3, 431-467.

Mardia, Kanti V., John T. Kent, and John M. Bibby, 1979, Multivariate Analysis, (Academic Press, London).

Markowitz, Harry M., 1952, Mean-Variance Analysis in Portfolio Choice and Capital Markets, *Journal of Finance* 7, 77-91.

Moskowitz, Tobias J., 2003, An Analysis of Covariance Risk and Pricing Anomalies, *The Review of Financial Studies* 16, 417-457.

Ou, Jane A., and Stephen H. Penman, 1989, Financial Statement Analysis and The Prediction of Stock Returns, *Journal of Accounting and Economics* 11, 295-329.

Pástor, Lubas, and Robert F. Stambaugh, 2000, Comparing Asset Pricing Models: An Investment Perspective, *Journal of Financial Economics* 56, 335-381.

Piotroski, Joseph D., 2000, Value Investing: the Use of Historical Financial Statement Information to Separate Winners from Losers, *Journal of Accounting Research* 38 (Supplement), 1-42.

Sharpe, William F., 1963, A Simplified Model for Portfolio Analysis, *Management Science* 9, 277-293.

Appendix: Construction of Characteristic Variables and Economic Distance

In this appendix, I provide the details on the construction of firm characteristic variables and economic distance.

The following price-related variables are extracted from CRSP monthly file:

- ME (SIZE): the natural log of the firm's market capitalization (closing price times shares outstanding).
- AVGME_12, AVGME_6, AVGME_3: the natural log of the average ME over the past 12, 6, and 3 months, respectively.
- RET2_3: the two-month cumulative return over the period $t-3$ to $t-2$. The most recent one month is skipped to avoid the short-run reversal effect, as documented in the momentum literature.
- RET4_6: the three-month cumulative return over the period $t-6$ to $t-4$.
- RET7_12: the six-month cumulative return over the period $t-12$ to $t-7$.
- RET_12 (MOM): the past one-year cumulative return; the sum of RET2_3, RET4_6 and RET7_12 (requiring that the stock has available returns of at least 6 months).
- RET_6: the past half-year cumulative return; the sum of RET2_3 and RET4_6.
- AVGRET_12, AVGRET_6, AVGRET_3: the average monthly return over the past 12, 6, and 3 months, respectively.

The following accounting-based fundamental variables are extracted from Compustat quarterly file:

- TA: the natural log of the firm's total assets (data item 44).
- BE: the natural log of the firm's book value of equity; for calculation of BE, see Appendix A of data definitions in Fama and French (2001).
- ROA: return on assets; income before extraordinary items (data item 8), plus interest expense (data item 22) and income statement deferred taxes (if available, data item 35) at time t, divided by total assets at time $t-1$.

- ROE: return on equity; earnings available for common equity (data item 25) at time t, divided by book value of equity at time $t-1$.
- RE/TA: retained earnings (data item 58) to total assets; the firm's life cycle stage measure is suggested in DeAngelo et al. (2006).
- AGR, SGR: total assets growth rate and net sales (data item 2) growth rate, respectively.
- LEV: leverage ratio, defined as total long-term debt (data item 51) to average total assets.
- LIQ: the natural log of that current ratio, which is current assets (data item 40) to current liabilities (data item 49).
- MARG: current gross margin ratio, defined to be one minus the ratio of the cost of goods sold (data item 30) to net sales.
- TO: the natural log of the accounting turnover ratio, which is net sales at time t divided by total assets at time $t-1$.
- ROA_12, ROE_12: the past one-year cumulative return on assets and return on equity, respectively.

In addition, I calculate the average number of the following variables over the past 4 quarters: TA, BE, ROA, ROE, RE/TA, AGR, SGR, LEV, LIQ, MARG, and TO.

The following fundamental-to-price ratio variables are extracted from CRSP/Compustat file:

- BM: the natural log of book-to-market equity (book value of equity to market value of equity); only firms with positive book value of equity are retained.
- S/P: the natural log of net sales to market value of equity.
- CF/FV: the natural log of cash flow to firm value; cash flow is the operating income before depreciation (data item 21) and firm value is defined as total assets, minus book value of common equity (data item 59), minus accounts payable (data item 46), plus market value of common equity.

- D/P: the natural log of the dividend yield, which is cash dividend to common equity (data item 20), divided by equity market capitalization.
- E/P: the natural log of earnings yield, which is income before extraordinary items available to common equity (data item 25), divided by equity market capitalization.

In addition, I calculate the average number of the following variables over the past 4 quarters: BM, S/P, CF/FV, D/P, and E/P.

The Euclidean economic distances are constructed by firm characteristic variables. To ensure the consistency of measurement scales, each variable is first Z-scored before calculating the distance. For example, I use ME, BM, and MOM to construct the distance between firms i and j at time t as the following:

$$d_{ij,t}(ME, BM, MOM) \triangleq \sqrt{(me_{i,t} - me_{j,t})^2 + (bm_{i,t} - bm_{j,t})^2 + (mom_{i,t} - mom_{j,t})^2},$$

where each coordinate is Z-scored as:

$$me_{i,t} \triangleq \frac{ME_{i,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} ME_{n,t}}{\sqrt{\frac{1}{N_t-1}\sum_{n=1}^{N_t}\left(ME_{n,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} ME_{n,t}\right)^2}},$$

$$bm_{i,t} \triangleq \frac{BM_{i,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} BM_{n,t}}{\sqrt{\frac{1}{N_t-1}\sum_{n=1}^{N_t}\left(BM_{n,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} BM_{n,t}\right)^2}},$$

