Spatial Dependences in the Stylized Hedge Fund Returns

Joung Keun Cho*

Abstract

We apply an exploratory spatial data analysis framework for integrating the time series of hedge fund returns to its neighborhood, mapping, and local analysis for the feasible spatial modeling. Our approach takes into account option-like features and serial correlations in the stylized hedge funds' risk-return payoffs. By comparing the classic risk factor analysis of hedge fund performance of ordinary least squares regression with spatial autoregressive models, we investigate each model's respective ability to estimate the stylized risk premiums. The time series analysis of hedge fund returns from the Barclays Hedge indicates that, for some of the sub-investment styles such as equity long-short, equity long-bias, event-driven arbitrage, convertible arbitrage, fixed-income arbitrage, distressed securities, multi-strategies, and commodity trading advisors, the spatial autoregressive modeling may provide consistent estimates of factor risk-premiums by correcting structural spatial dependence through the measure of endogeneity of implied volatilities. We employ spatial specifications including spatial lag (SLM) and spatial error (SEM) models to minimize the overestimation bias in factor risk premiums by exploring some practical implications in an *ad hoc* screening through the missing spatial autoregressive heterogeneity in the ordinary least squares approach. Both SLM and SEM models are applied to verify a 'meant-to-be' spatial dependence to a relatively short time series of a recently failed credit hedge fund previously marketed its vanishingly rare talent of return predictability and consistency.

Key words: *Spatial Dependence, Spatial Lag, Spatial Error, Hedge Fund Performance Attribution*

JEL Classification: G23, G11

* Institutional Advisory to QCAM Currency Asset Management AG, U.S. Tax Advisory to Sellymon.com, Assistant Professor of Finance, School of Business, Seokyeong University 714 Hanlim Hall, 124 Seokyeong-ro, Seoul 02713, Korea. E-mail: joungkec@alumni.cmu.edu

[1] Introduction

Many hedge fund strategies exhibit non-linear risk-return payoffs as manifested through significant betas on option-based risk factors. Agarwal and Naik (2004) show that some equity-based hedge fund investment styles tend to show a relationship with option-based risk factors that consist of returns obtained by buying and selling one-month later liquid put and call options on the S&P 500 index. For instance, payoffs of event arbitrage, corporate restructuring, event-driven, relative value arbitrage, and convertible arbitrage trades resemble that from writing a put option on the equity market index and end up with a typical short-volatility risk-return profile. These are trading strategies collecting small but continuous option premiums based on normal economic activity, while susceptible to a large loss during the tail-events in the equity market. It may also be from the managers' intentional efforts to calibrate a payoff structure similar to that from writing a put option to improve their Sharpe ratios or to assimilate it to their incentive contracts. In practice, the standard assumption of normal, homoscedastic and uncorrelated error terms that lead to the best linear unbiased estimator (BLUE) characteristics of ordinary least squares (OLS) estimators are not necessarily satisfied in the performance attribution analysis by the real hedge funds data. Then checking the degree and extent to which these assumptions are violated may be the first step in the analysis of hedge fund returns. Further attention is required whenever the errors in the variables in the model show some degree of spatial dependence.

Various works of literature have pointed out that the returns of the hedge funds are serially correlated (i.e., autocorrelated) and systematically distorted due to their non-normal nature of return distributions. Cho and Kim (2017) study the market timing ability of Korean equity long-short hedge funds by introducing the concept of style-tilting volatility (STV) as a measure to quantify the amount of variation or dispersion of a set of individual fund's style exposures. Some hedge funds demonstrated enhanced risk-adjusted returns through a wide range of volatility timing behavior, while their active style bets did not necessarily result in performance persistence compared to the peer managers. Cho (2019) elaborates that the time-varying nature of the conditional moderation effect in the hedge fund managers by adaptively altering the subsequent month's risk-on positions proactively and not enough overall consistency in their exposures is revealed even in the list of well-documented systematic risk factors. The fact that the hedge funds' trading strategies adopt nonlinear derivative contracts or may engage in market or factor-timing leads to an asymmetric return distribution and the probable fat tails, which is far from Gaussian as the higher moments such as skewness and excess kurtosis significantly deviate from zero. A negative skewness and positive excess kurtosis (i.e., leptokurtic) are indicative of a higher probability of large losses than stipulated at Gaussian return distributions. Since the higher moments of the return distribution are not considered in the Sharpe ratio formula, the underestimation of the high tendency of large losses for any qualified hedge fund portfolio is also probable.

A positive autocorrelation characterized by a significantly positive beta coefficient in the time series of returns on its own lagged value implies that the hedge fund returns exhibit a trend such that a positive (negative) return will be followed by another positive (negative) return. Autocorrelation may result from difficulties in the periodic valuation of the investment holdings. For any illiquid positions such as private debt or direct lending to small and medium-sized companies, the fund manager might attempt to smooth the contemporaneous returns concerning the previous months' records so that an estimation of the market value might be represented by a positive first-order autocorrelation. These autocorrelations lead to an underestimation of the return volatilities and an overestimation of the Sharpe ratios. While the variance in the hedge fund returns often changes over time, the typical phenomena of volatility clustering imply the persistence of volatility so that the volatility shocks this month might influence the expectation of volatility many periods in the future. As this requires a non-linear specification between returns and volatility, some modeling efforts for estimating the level of returns (e.g., an ARMA model) and the volatility modeling (e.g., a GARCH-family) have been combined to capture the level of persistence in risk-return profiles.

While Hengl et al. (2007), Harris et al. (2010), and Meng (2014) elaborate the practical applicability of a hybrid data interpolation techniques by combining interpolation purely based on point observations and the correlation between response variables and auxiliary variables together with spatial autocorrelation within the response variable, this article extends the lead taken by Inoue, Shimizu, and Kigoshi (2010) and Selby, Kockelman, and Kara (2013) in two dimensions. Inoue et al. (2010) consider spatial distribution and temporal changes in the modeling Tokyo residential land prices by using spatio-temporal analysis with universal kriging. The parameters of the spatio-temporal covariance function are estimated separately after the initial hedonic land price regressions and use the residuals of land price model by weighted least squares criteria. The subsequent estimation of the parameters of generalized least squares was conducted by overlaying the estimated covariance structure. The stability of estimation might be proved once the parameters converge after the several iterations of the entire estimation of the land price model and the spatio-temporal covariance functions. Since the algorithms are based on asymptotic, the model performance in small data sets may be suspect. Selby et al. (2013) consider two algorithms of universal kriging and geographically weighted regression in their spatial prediction of traffic levels in Texas regions that can be defined as "local", because spatially neighboring traffic data are used to calculate the traffics in unmeasured locations.