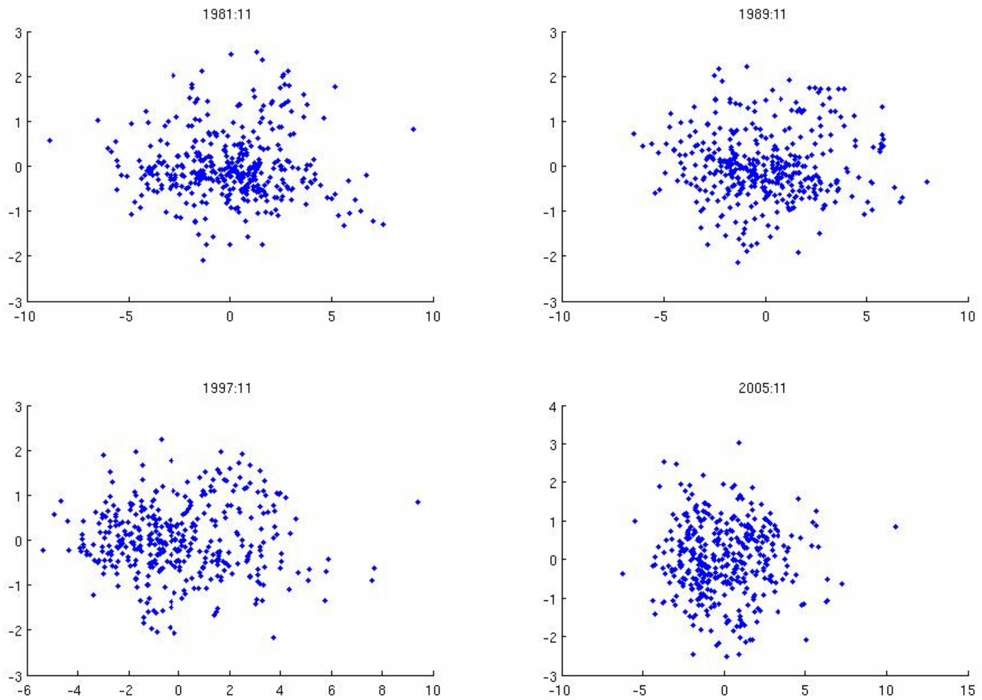
and

$$mom_{i,t} \triangleq \frac{MOM_{i,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} MOM_{n,t}}{\sqrt{\frac{1}{N_t-1}\sum_{n=1}^{N_t}\left(MOM_{n,t} - \frac{1}{N_t}\sum_{n=1}^{N_t} MOM_{n,t}\right)^2}}.$$

Figure 1: Classical multidimensional scaling (CMDS) representation of economic distance

This figure plots CMDS representation of distances among non-financial firms in the S&P500 index on the plane (panel A) and the surface (panel B). Euclidean distances in original space with 12 coordinates are based on accounting-based fundamentals, including capital structure, accounting liquidity, and operating efficiency (see text for details). I choose four representative months and the goodness-of-fit statistics range from 80 percent to 90 percent.

Panel A: Configuration in 2-Dimension



Panel B: Configuration in 3-Dimension

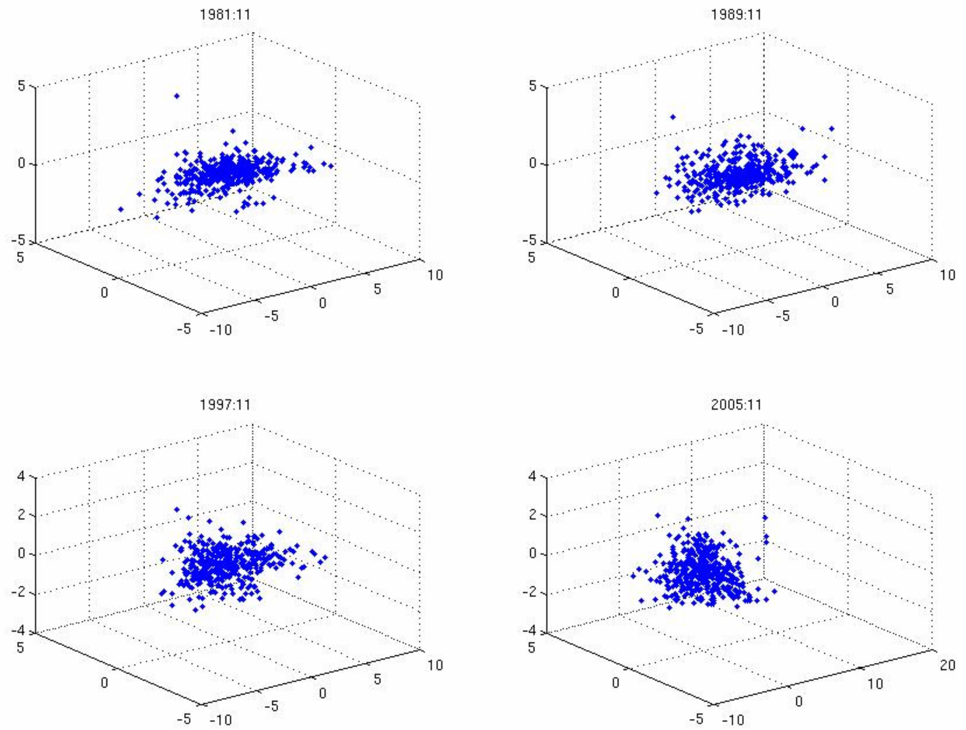


Figure 2: Characteristic-based spatial covariance and histogram of economic distance over the full sample (1975/12 – 2006/11)

The top panel plots the spatial covariances (solid line) and the bootstrap acceptance region of spatial independence (dotted lines) against the characteristic-based economic distances for non-financial firms in the S&P 500 index from December 1975 to November 2006. Risk-adjusted abnormal returns are estimated by characteristic-matched portfolio returns (Daniel et al., 1997), and spatial covariances are nonparametrically estimated using a local average method. The bottom panel plots the cumulative empirical distribution of "SMCE" distances based on the following characteristics: size (ME, AVGME_3, AVGME_6, AVGME_12), momentum (RET2_3, RET4_6, RET7_12, RET_6, RET_12), average return (AVGRET_3, AVGRET_6, AVGRET_12), capital structure (TA, BE, LEV), accounting liquidity (LIQ), operating efficiency (MARG, TO), and the average level of each of accounting variables over the past four quarters.