Firstly, we broaden the scope of the paper by considering both nuisance and substantive spatial dependence in hedge fund returns. There is considerable evidence from behavioral finance that implied volatilities are clustering with asymmetrical in upside and downside potentials. Nuisance here refers to model residuals from the motivation to correct the effect of spatial dependence of hedge fund returns through the adjustments that incorporate the spatial autocorrelation in an error-term of the factor-risk models. Secondly, the time series of hedge fund returns are closely related to the risk-return profile of the market; thus, it is challenging but equally doable that the pricing model can be constructed based on this 'geographical' market return-risk framework. Since most information on monthly hedge fund returns is in the form of point, not an area data, however, in an attempt to observe the spatial dependences in return time series, we established a process for the creation of a market risk-return grid[1] of the Cartesian coordinates of (x_i, y_i). This process involves an intrinsically symmetric weights matrix $\boldsymbol{W}$ derived from the queen contiguity for the Thiessen polygon tessellations constructed on the observed point representations of the hedge fund returns on the Cartesian coordinates of Eastings and Northings, e.g., equity market index and the index of implied volatilities. Following our initial expectation, including explicit spatial terms of both endogenous and instrumental variable ($\boldsymbol{IV}$) to the right-hand side of the model specifications might allow the consistent estimation vs. non-spatial OLS approach. Any substantive autocorrelation in returns may cause model bias, which might be due to the spatial dependencies through lagging or from the error terms with all the included covariates. A typical inconsistency observed in OLS estimates is due to the multidirectional dependency in the spatial data, which makes the important distinction between spatial autocorrelation and autocorrelation observed from the financial time-series. Our analytical methodology attempts to overcome the difficulties of analyzing financial time series data by introducing nuisance and substantive spatial dependence parameters in the model specifications. This type of spatial autocorrelation might be considered as substantive in the form of the spatial dependence from the nature of spatial spillover with associated economic processes behind it. Meanwhile, the rows of the neighborhood matrix $\boldsymbol{W}$ sum to 1, which means that $\boldsymbol{W}$ is always row-standardized.

While the purpose of this study is to apply the exploratory spatial data analysis (ESDA) methodologies to the performance attribution analysis of hedge fund returns and subsequent application to the feasible a priori spatial detection of any clues of irregularity in its return generation process, comparative analysis with the widely adopted framework of performance measures such as those from Getmansky, Lo, and Makarov (2004) is temporarily outside the scope of this article. The paper proceeds as follows. Section 2 models to measure the impact of spatial nuisance and substantive characteristics. Section 3 presents the empiri-

1) This is an imaginary Cartesian coordinates of (x_i, y_i) with the natural logarithm of the S&P 500 index as x-coordinate similar to Eastings and the natural logarithm of the implied volatility (VIX) index as y-coordinate similar to Northings to see whether the various hedge fund returns are systematically related to their own returns in the adjacent distances in this Cartesian equity market-implied volatility map as shown in the second and third panels of Exhibit 1 and 2 in Section 3.

cal methodology in our spatial lag and error models applied to the Barclays Hedge Fund Indices. In Section 4, discusses an *ad hoc* case experimentation to a recently failed credit hedge fund for possible qualitative application of the irregularity detection through spatial specifications. Section 5 concludes.

[2] Spatial Lag and Spatial Error Models

A large number of OLS diagnostics assume normal error distributions. Because it is hard to assess the extent to which this may be violated with the unobservable errors, the regression residuals are adopted for testing the degree of non-normal errors. For instance, a low probability of the Jarque-Bera test indicates a rejection of the null hypothesis of error normality. However, the subsequent tests for heteroscedasticity and spatial dependence should be interpreted with rigors, as they are also subject to the same normality principle. OLS can be used for modeling hedge fund returns as the model is represented using matrix notation in Eq. 1.1 and the estimated β coefficients in Eq. 1.2.

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I) \qquad \text{(Eq. 1.1)}$$

$$\beta = (X'X)^{-1}X'y \qquad \text{(Eq. 1.2)}$$

where y is an (N×1) vector of observed monthly log-returns of hedge fund index; X is an (N×K) vector of factor-risk premiums; β is a (K×1) vector of estimated coefficients, and ε is an (N×1) assumed-to-be independent error vector as the net effect of all the other factors affecting hedge fund returns but omitted from this classic specification. An OLS-based factor-risk model is served as the benchmark against which the subsequent spatial econometric models to be evaluated.

Spatial dependence is the situation where the hedge fund returns (or the error terms) at each location in the Cartesian coordinates of the market index and the implied volatilities (x_i, y_i) is correlated with the returns (or values for the error terms) at other locations. Global spatial models address spatial dependence or autocorrelation in the spatial processes. For example, Anselin (1988) assumes spatial autocorrelation is in either the response variable or the error terms, and the corresponding models are usually calibrated by maximum likelihood (ML) rather than the OLS technique as some OLS assumptions (e.g. independently and identically distributed residuals) are violated. One potential intuition is that the financial market participants could anticipate similar returns in a previously observed neighborhood of the market risk-return framework with the reference of determining contemporaneous risky asset returns due to the uncertainties in the neighborhood characteristics of risk premiums. Thus, it might be the expectation of the "déjà vu" value of returns near to the particular Cartesian locations of the levels of the market and implied volatilities.

The presence of spatial autocorrelation in the time series of hedge fund returns can be determined by estimating the global Moran's I test for residuals as in Eq. 2, where ε denotes a residual vector, W is an exogenous spatial weights matrix, N is the number of observations, and S is a standardized factor defined as the sum of all elements in the given weights matrix. If the statistic is significant then the null hypothesis of no spatial autocorrelation is rejected and the return process of hedge funds imposes spatial autocorrelation.

$$I = [N/S]\{[\varepsilon' W \varepsilon]/\varepsilon'\varepsilon\} \quad \text{(Eq. 2)}$$

The test statistics of spatial autocorrelation is based on independence assumptions, which may result in biased inferences: Under the influences of spatial errors, OLS estimates become inefficient. Under the spatial autoregressive effects, OLS estimates are biased and thus result in incorrect inferences. The random regression error process might be heteroscedastic when it does not have a constant variance over all observations which results in unbiased but inefficient OLS estimates. Any inference based on the usual t- and F-statistics will be misleading as well, and the R^2 measure of the goodness-of-fit will be spurious. In the analysis of hedge fund return attributions, this might be observed when there are systematic regional differences in the relationships with volatilities (i.e., spatial regimes of risk-on and off), or when there is a continuous spatial drift in the parameters (i.e., spatial expansion and clustering). The presence of any of these spatial effects would make an OLS misspecified. Whenever both heteroscedasticity and spatial dependence may be present in the data process, the heteroscedasticity tests tend to be highly sensitive to the presence of spatial dependence (Anselin and Rey, 2014). Therefore, any test indicative of heteroscedasticity may also come from the presence of spatial dependence (or the other way around).

When spatial autocorrelation such that the spatial lag term contains the effects to the fund returns from the previously watched neighboring observations, which in turn may contain the spatial lag for their neighbors' neighbors, leading to the issue of simultaneity. This simultaneity mostly results in a non-zero correlation between the spatial lag and the error term, which violates an assumption of independence between them. Then, the OLS estimation will be inconsistent and biased due to the dependence of error term, and all the included covariates and any subsequent inference will be flawed again. Instead of OLS, specialized estimation methods that properly accounting for the spatial simultaneity in the model are necessary. These methods are either based on the generalized method of moment principle or the application of instrumental variable (IV) estimations in a Spatial Two-Stage Least Squares (S2SLS) or both.

The spatial error model (SEM) is designed to capture the influence of unmeasured exogenous risk factor premiums since the spatial clustering of the hedge fund returns is manifested through the error term ε, thus not explained by the measured covariates of explanatory risk premiums. SEM contains the a priori pivotal spatial variable in the right-hand side (RHS) of the model with subtle spatial features to the regression residuals and the error may follow a spatial autoregressive specification as in Eq. 3.1.

$$\boldsymbol{E}(\boldsymbol{y}) = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \boldsymbol{u} = \lambda \boldsymbol{Wu} + \varepsilon, \varepsilon \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \qquad \text{(Eq. 3.1)}$$

$$\boldsymbol{E}[\varepsilon\varepsilon'] = \sigma^2 [(\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}((\boldsymbol{I} - \lambda\boldsymbol{W})^{-1})'] \qquad \text{(Eq. 3.2)}$$

where $\boldsymbol{y}$ is an (N×1) vector of observations on the hedge fund returns; $\boldsymbol{X}$ is an (N×K) matrix of observations on the explanatory risk factor premiums; $\boldsymbol{\beta}$ is a (K×1) vector of regression coefficients; $\boldsymbol{u}$ in an (N×1) vector of spatially autocorrelated error terms; $\boldsymbol{W}$ is the weights matrix; $\boldsymbol{Wu}$ is a spatial lag for the errors; λ is the spatial autoregressive coefficient; and ε as a vector of idiosyncratic errors. Then, $\lambda\boldsymbol{Wu}$ captures the spatial autocorrelation between the neighboring error terms $\boldsymbol{u}$. Since an ideal spatial error model implies no distinctive effects of the neighboring dependent variable, the idiosyncratic errors should be heteroscedastic $\boldsymbol{E}[\varepsilon_i^2] = \sigma_i^2$, but uncorrelated $\boldsymbol{E}[\varepsilon_i \varepsilon_t] = \boldsymbol{0}$. The observed spatial clustering in the hedge fund returns might be accounted for by the spatial patterning of measured and unmeasured exogenous factor risk premiums. Eq. 3.2 shows the variance-covariance matrix which tends to be heteroscedastic as well. Therefore, the expected hedge fund returns $\boldsymbol{E}(\boldsymbol{y})$ for each Cartesian coordinates $(\boldsymbol{x}_i, \boldsymbol{y}_t)$ are affected by the stochastic errors at all other nearby returns through the spatial multiplier $(\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}$. The value of $|\lambda|$ implies the magnitude of spatial multiplier effects.

By incorporating the influence of unmeasured exogenous risk factor premiums for stipulating an additional effect of neighboring attribute values, the spatial lag model (SLM) in Eq. 4 accounts for the potentially confounding effect of spatial autocorrelation in the hedge fund returns. Then insignificant spatial dependence would remain in the residuals of our factor risk premium models. As Anselin and Rey (2014) proposed a decision-making process for selecting an appropriate analysis of spatial dependence, the Lagrange Multiplier (LM) test for spatial error autocorrelation in the SLM is a useful diagnostic for this.

$$\boldsymbol{E}(\boldsymbol{y}) = \rho\boldsymbol{Wy} + \boldsymbol{X\beta} + \boldsymbol{u} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{X\beta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{u}, \quad \boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}) \qquad \text{(Eq. 4)}$$

where $\boldsymbol{W}$ as an (N×N) exogenous weights matrix that specifies the assumed spatial structure or connections between the observed fund returns, and $\boldsymbol{Wy}$ is an (N×1) vector of neighboring hedge fund returns accounting for spatial dependencies. The scalar parameter ρ is a spatial autoregressive coefficient indicating the effect of the hedge fund returns in the neighboring previously observed hedge fund returns in the Cartesian coordinates of the market and implied volatilities $(\boldsymbol{x}_i, \boldsymbol{y}_t)$. $\boldsymbol{X\beta}$ is an (N×K) matrix of observations on the exogenous explanatory risk factor premiums multiplied by a (K×1) vector of regression coefficients $\boldsymbol{\beta}$ for each $\boldsymbol{X}$, and $\boldsymbol{u}$ is an (N×1) vector of normally distributed random error and might be decomposable into a spatially lagged term ρWy of the weighted average of neighboring values y and an idiosyncratic error term ε. The strength of the spatial dependence among the historical hedge fund returns ρWy is correlated with the expectations of contemporaneous hedge fund returns $E(y)$.

While the inclusion of Wy allows assessing the significance of the residual non-spatial variables, the presence of a spatial lag term is mathematically equivalent to the existence of endogenous variables on the RHS in simultaneous equation systems. As these two $(I-\rho W)^{-1}$ terms in Eq. 4 are called spatial multipliers (Anselin and Rey, 2014), the expected monthly returns of a hedge fund at one location (x_i, y_i) depend on the value of all the nearby Cartesian coordinates (x_l, y_l) in this model. Cartesian coordinates here mean a polygon structure as we have constructed asymmetric Thiessen polygon tessellations from the point representations of the historical hedge fund returns and derived a first-order queen contiguity weights matrix W from the center of these polygons. The rows of the neighborhood weights matrix W are row-standardized and always sum to 1. Since β as the vector of marginal implicit sensitivities to hedge fund returns in a traditional factor risk model, $(I-\rho W)^{-1}\beta$ implies that the spatial marginal sensitivities to risk premiums will be smaller than traditional performance attribution analysis when $|\rho| < 1$. Therefore, the expected marginal effect on hedge fund returns consists of both a direct effect due to the change in the contemporaneous risk premiums as well as the induced effects due to marginal changes related to neighboring historical fund returns. These induced effects imply that any serious expectation of contemporaneous hedge fund returns is restricted primarily from the historical observations experienced by the neighboring returns at the nearby Cartesian coordinates (x_l, y_l).

By substituting the reduced form $u = (I-\lambda W)^{-1}\varepsilon$ into the SLM results in a spatial lag component with an error variance-covariance $(I-\lambda W)^{-1}\Sigma(I-\lambda W')^{-1}$.

$$y = \rho Wy + X\beta + (I-\lambda W)^{-1}\varepsilon \quad \text{(Eq. 5.1)}$$

The main complication is the existence of endogenous spatially lagged hedge fund returns Wy on the RHS of the equation. After rearranging the terms as in Eq. 5.2, the first and second-order spatially lagged hedge fund returns show up on the RHS together with a non-spatially correlated error term as in Eq. 5.3.

$$(I-\lambda W)(I-\rho W)y = (I-\lambda W)X\beta + \varepsilon \quad \text{(Eq. 5.2)}$$

$$y = (\lambda+\rho)Wy - \lambda\rho W^2 y + X\beta - \lambda WX\beta + \varepsilon \quad \text{(Eq. 5.3)}$$

While the presence of $\lambda+\rho$ and $\lambda\rho$ terms in the same equation would make these coefficients unidentifiable, additional information in $\lambda WX\beta$ and by dividing the estimate for $\lambda\beta$ by the matching value of estimated $\hat{\beta}$, a non-unique estimate for λ and ρ might be obtained[2].