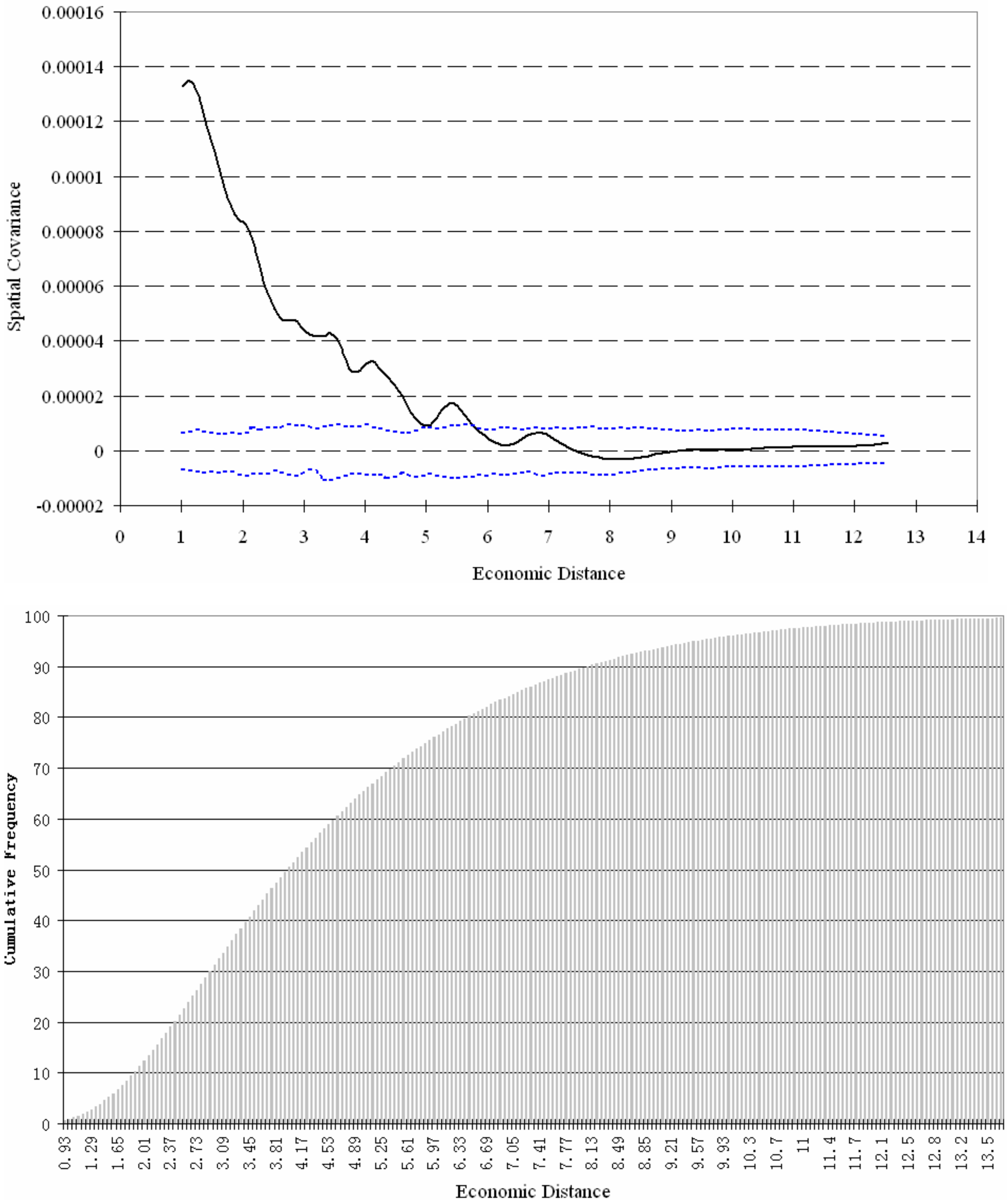


Figure 3: Test common shocks hypothesis for the characteristic-based spatial covariance pattern over the full sample (1975/12 – 2006/11)

This figure plots the spatial covariances (solid line) and the bootstrap acceptance region of common shocks hypothesis (dotted lines) for non-financial firms in the S&P 500 index from December 1975 to November 2006. Risk-adjusted abnormal returns are estimated by characteristic-matched portfolio returns (Daniel et al., 1997), and spatial covariances are nonparametrically estimated using a local average method. Firm characteristics are the same as those used in Figure 2. To test the null hypothesis, I simulate an aggregate market shock, plus ten industry-specific common shocks (allowing arbitrary correlations across industries). The Fama-French 10-industry classification is used throughout the simulation.

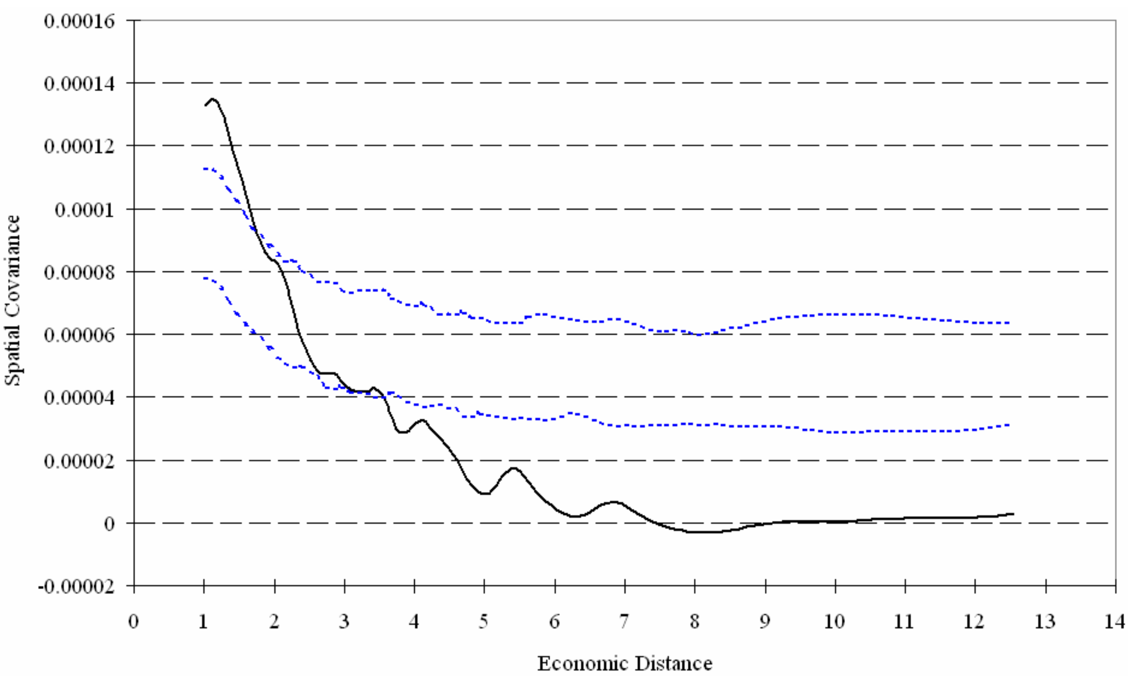


Figure 4: Three-dimensional spatial covariance surfaces over the full sample (1975/12 – 2006/11)

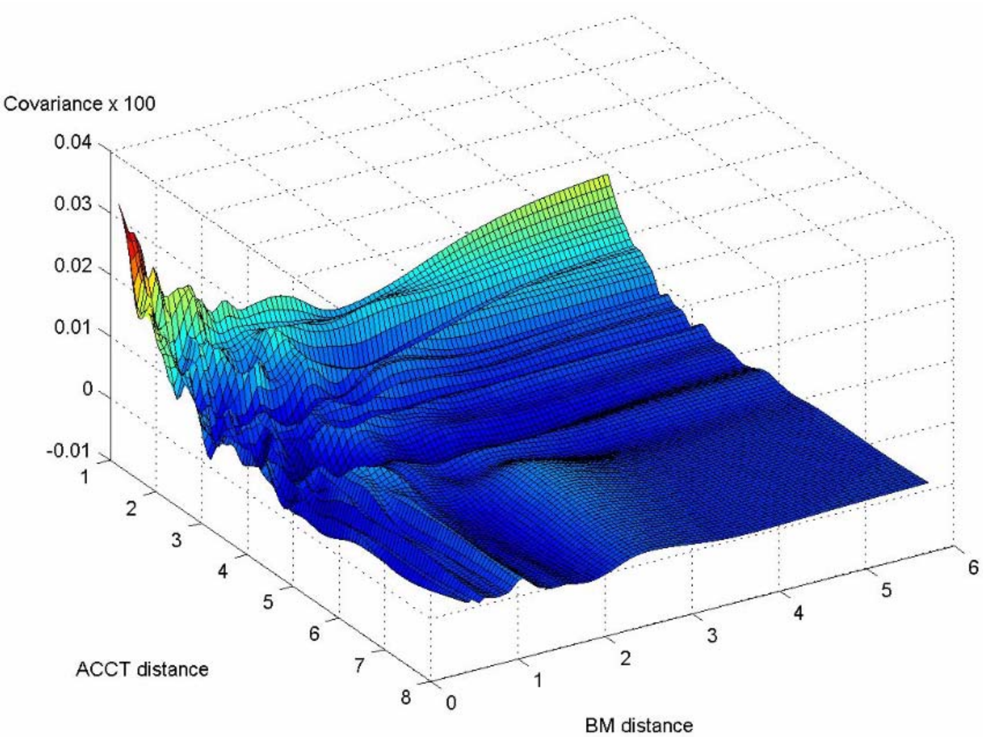
The figure plots the three-dimensional surfaces of spatial covariance (in percent) on various combinations of characteristic-based economic distance for non-financial firms in the S&P500 index from December 1975 to November 2006. Risk-adjusted abnormal returns are estimated by characteristic-matched portfolio returns (Daniel et al., 1997), and spatial covariances are estimated by two-dimensional nonparametric regressions.

In Panel A, "BM" distance is constructed from fundamental-to-price ratios, including book-to-market (BM), sales to price (S/P), cash flow to firm value (CF/FV), dividend yield (D/P), earnings yield (E/P), and the average level of each of them over past four quarters; "ACCT" distance is constructed from accounting-based fundamentals, including total assets (TA), book equity (BE), leverage ratio (LEV), ratio of retained earnings to total assets (RE/TA), return on equity (ROE), return on assets (ROA), sales growth rate (SGR), assets growth rate (AGR), current ratio (LIQ), accounting turnover ratio (TO), current gross margin ratio (MARG), and the average level of each of them over the past four quarters.

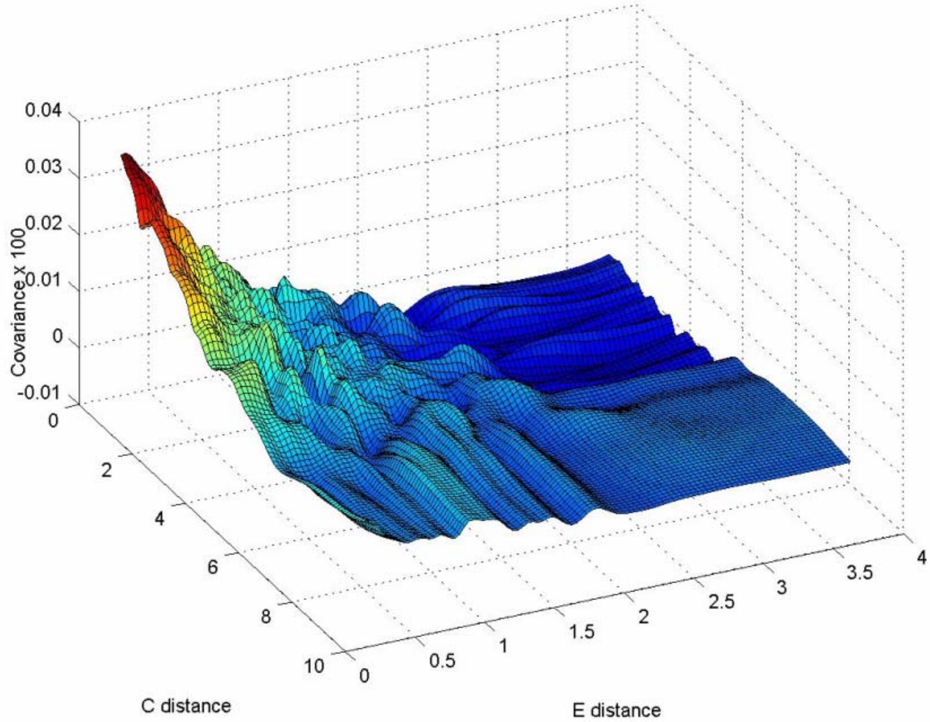
In Panel B, "C" distance is constructed from capital structure, including total assets, book equity, and leverage ratio; "P" distance is constructed from profitability, including the ratio of retained earnings to total assets, return on equity, return on assets, sales growth rate, and assets growth rate; "E" distance is constructed from accounting liquidity and operating efficiency, including current ratio, accounting turnover ratio, and current gross margin ratio. I also include the average level of each of these accounting variables over the past four quarters.