2) Anselin and Rey (2014) explain that as the constant term provides an estimate for the product $(I-\lambda)\beta_0$, the estimated

With the generalized S2SLS estimator for $\hat{\rho}$ and the Generalized Method of Moments (GMM) estimator for $\hat{\lambda}$ and pre-multiplying both sides of the spatial filter expression by the inverse matrix $(I-\rho W)^{-1}$ yields Eq. 6.

$$y = (I-\rho W)^{-1}X\beta + (I-\rho W)^{-1}(I-\lambda W)^{-1}\varepsilon \quad \text{(Eq. 6)}$$

Since $E[\varepsilon]=0$, the variance-covariance matrix for the error term no longer imposes a complexity. Eq. 5.1 might be generalized[3] with an additional set of endogenous explanatory variables Y with associated coefficient vector γ:

$$y = \rho W y + X\beta + Y\gamma + (I-\lambda W)^{-1}\varepsilon \quad \text{(Eq. 7)}$$

[3] Application to the Barclays Hedge Fund Index

Due to the parameter uncertainty, the usual factor risk models do not immediately applicable to the analysis of single hedge funds. The stability of parameters also depends on the length of single managers' track records. The choice of the proxies for average hedge fund returns in this study is one of the practitioners' standard Barclays Hedge Fund Index (BHFI)[4] and its eighteen sub-indices. We first divide eighteen sub-indices into equity- and arbitrage-focused strategies to examine whether the investment style differences are related to the magnitude of spatial dependencies in hedge fund performance. While equity strategies such as equity long-short (ELS), equity long bias (ELB), Pacific Rim equity (PREQ), and technology equities (Tech) are screened, arbitrage strategies such as event-driven (AED), convertible arbitrage (ACNV), fixed income (AFI), distressed securities (Dist), multi-strategies (MS), BTOP50, and the Barclays CTA (CTA) are screened out of the sub-index pools.

Global Moran's I test was conducted first to verify whether there was a spatial dependence as presented in Exhibit 1 and 2 using monthly hedge fund returns data from January 2003 to December 2018[5]. The first panels in

$\widehat{\beta_0}$ can be obtained by means of an estimated $\hat{\lambda}$. In addition, the same spatial weights matrix W in the lag and error specifications in actual calibration was adopted since assuming two different spatial weights for the lag and error processes in the same hedge fund returns doesn't make a lot of sense.

3) Kelejian and Prucha (1998) explains how a consistent estimator for the spatial autoregressive coefficient ρ and the parameter λ can be obtained by 2SLS specification. Their algorithms were further generalized to cover the heteroscedasticity in the error term by Kelejian and Prucha (2010) and Drukker et al. (2013).

4) "Certain biases do inflate performance while others may skew index performance downwards. The Barclays Hedge Fund Indices may only be able to provide some biased snapshot of the 'true' hedge fund universe, they are not including defunct funds, thus do not account for the survivorship bias. Furthermore, these indices didn't explicitly account for the backfill and the incubation biases that arise due to the voluntary nature of self-reporting in hedge fund databases." Cho, J.K. and Kim, G.W. "Making a Bet at a Right Time: Style and Volatility Timing Abilities of Korean Equity Hedge Funds." Asset Management Review, Vol. 5, Issue 2, December 2017

Exhibit 1. Moran's *I*, LISA Cluster & Significance Maps, and 3D Return Plots – Equity Strategies

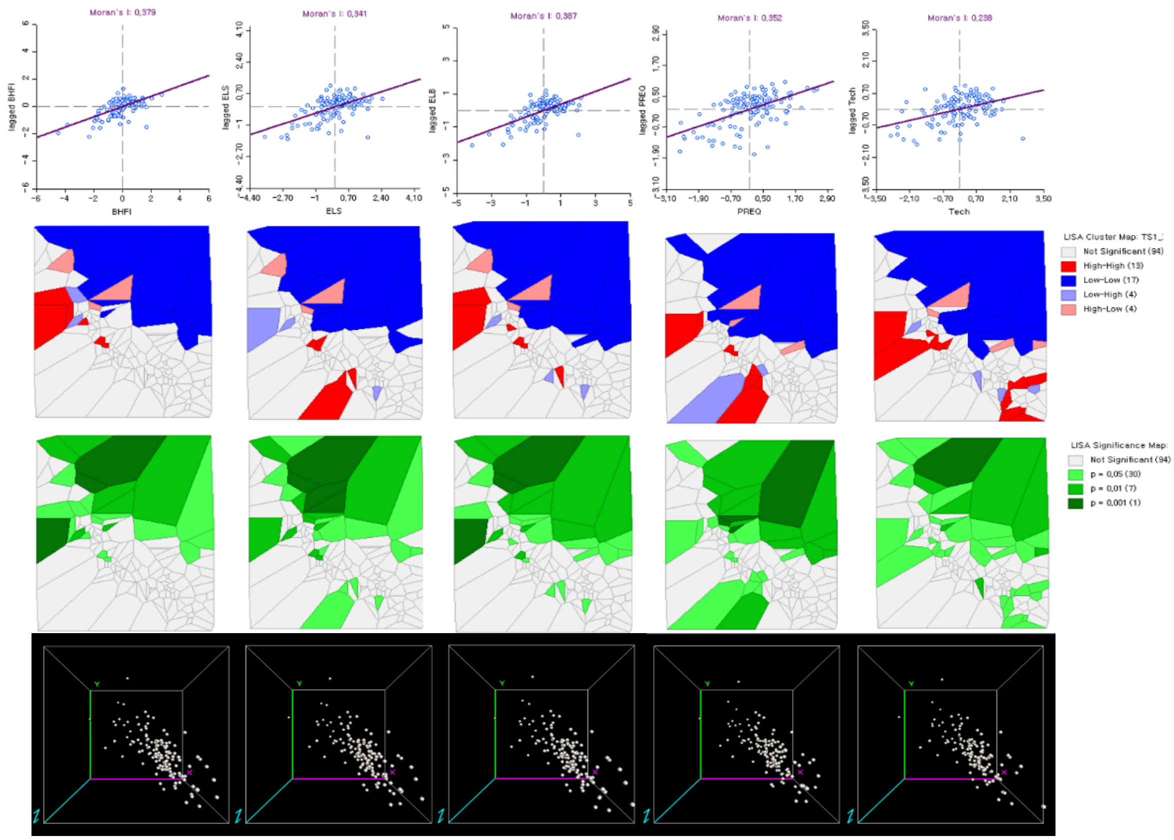
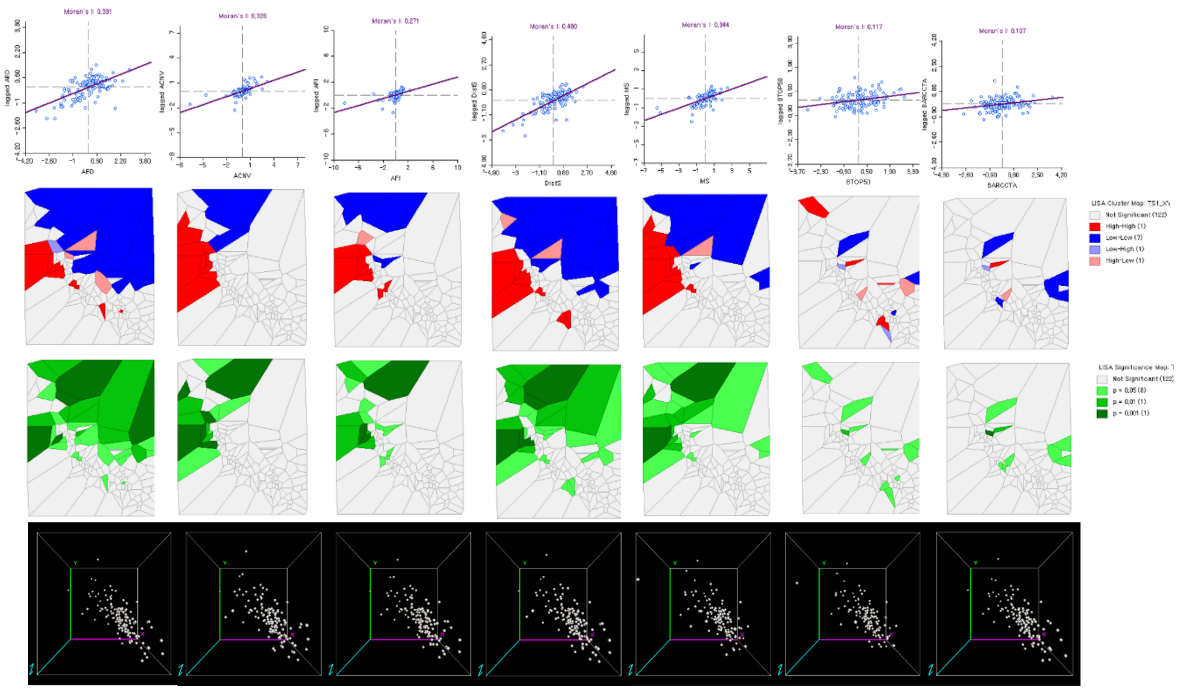