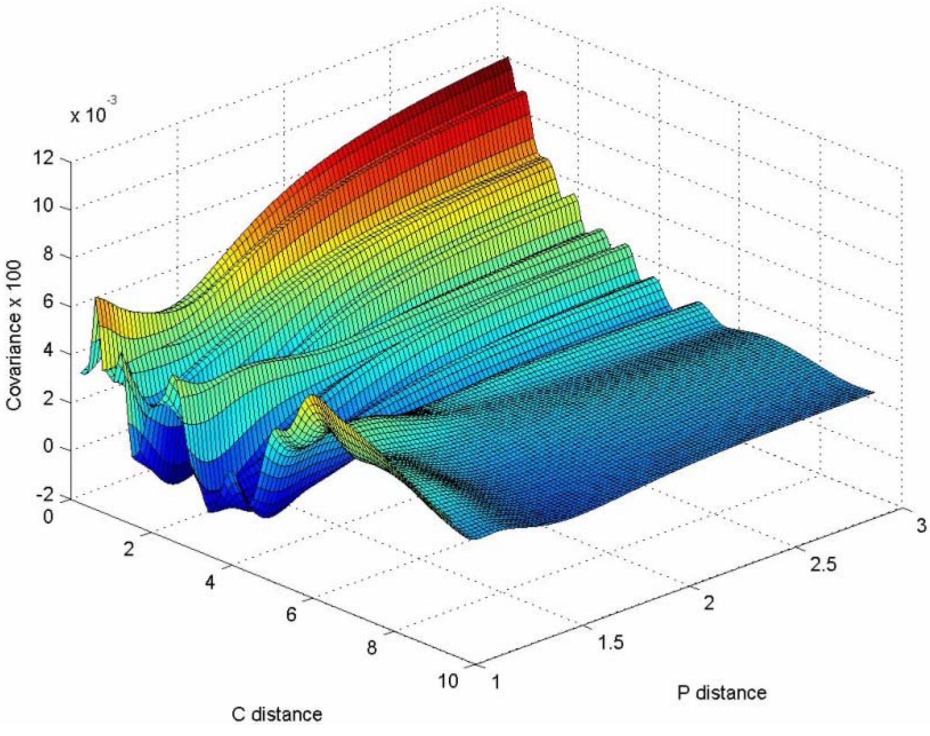
In Panel C, "SM" distance is constructed from price-related characteristics, including size (ME, AVGME_3, AVGME_6, AVGME_12), momentum (RET2_3, RET4_6, RET7_12, RET_6, RET_12), and past average return (AVGRET_3, AVGRET_6, AVGRET_12). Appendix provides the details of constructing these variables in parentheses.

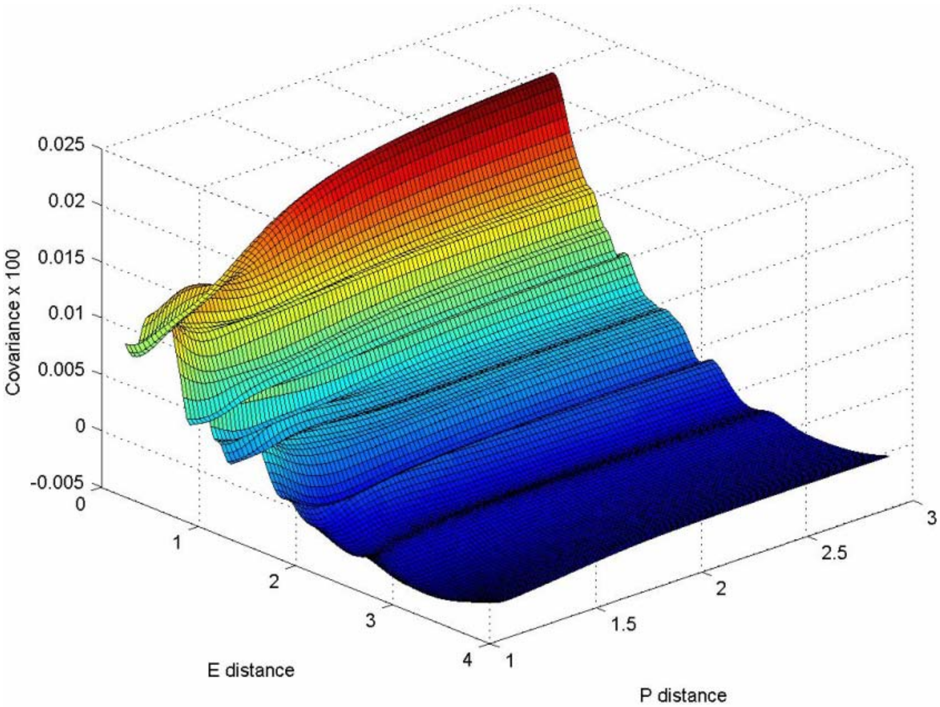
Panel A: "BM" vs. "ACCT" distance



Panel B: “C” vs. “E” vs. “P” distance







Panel C: "SM" vs. "CE" distance

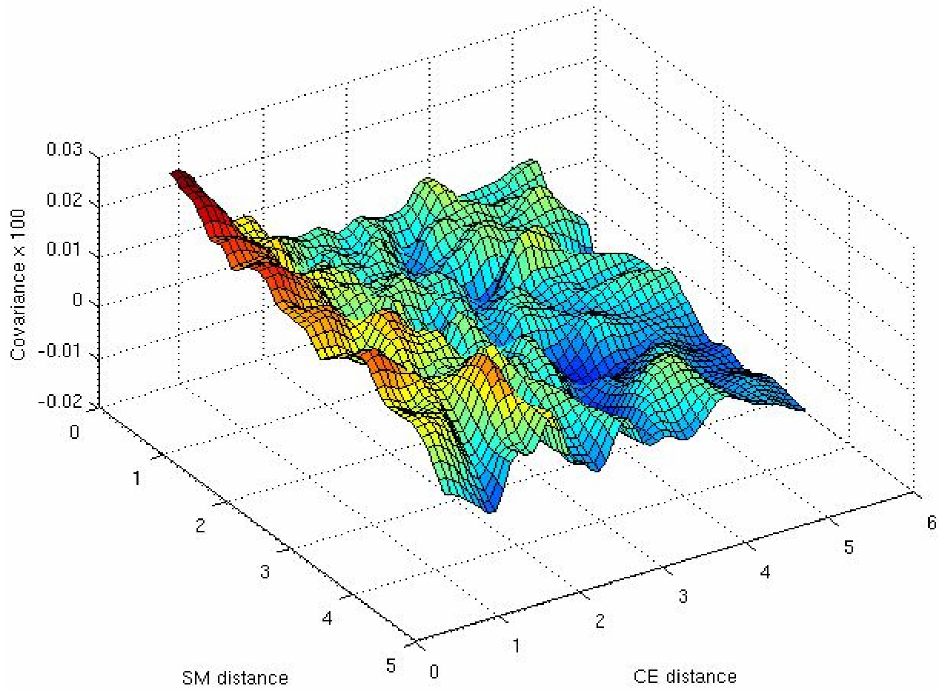


Figure 5: Characteristic-based spatial covariance over the full sample (1975/12 – 2006/11): an alternative risk-adjustment method for abnormal returns

This figure plots the spatial covariances (solid line) and the bootstrap acceptance region of spatial independence (dotted lines) against the characteristic-based economic distances for non-financial firms in the S&P 500 index from December 1975 to November 2006. Risk-adjusted abnormal returns are estimated from the conditional Fama-French four-factor regression, and spatial covariances are nonparametrically estimated using a local average method. Firm characteristics are the same as those used in Figure 2.

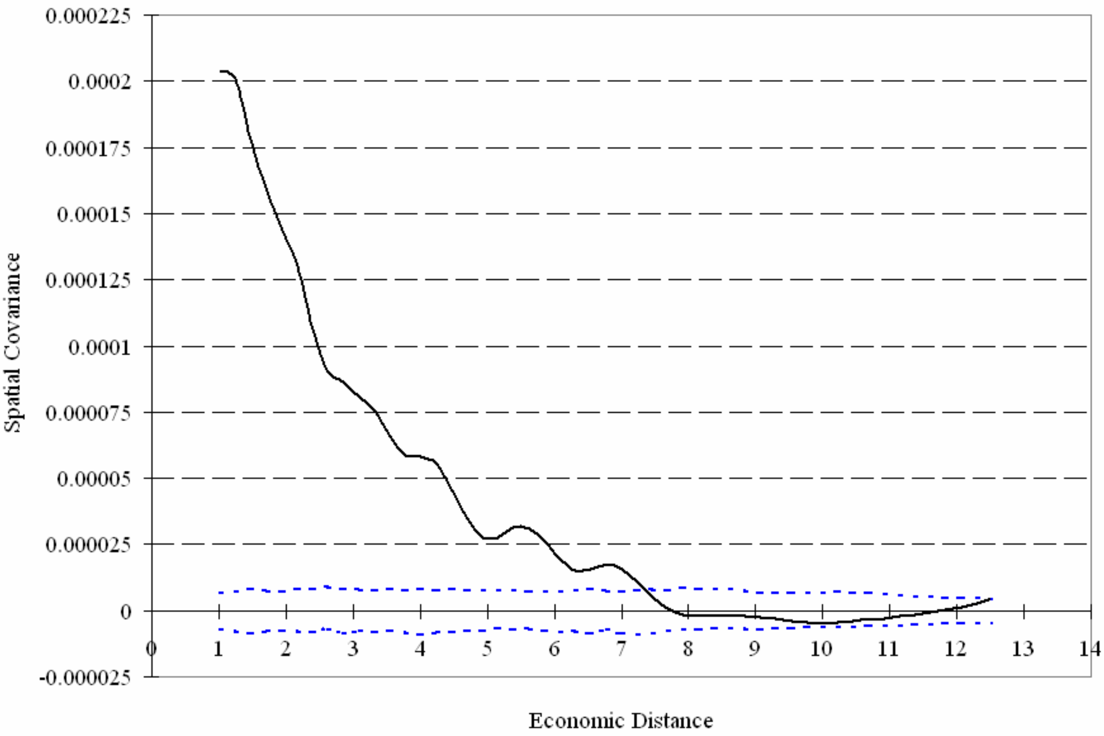


Figure 6: Test common shocks hypothesis for the characteristic-based spatial covariance pattern over the full sample (1975/12 – 2006/11): different industry specifications

This figure plots the spatial covariances (solid line) and the bootstrap acceptance region of common shocks hypothesis (dotted lines) for non-financial firms in the S&P 500 index from December 1975 to November 2006. Risk-adjusted abnormal returns are estimated from the conditional Fama-French four-factor regression, and spatial covariances are nonparametrically estimated using a local average method. Firm characteristics are the same as those used in Figure 2. To test the null hypothesis, I simulate an aggregate market shock plus industry-specific common shocks (allowing arbitrary correlations across industries), based on Fama-French 5- and 49-industry and Moskowitz-Grinblatt 20-industry classifications, respectively.

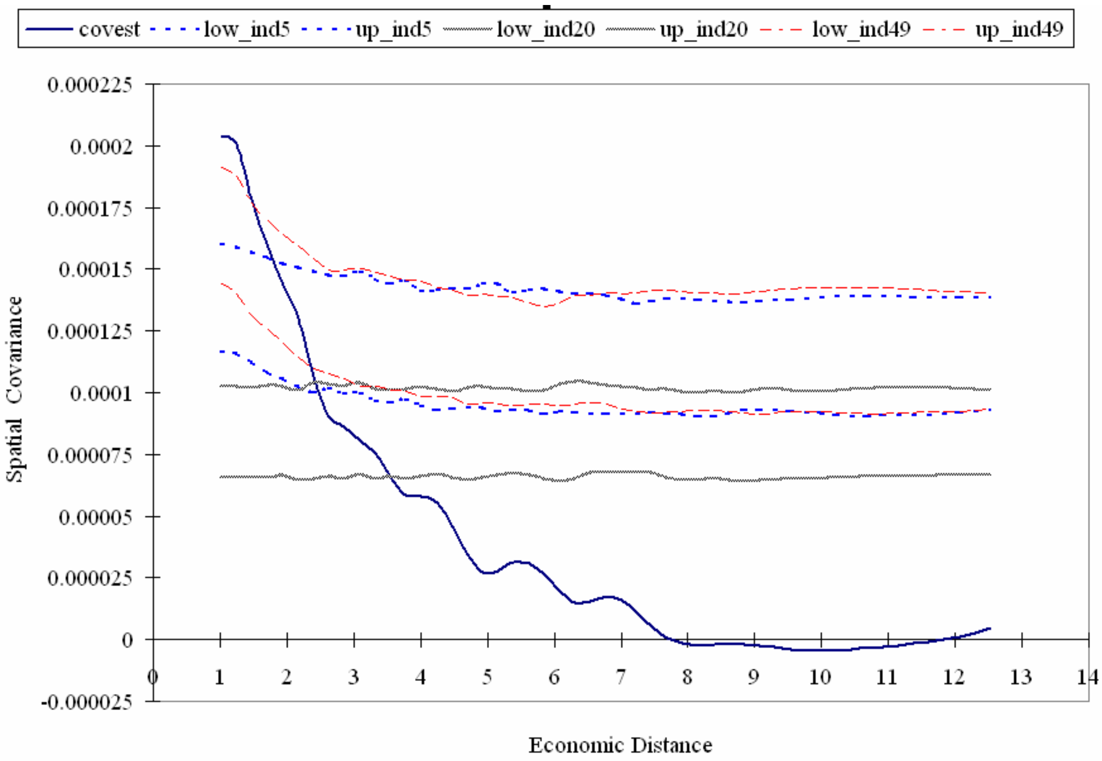


Table 1: Summary statistics of S&P 500 stocks

This table reports the first quartile, mean, median, third quartile, and standard deviation of the characteristics of the non-financial firms in the S&P 500 index from 1975/12 to 2006/10. At the end of each month, I obtain market equity ("ME"), size decile ranking by NYSE decile breakpoints ("ME_rank"), book-to-market equity ("BM"), book-to-market decile ranking by NYSE decile breakpoints ("BM_rank"), momentum ("RET_12"), the leverage ratio ("LEV"), the current ratio ("LIQ"), the accounting turnover ratio ("TO"), and the current gross margin ratio ("MARG"). Panels A, B, and C report summary statistics in each sub period, and Panel D reports summary statistics over all periods.