Exhibit 1 and 2 show the Moran Scatter Plot together with the calculated Moran's *I* values. Moran Scatter Plot shows the horizontal axis in the normalized contemporaneous and neighboring hedge fund returns. The first and third quadrants represent areas of values with positive correlations (both high-high and low-low) and the remaining quadrants represent areas in negative correlation. For example, the monthly returns of the BHFI index show Moran's *I* of 0.379 ($p < 0.01$) between the contemporaneous index returns and the neighboring (thus spatially lagged) returns. Moran's *I* statistic was further estimated to 0.341 (ELS), 0.387 (ELB), 0.352 (PREQ), 0.238 (Tech), 0.381 (AED), 0.325 (ACNV), 0.271 (AFI), 0.48 (Dist), 0.344 (MS), 0.117 (BTOP50), and 0.107 (CTA), which suggest the presence of spatial autocorrelation in the monthly equity- and arbitrage-focused strategy returns.

The Local Indicators of Spatial Association (LISA) significance maps in the third panels

5) The data coverage encompasses the periods of sporadic market stress triggered by the U.S. sub-prime mortgage crisis, the fiscal crisis in the periphery Eurozone, the Chinese deceleration in the summer of 2015, unfounded fears of recession in the U.S. in January 2016, the Brexit referendum in June 2016, and the implosion of the volatility market in February 2018 until the sudden recessionary concerns at the end of 2018, offering relatively long data samples for a robust analysis of hedge fund returns.

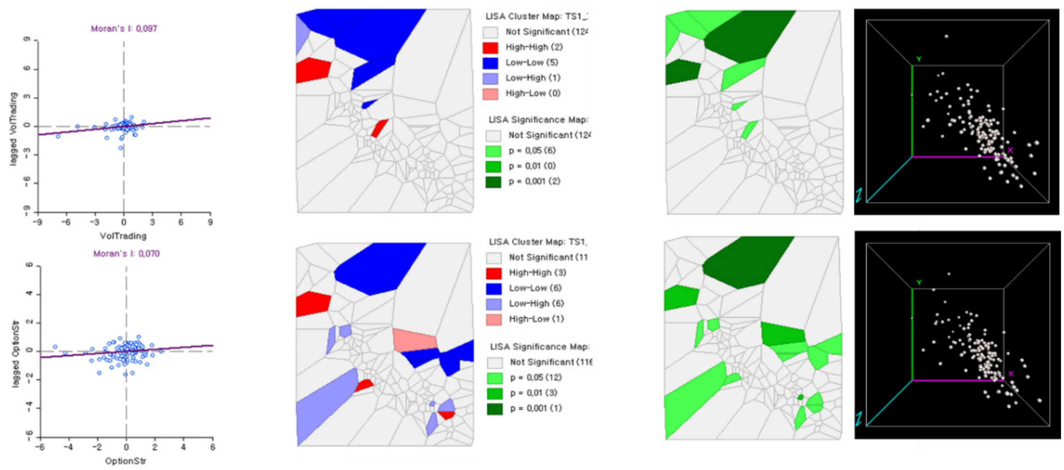
Exhibit 2. Moran's I, LISA Cluster & Significance Maps, and 3D Return Plots – Arbitrage Strategies



of Exhibit 1 and 2 show the locations of polygon tessellations with a significant local statistic. The spatially heterogeneous degree of significance is reflected in the increasingly darker shades of green. The map starts with $p < 0.05$ and shows all the categories of significance that are meaningful for the given number of permutations. Since there were 999 permutations in Exhibit 1 and 2, the smallest p-value is 0.001, with the darkest shade of green locations.

The second panels of Exhibit 1 and 2 show the cluster map of selected equity- and arbitrage-focused strategies which augments the significant locations with an indication of the type of spatial association, based on the location of the value and its spatial lag in the Moran scatter plot in four categories; dark red for the high-high clusters, dark blue for the low-low clusters, light blue for the low-high spatial outliers, and light red for the high-low spatial outliers. For instance, the ACNV index in Exhibit 2 shows positive high-high (red) and low-low (blue) areas in both lower and upper quadrants at the second panel between the reverse 45-degree line. The represented red and blue area of ACNV is comparable to those of BHFI where the area of red (high-high) is smaller, while the area of blue is much larger in BHFI. While the positive correlation of high-high area implies the locations with high hedge fund returns at a high level of similarity with its neighboring observed historical returns (Hot Spot), the low-low area implies the locations with low hedge fund returns at a high level of similarity with its neighboring historical returns (Cold Spot). Therefore, for the hedge fund indices such as BHFI, ELB, Tech,

Exhibit 3. Moran's *I*, LISA Cluster & Significance Maps, and 3D Return Plots – Vol & Option Traders



AED, ACNV, AFI, Dist, and MS, the Hot Spot can be expected when the equity market index level is relatively low and the implied volatility index is relatively high.

The second (fourth) quadrant represents the locations of spatial outliers with high (low) hedge fund returns and low (high) level of similarity with its surroundings thus high-low (low-high). CTA distinguishes itself from all other strategies as it is known as trend-following, long-volatility trades. Moran's *I* statistics are relatively lower than others for both BTOP50 and CTA at 0.117 and 0.107, respectively.

Even though we initially classified the strategies of Barclays Hedge Funds into equity- and arbitrage-focused, we would doubt the practical benefits out of the portfolios of hedge fund strategies since most of the equity and arbitrage-focused strategies show very similar LISA clustering patterns. In that sense, BTOP50 and CTA might be more relevant in the argument in favor of hedge fund portfolio diversifications. The last panels of Exhibit 1 and 2 show 3-dimensional scatter plots with the monthly log-returns of the S&P 500 index in the X-axis, the monthly log-returns of VIX in the Y-axis, and the monthly hedge fund returns in the Z-axis. Moran's *I* statistic does not provide a meaningful implication in portfolio diversification as we can see in Exhibit 3 of Barclays Volatility and Option Trading strategies. While the Moran's *I* statistic is quite low at 0.097 for Vol Trading and 0.070 for Options Trading indices, the major distribution of the Hot and the Cold Spots are quite similar to each other but with those of equity and arbitrage strategies with fewer implications for these short volatility trades to the potential diversification benefits.

Global factor definitions are consistent with a global market for risk, where hedge funds operate. With Fama and French [2012]'s six global equity risk factors[6] such as market excess returns (xMKT), size (SMB), value (HML), quality (RMW), conservativeness (CMA), and

6) http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Exhibit 4. Results of Estimating Equity Hedge Fund Returns by OLS and Spatial Error & Lag Models

	Barc Hedge Indx (BHFI)		Equity Long-Short (ELS)		Equity Long-Bias (ELB)		Pacific Rim Ety (PREQ)			Technology Eqty (Tech)	
Variable	OLS	SLM	OLS	SEM	OLS	SLM	OLS	SLM	SEM	OLS	SEM
Intercept	0.0008	0.0007	0.0017***	0.0019**	0.0007	0.0007	0.0014	0.0011	0.0016	0.0043***	0.0045***
xMKT	0.3341***	0.3217***	0.2611***	0.2521***	0.5508***	0.5299***	0.3229***	0.2979***	0.3112***	0.3902***	0.3902***
SMB	0.2068***	0.1998***	0.1356***	0.1197**	0.2765***	0.2663***	0.1859**	0.1696**	0.169**	0.1196	0.1044
HML	0.0542	0.0507	0.0000	0.0055	-0.0229	-0.0268	0.0528	0.0649	0.1056	-0.2594***	-0.2450***
RMW	-0.0190	-0.0193	-0.1395**	-0.1575***	-0.1895**	-0.1884**	-0.1651	-0.1524	-0.139	-0.2681**	-0.2819**
CMA	-0.2644***	-0.2614***	-0.1638***	-0.2095***	-0.2125***	-0.2118***	-0.1285	-0.1342	-0.1739*	-0.3533***	-0.4061***
WML	0.0417**	0.0398**	0.0759***	0.0728***	0.0524**	0.0470**	0.0709*	0.0695**	0.0859**	0.1377***	0.1414***
PTFSBD	-0.0067*	-0.0051	-0.0059	-0.0060*	-0.0092*	-0.0067	-0.0084	-0.0043	-0.0066	-0.0157*	-0.0159**
PTFSFX	0.0048	0.0045	0.0055	0.0069**	0.0082*	0.0079**	0.0038	0.0032	0.0047	0.0192***	0.0182***
PTFSCOM	-0.0083**	-0.0077**	-0.0091**	-0.0086**	-0.0101*	-0.0098**	-0.0056	-0.0053	-0.0067	-0.0091	-0.0088
PTFSIR	-0.0026	-0.0024	0.0012	0.0019	-0.0012	-0.0004	0.0013	0.0016	0.0007	0.0109*	0.0122**
PTFSSTK	-0.0009	0.001	0.0042	0.0042	0.0019	0.0048	-0.0012	0.0087*	-0.0003	0.003	0.0057
Lambda (λ)				0.4050***					0.3817***		0.3028**
Rho (ρ)		0.1207**				0.1274***		0.2578***			
Moran's I	0.079*		0.149***		0.071*		0.1719***			0.1218***	
Adj-R^2	0.92	0.93	0.84	0.867	0.932	0.941	0.659	0.713	0.718	0.728	0.764
Log-Like	506.2	509.14	497.32	501.71	460.66	464.69	403.05	407.65	407.71	401.06	403.61
AIC	-988.4	-992.29	-970.63	-979.42	-897.32	-903.38	-782.1	-789.3	-791.41	-778.11	-783.22
JB (d.f. 2)	11.17***		1.4491		16.63***		16.34***			1.76	
B.P. (d.f. 11)	36.93***	44.80***	39.73***	22.16**	22.69**	27.44***	37.61***	29.90***	31.98***	7.43	5.17
K.B. (d.f. 11)	21.74**		32.35***		13.65		22.79**			9.88	
L.M.-Error			7.92***				10.49***			5.27**	
L.M.-Lag	6.12**				8.63***		10.71***				
L.R. (d.f. 1)		5.89**		8.78***		8.06***		9.20***	9.31***		5.107**

*Note: * p < 0.10, ** p < 0.05, *** p < 0.01*

the momentum (WML) premiums, five primitive trend-following strategies (PTFS) look-back straddle returns of bond (PTFSBD); currency (PTFSFX); commodity (PTFSCOM); short-term interest rate futures (PTFSIR); and stock index (PTFSSTK) proposed by Fung and Hsieh [2001][7] were incorporated together with the VIX[8] for its endogeneity character. As some of the factors didn't come up with the explicitly definable returns, not all of the risk factors offer valid return premiums. Meanwhile, identifying the significance of factor-conditioned alpha based on various risk factor models per hedge fund style is not the major focus of this paper.

Evaluating the relative performance of models begins with a comparison of parameter estimates for the OLS and spatial autoregressive models. Based on these empirical backgrounds, we now demonstrate the performance of equity hedge funds by OLS, SLM, and SEM in Exhibit 4 and the performance of arbitrage hedge funds in Exhibit 5 with the regression of excess monthly hedge fund returns against eleven risk factor premiums elaborated at Eq. 1.1 (OLS), Eq. 3.1 (SEM), and Eq. 4 (SLM).

To quantify the degree of linkage between the neighboring polygon tessellations of historical returns, the neighborhood structure is to be converted by a weight matrix chosen. In our case, the results are reported using the row standardized way with the implication of allocating more weights to the residuals with relatively fewer neighboring observations of

7) http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-Fac.xls.
8) VIX White Paper, www.cboe.com/micro/vix/vixwhite.pdf

Exhibit 5. Results of Estimating Arbitrage Hedge Fund Returns by OLS and Spatial Error & Lag Models