	ME	ME_rank	BM	BM_rank	RET_12	LEV	LIQ	TO	MARG
					Panel A: 1975/12 - 1986/2				
Q1	361.5	7.0	0.63	3.0	-0.037	0.09	1.40	0.20	0.19
mean	1819.5	7.9	1.17	5.1	0.125	0.19	2.11	0.33	0.08
median	850.0	9.0	1.01	5.0	0.122	0.18	1.87	0.31	0.27
Q3	1792.8	10.0	1.43	8.0	0.286	0.27	2.42	0.41	0.37
STD	3994.2	2.2	1.12	2.9	0.280	0.13	4.81	0.23	24.56
					Panel B: 1986/3 - 1996/6				
Q1	1231.2	7.0	0.37	2.0	-0.049	0.08	1.12	0.14	0.21
mean	5471.6	8.3	0.72	4.7	0.102	0.19	1.74	0.27	0.33
median	2701.6	9.0	0.59	4.0	0.115	0.18	1.52	0.25	0.32
Q3	5709.3	10.0	0.91	7.0	0.269	0.28	2.04	0.35	0.45
STD	9073.4	1.8	0.54	2.9	0.275	0.14	1.06	0.20	0.27
					Panel C: 1996/7 - 2006/10				
Q1	3997.9	8.0	0.23	2.0	-0.095	0.09	1.02	0.10	0.24
mean	19671.5	8.8	0.52	4.1	0.065	0.20	1.75	0.23	0.40
median	8151.3	9.0	0.39	3.0	0.096	0.18	1.41	0.20	0.37
Q3	17457.1	10.0	0.64	6.0	0.266	0.29	2.00	0.30	0.56
STD	38192.0	1.3	0.51	2.8	0.368	0.14	1.82	0.19	0.24
					Panel D: 1975/12 - 2006/10				
Q1	977.2	8.0	0.35	2.0	-0.059	0.09	1.16	0.13	0.21
mean	9153.6	8.3	0.79	4.6	0.097	0.19	1.86	0.28	0.27
median	2818.9	9.0	0.61	4.0	0.111	0.18	1.59	0.25	0.31
Q3	7634.7	10.0	1.03	7.0	0.273	0.28	2.20	0.36	0.46
STD	24308.5	1.8	0.81	2.9	0.312	0.14	3.04	0.21	13.73

Table 2: Ex-post mean, volatility, and other characteristics of global minimum variance portfolios

This table shows the performance of global minimum variance (GMV) portfolios under various covariance estimators: the characteristic-based spatial model (Bess), Sharpe's single index model (F1), the Fama-French three-factor model (FF-3), and the Fama-French three-factor plus Carhart momentum factor model (FF-4). All are estimated on a rolling-window basis using the monthly data in the prior 60 months.

In Panels A and C, portfolios are formed monthly at the end of month *t* from 1981/12 to 2006/10 and held for one month, imposing unconstraint and no short sale constraint, respectively. In Panels B and D, portfolios are formed at the end of each year from 1981 to 2005 and held for one year.

The following variables are calculated to measure out-of-sample performances of GMV portfolios: the annualized mean return in percent ("μ"), the annualized return volatility in percent ("σ"), the annualized Sharpe ratio ("SR"), the average number of stocks to short ("No. Short"), the average number of stocks to long ("No. Long"), the average of the lowest portfolio weight in percent ("minw"), the average of the highest portfolio weight in percent ("maxw"), the average of short interest in percent ("SI"), and the average of portfolio turnover ("PTO"). As a benchmark, I also report the naïve diversification strategies: equal-weighted (EW) and value-weighed (VW) portfolios of non-financial stocks in the S&P 500 index.

	μ	σ	SR	No. Short	No. Long	minw	maxw	SI	PTO
Panel A: Unconstraint, holding period 1 month									
Bess	9.04	13.84	0.483	13.0	383.8	-0.061	3.091	-0.34	0.09
F1	15.68	13.88	0.755	169.2	231.3	-1.262	3.701	-49.19	0.23
FF-3	13.57	12.12	0.690	162.9	237.6	-2.093	4.276	-67.65	0.33
FF-4	13.60	11.75	0.715	160.7	239.8	-2.370	4.441	-70.19	0.36
Panel B: Unconstraint, holding period 12 months									
Bess	11.02	13.41	0.642	14.8	386.8	-0.070	3.068	-0.41	0.34
F1	14.83	16.37	0.564	170.3	230.7	-1.223	3.541	-49.60	0.84
FF-3	12.52	15.44	0.456	163.3	237.6	-2.011	4.181	-69.04	1.27
FF-4	12.76	15.09	0.482	162.0	239.0	-2.230	4.377	-71.81	1.33
Panel C: No short sale constraint, holding period 1 month									
Bess	15.30	13.96	0.723	0	397.3	0.028	1.970	0	0.08
F1	14.24	12.71	0.711	0	58.4	0	8.321	0	0.15
FF-3	13.30	11.74	0.690	0	50.8	0	9.813	0	0.20
FF-4	13.39	11.74	0.697	0	49.6	0	9.695	0	0.21
EW	15.27	16.26	0.619	0	400.5	0.250	0.250	0	0.06
VW	13.53	14.72	0.566	0	400.5	0.005	4.682	0	0.02
Panel D: No short sale constraint, holding period 12 months									
Bess	14.83	12.41	0.750	0	397.2	0.029	1.898	0	0.25
F1	13.80	15.07	0.550	0	59.2	0	7.902	0	0.53
FF-3	12.53	15.64	0.454	0	51.0	0	9.177	0	0.70
FF-4	12.33	15.26	0.452	0	49.6	0	8.984	0	0.73
EW	14.89	13.09	0.719	0	400.9	0.249	0.249	0	0.19
VW	13.86	14.21	0.586	0	400.9	0.004	4.584	0	0.05