	Event-Driven Arb (AED)		Convertible Arb (ACNV)		Fixed Income Arb (AFI)		Distressed (DistS)		Multi-Strategies (MS)		BTOP50		Barclays CTA	
Variable	OLS	SLM	OLS	SLM	OLS	SLM	OLS	SLM	OLS	SLM	OLS	SLM	OLS	SLM
Intercept	0.0022***	0.0019**	0.002	0.0007	0.0008	0	0.0007	0.0006	0.0012	0.0006	0.0000	0.0000	0.0007	0.0006
xMKT	0.2707***	0.2484***	0.1485***	0.1069***	0.099***	0.0657**	0.2468***	0.1692***	0.1687***	0.1453***	0.1326***	0.1341***	0.1334***	0.1341***
SMB	0.2477***	0.2400***	0.2833***	0.2148***	0.1792**	0.1573**	0.2215**	0.1826**	0.2199***	0.1938***	-0.0982	-0.1048	-0.0027	-0.006
HML	0.0702	0.0657	0.1698	0.1550*	0.1345	0.097	0.2732**	0.2257**	0.0462	0.0305	-0.0555	-0.0447	-0.0104	-0.0015
RMW	-0.0551	-0.0582	0.0863	0.034	0.08	0.0647	-0.099	-0.1246	0.0500	0.0329	0.3281*	0.3147**	0.2134*	0.2044*
CMA	-0.2247***	-0.2214***	0.4650***	-0.4451***	-0.2541**	-0.2545**	-0.3271**	-0.3244***	-0.2737***	-0.2728***	0.1168	0.0754	0.0168	-0.0154
WML	0.0192	0.0173	-0.059	-0.0399	0.0336	0.0198	0.0718*	0.0387	0.0239	0.0239	0.1166**	0.1164**	0.0933**	0.0960***
PTFSBD	-0.0152**	-0.0136**	0.0033	0.0079	-0.0006	0.0013	-0.0207**	-0.0166**	-0.0035	-0.0011	0.0172	0.0154	0.0125	0.011
PTFSFX	0.004	0.0037	-0.0103	-0.0115*	-0.0106	-0.0111*	0.0026	0.0015	0.0027	0.0019	0.0208**	0.0202**	0.0233***	0.0229***
PTFSCOM	-0.0109*	-0.0098*	-0.0043	0.0008	-0.0005	0.0013	0.0150*	-0.0127*	-0.0033	-0.0014	0.0110	0.0119	0.0170**	0.0186**
PTFSIR	0.0017	0.0018	-0.0150**	-0.0177***	-0.0235***	-0.0232***	-0.0115*	-0.0095*	-0.0092**	-0.0092***	0.0060	0.0064	0.0083	0.0077
PTFSSTK	0.0056	0.0087*	-0.004	0.0008	0.0013	0.0021	0.003	0.0068	-0.0022	0.0012	0.0007	0.0029	-0.0008	0.002
Lambda (λ)														
Rho (ρ)		0.2132***		0.4702***		0.3567***		0.4297***		0.3346***		0.3086***		0.3577***
Moran's I	0.1194**		0.190***		0.109***		0.13***		0.159***		0.08*		0.119***	
Adj-R^2	0.779	0.811	0.583	0.689	0.485	0.574	0.631	0.731	0.693	0.752	0.207	0.314	0.289	0.4
Log-Like	449.65	453.68	384.94	395.34	402.09	407.12	384.03	396.65	456.28	463	359.42	362	399.61	403.34
AIC	-875.3	-881.35	-745.87	-764.69	-780.17	-788.25	-744.06	-767.3	-888.56	-899.99	-694.83	-660.54	-775.21	-780.68
JB (d.f. 2)	2.39		167.67***		394.64***		52.33***		97.62***		7.62**		2.8	
B.P. (d.f. 11)	21.59**	30.26***	187.24***	208.04***	307.61***	361.54***	27.43***	20.99**	131.92***	154.55***	6.45	6.16	6.16	6.28
K.B. (d.f. 11)	16.45		53.87***		63.04***		11.52		43.41***		4.71		4.86	
L.M.-Error														
L.M.-Lag	8.89***		19.88***		9.81***		27.24***		13.44***		4.65**		6.77***	
L.R. (d.f. 1)		8.06***		20.82***		10.08***		25.23***		13.43***		5.19**		7.47***

*Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

historical hedge fund returns. We follow the normality approach in estimating Moran's I in Exhibit 4 and 5. Except for BHFI, ELB, and BTOP50 indices, the test statistics of 9 hedge fund indices show that the null hypothesis of no spatial effects has to be rejected at $p <$ 0.05. While these results are qualitatively insensitive to the neighboring structure and weights style (Anselin and Rey, 2014), Moran's I statistics of regression residuals are calculated as 0.149 (ELS), 0.172 (PREQ), 0.119 (AED), 0.19 (ACNV), 0.11 (AFI), 0.13 (Dist), 0.159 (MS), and 0.119 (CTA) and are significant at $p <$ 0.01. In other words, Moran's I reject the null hypothesis of no spatial autocorrelation between the regression residuals of the OLS model. It suggested that a conventional OLS model might not be appropriate; therefore, alternative modeling should be considered.

Since the existence of spatial autocorrelation in the monthly hedge fund returns and ε may justify the use of spatial autoregressive models to control for any potential estimation bias in the OLS coefficients, the significance of the lag parameters ρ and λ in the SLM and SEM models is indicative of the extent that the spatial autocorrelation effect in the hedge fund returns and the errors ε have been controlled for. An examination of the coefficients is supposed to indicate whether the use of the spatial lag term $\rho W y$ in the SLM is effective in reducing the bias in the OLS coefficients. Consequently, $(I - \rho W)^{-1} X\beta$ term from Eq. 4 suggests that the estimated significant coefficients in the SLM (SEM) should be smaller in their magnitude when compared to the OLS counterparts when $|\rho| < 1$. In the presence of strong spatial autocorrelation, likely, the spatial regression model will significantly outperform the OLS counterpart, which is the case as shown in Exhibit 4 and 5.

When the Moran's I test rejects the null hypothesis of no spatial dependence, the Lagrange Multiplier (LM) tests for spatial effects mostly suggest in favor of SLM rather than SEM, which implies the major effect of the misspecification may pertain to spatial autocorrelation rather than heteroscedasticity. While the LM-Lag statistic tests the null hypothesis of no spatial autocorrelation in the hedge fund returns, the LM-Error statistic tests the null hypothesis of no spatial error autocorrelation. In Exhibit 4 and 5, the LM-Lag test statistic was found to be significant at a 5% or above significance level with BHFI, ELB, AED, ACNV, AFI, Dist, MS, BTOP50, and CTA indices. The LM-Error test statistic was found to be signifi cant at a 5% or above significance level with ELS and Tech. For PREQ, both test statistics of LM-Error and LM-Lag were equally significant. The LM-test is asymptotically distributed as $\chi^2(1)$ and indicates the existence of spatial autocorrelation in the residuals in most of the equity- and arbitrage-focused hedge fund strategies. Since the spatial lag parameters are positive and significant, the result from the SLM indicates that spatial autocorrelation in hedge fund returns exists for those historical returns located very close to each other. The persistence of volatility might be a feasible description since the volatility shocks in the previous month might influence the expectations of volatility in many periods thereafter. Both spatial models perform better than the OLS counterpart in the improvement of goodness-of-fit (R^2), higher log-likelihood statistic (Log-Like), and the lower AIC measures.