Table 3: Ex-post mean, volatility, and other characteristics of optimal tangency portfolios

This table reports the optimal tangency portfolio performance under various covariance estimators, based on the same ex-ante information set of simulated expected returns. At the end of each month *t* from 1981/12 to 2006/10, I first simulate a firm's expected return for next month *t*+1 assuming its realized returns in the prior 60 months are generated from a Normal distribution. Then I use the following models to predict covariances in the next month: the characteristic-based spatial model (Bess), Sharpe's single index model (F1), the Fama-French three-factor model (FF-3), and the Fama-French three-factor plus Carhart momentum factor model (FF-4). All are estimated on a rolling-window basis using the monthly data in the prior 60 months. Using these estimates, I compute the optimal tangency portfolio weights under different investment scenarios: fully unconstraint (Panel A), no short sale constraint (Panel B), 120/20 active strategy constraint (Panel C), and Regulation T constraint (Panel D). The tangency portfolio is held for one month and reformed at the end of month *t*+1.

I repeat the simulation experiment 1000 times. In each trial, the following variables are calculated to measure the out-of-sample performance of optimal tangency portfolios: the annualized mean return in percent ("μ"), the annualized return volatility in percent ("σ"), the annualized Sharpe ratio ("SR"), the average number of stocks to short ("No. Short"), the average number of stocks to long ("No. Long"), the average of the lowest portfolio weight in percent ("minw"), the average of the highest portfolio weight in percent ("maxw"), the average of short interest in percent ("SI"), the average of portfolio turnover ("PTO"), annualized fee in percent ("Δ1, Δ5, Δ10") that an investor with quadratic utility and constant relative risk aversion of γ=1, 5, 10, respectively, would be willing to pay to switch from the factor-based covariance estimator to the characteristic-based spatial covariance estimator, and annualized fees in percent under transaction costs ("TC_Δ1, TC_Δ5, TC_Δ10").

The table reports the average number of these variables over 1000 trials. "p-value" represents the proportion of trials in which the factor-based covariance model has a higher Sharpe ratio than the characteristic-based spatial covariance model.

	Panel A: Unconstraint				Panel B: No Short Sale Constraint			
	Bess	F1	FF-3	FF-4	Bess	F1	FF-3	FF-4
μ	15.63	12.32	286.34	-60.22	15.82	15.14	15.01	14.89
σ	19.93	581.36	1876.96	723.47	13.94	16.77	16.99	16.77
SR	0.507	0.063	0.004	-0.032	0.742	0.588	0.572	0.572
No. Short	153.7	190.2	190.3	190.0	0	0	0	0
No. Long	247.2	210.7	210.7	210.9	174.7	32.3	25.2	24.7
minw	-3.66	-63.51	-202.10	-132.53	0	0	0	0
maxw	5.15	66.14	204.54	139.61	3.10	15.37	17.80	17.59
SI	-123.85	-2552.93	-8727.70	-5485.13	0	0	0	0
PTO	4.07	94.24	230.20	242.66	1.32	1.55	1.60	1.61
Δ_1	-	67.53	63.41	72.28	-	1.18	1.36	1.44
Δ_5	-	90.63	85.47	89.90	-	6.11	6.54	6.26
Δ_{10}	-	94.03	89.50	92.78	-	16.14	16.46	16.05
TC_Δ_1	-	61.12	67.29	80.93	-	1.29	1.49	1.58
TC_Δ_5	-	80.63	81.44	89.69	-	5.73	6.17	5.92
TC_Δ_{10}	-	85.15	84.78	91.54	-	14.85	15.20	14.77
p-value	-	0	0	0	-	0.026	0.012	0.014

Table 3: continued

	Panel C: 120/20 Active Strategy Constraint				Panel D: Regulation T Constraint			
	Bess	F1	FF-3	FF-4	Bess	F1	FF-3	FF-4
μ	15.96	14.67	13.99	14.05	16.20	15.36	14.10	14.40
σ	14.15	16.58	17.57	17.47	15.14	17.92	18.81	18.25
SR	0.741	0.558	0.491	0.498	0.708	0.550	0.462	0.493
No. Short	58.1	24.4	16.0	15.1	106.0	66.8	48.9	46.5
No. Long	204.0	77.7	56.3	54.9	219.1	123.2	100.2	98.1
minw	-1.59	-3.85	-4.86	-5.06	-2.33	-3.66	-4.89	-5.18
maxw	3.18	9.19	11.86	12.02	3.63	6.64	8.61	9.02
SI	-20.00	-20.00	-20.00	-20.00	-49.99	-50.00	-50.00	-50.00
PTO	1.82	1.89	2.03	2.05	2.59	2.52	2.73	2.76
Δ_1	-	1.72	2.58	2.50	-	1.36	2.79	2.38
Δ_5	-	5.94	7.79	7.62	-	6.63	8.59	7.59
Δ_{10}	-	15.60	16.90	16.78	-	16.99	17.86	17.08
TC_Δ_1	-	1.74	2.66	2.59	-	1.29	2.82	2.43
TC_Δ_5	-	5.51	7.31	7.13	-	5.84	7.83	6.89
TC_Δ_{10}	-	14.07	15.40	15.25	-	14.82	15.81	14.96
p-value	-	0	0	0	-	0	0	0

Table 4: Summary statistics of covariance matrix condition number

At the end of each month *t* from 1981/12 to 2006/10, I estimate the return covariance matrix of non-financial firms in the S&P 500 index. The matrix condition number in the second-norm is defined as the ratio of the maximum to minimum eigenvalues of the estimated covariance matrix. The time-series of these matrix condition numbers are recorded. This table reports summary statistics.

The following covariance models are used to estimate covariance matrix: the characteristic-based spatial model (Bess), Sharpe's single index model (F1), the Fama-French three-factor model (FF-3), and the Fama-French three-factor plus Carhart momentum factor model (FF-4), all of which are estimated on a rolling-window basis using the monthly data in the prior 60 months.

	Mean	Median	Q1	Q3	Min	Max	StdDev
Bess	39.9	35.4	25.2	46.4	13.3	206.1	26.4
F1	865.2	873.2	639.5	1068.2	292.7	2168.6	312.9
FF-3	1093.9	1058.2	810.0	1433.9	376.5	2654.3	415.6
FF-4	1179.1	1147.4	831.0	1518.4	394.6	2826.6	462.5