The studentized Breusch-Pagan (B.P.) tests for OLS reported in Exhibit 4 and 5 show that the present of heteroscedasticity: The values for the B.P. test are 36.93 (BHFI), 39.73 (ELS), 22.69 (ELB), 37.61 (PREQ), 21.59 (AED), 187.24 (ACNV), 307.61 (AFI), 27.43 (Dist), and 131.92 (MS), while the values for Koenker-Bassett (K.B.) test are 21.74 (BHFI), 32.35 (ELS), 22.79 (PREQ), 53.87 (ACNV), 63.04 (AFI), and 43.41 (MS). These are highly significant for a χ^2 variate with 11 degrees of freedom, strongly indicating the presence of heteroscedasticity. This may be problematic because the derivation of the maximum likelihood (ML) estimate assumes a constant error variance, which suggests the spatial two-stage least squares (S2SLS) model might be more appropriate. No evidence of significance in both statistics was detected at Tech, BTOP50, and CTA indices. Furthermore, the large difference between B.P. and K.B. test statistic also confirms potential error non-normality, since, under the Gaussian, the value for both should roughly be the same (Anselin and Rey, 2014).

The ones in the bottom are the regression coefficients coming with asymptotic standard error from the analytical computation of the asymptotic variance and an asymptotic t-test with its p-value. We also get the likelihood ratio (L.R.) test on the spatial parameter. This is a test on the null hypothesis that (ρ or λ) = 0 and the values of 5.89 (BHFI), 8.78 (ELS), 8.06 (ELB), 9.20/9.31 (PREQ), 5.11 (Tech), 8.89 (AED), 20.82 (ACNV), 10.08 (AFI), 25.23 (Dist), 13.43 (MS), 5.19 (BTOP50), and 7.47 (CTA) are highly significant for a χ^2 variate with 11 degrees of freedom, strongly indicating the presence of remaining spatial dependence. Since this test is based on the ML estimation results, it may also be susceptible to the presence of heteroscedasticity.

When compared to the coefficients of the explanatory variables of the OLS model, the absolute values of the equivalents of the SLM and SEM were on the decline. Rho (ρ) as a spatial multiplier reveals the indirect or external effects of spatial interaction among the monthly hedge fund returns on each other. This could be described as the spatial spillover suggesting that, for instance, the monthly ACNV returns were positively influenced by those of the adjacent previous months' returns at an average of 47.02%, when other covariates were controlled for. An average of a 0.215% increase in the strategy returns would be observed if the size premiums were increased by 1%, all else being equal, which is smaller than 0.283% estimated by OLS. The straddle returns of short-term interest rate futures (PTFSIR) had a negative impact (-0.018%) on the returns of ACNV at a 5% significance level. This suggests that the premium of PTFSIR has a diminishing marginal impact on the ACNV returns despite the claims of these convertible arbitrage managers for their portfolio hedging practice to the fluctuations of short-term interest rates. Conservativeness premiums (CMA) had also a negative effect on the strategy returns, suggesting that a strategy returns would be lower by 0.445% every month if the conservativeness premiums were increased by 1%, ceteris paribus. Likewise, the monthly AED returns were positively influenced by those of the adjacent months' returns at an average of 21.32% when other factor risk premiums were controlled. An average of a 0.24% increase in the strategy returns would be observed if the size premiums were increased by 1%, all else being equal. Again, the CMA premium could harm the strategy returns, suggesting that strategy returns would be lower by 0.221% monthly if the conservativeness premiums were increased by 1%.

We see that the most spatial autoregressive coefficient (ρ, λ) is highly significant. Several of the coefficient estimates are quite different from the OLS results in their magnitude and significance. Note that the estimates of the significant coefficients in the SLM are all smaller in absolute value relative to the OLS counterparts, which might imply the non-spatial model absorbs some degree of spatial autocorrelation in its estimates for the OLS regression coefficients. The largest change is for the xMKT term, which goes from 0.1485 in OLS to 0.1069 (ACNV) in SLM and from 0.2468 in OLS to 0.1692 (Dist), both are highly significant. Other coefficients whose estimates are affected include HML in Dist (going from 0.2732 in OLS to 0.2257) and WML in BHFI (going from 0.0417 to 0.0398). The lack of significance of PTFSIR (BHFI, ELS, ELB, PREQ, AED, BTOP50, and CTA) and PTFSCOM (PREQ, Tech, ACNV, AFI, MS, and BTOP50) mostly remains. While the coefficients of PTFSBD in BHFI and ELB are no longer significant in the SLM, which would illustrate the effect that spatial autocorrelation might have on OLS estimates (Anselin and Rey, 2014), the coefficients of PTFSSTK in AED, PTFSFX in ACNV/AFI, and RMW in BTOP50 recover their significance upon the inclusion of spatial specifications.

One of the insights from spatial dependence is that the value of the hedge fund return at any given location depends on the factor risk premiums at all other observations in the system. In particular, Kim et al. (2003) argue that if the column vector x_h experiences an equal unit change at all locations, the resulting total effect on y is ($\beta_h/(1-\rho)$). In the SLM, the total effect of a change in the explanatory variables on the dependent variable, ($\beta_h/(1-\rho)$), is consisted of one due to the direct effect of x_h at each location, β_h and the other due to the indirect effect driven by the spatial multiplier, $\beta_h\rho/(1-\rho)$. Since $|\rho|<1$ in practice, the distance decay effect will tend to die out for an order of contiguity well below n, i.e., if locations are far enough apart, there will only be a negligible spatial correlation between these hedge fund returns. The partial derivative of a column vector concerning all the elements of a row vector is a matrix since the effect must be assessed of a change in every one of the elements of the row in each element of the column. More precisely, $\delta W_{x_h}/\delta x'_h = W$. With $|\rho|<1$, the spatial multiplier in effect amplifies the direct effect of x_h on y at each location. Considering the monthly estimate for the explanatory variable xMKT, which yields 0.099% (AFI) in the non-spatial OLS model and consistently gives 0.0657% in the SLM. The value of 0.0657% is a direct effect and the total effect of a change in a continuous explanatory variable can be computed as $\hat{\beta}/(1-\hat{\rho})$. Using the estimated value for $\hat{\rho}$ from the first-order spatial lag instruments, this yields 0.0657/(1 - 0.3567) = 0.1021% for xMKT, a more than 55% increase in the magnitude of the original estimate. Of the total effect, 0.0364% per month is due to the spatial multiplier. The estimated total effect of 0.1021% is also higher than the estimate of 0.099% obtained in the non-spatial OLS model. This illustrates the extent to which the spatial multiplier changes the analytical interpretation of the marginal effects of the excess market risk premium to the hedge fund performance.

Any time we consider a sum of squares in the spatial domain, a sum of squares of spatially weighted residuals are the proper ones to use. Typically, the way we measure the relative importance of the lag vs. the rest is by looking at L.R. test or in an *ad hoc* manner looking at how the adj- R^2 changes. The adj- R^2 is some correction in favor of parsimony. As we load up the regression, our R^2 goes up but if we take something out of that R^2, we want to make sure that we are adding significant new variables. This is something along the same lines with the Akaike Information Criterion (AIC), which corrects the Log-Likelihood for the $2L$. $2K$ is the number of variables in the model: $AIC=-2L-2K$. Since the real purpose of including the spatial lag operator is to get a consistent estimate of β, ignoring the spatial dependence can give us a highly misleading inference. For instance, if we translate these coefficients and multiplying with dollar values, they could be very different, which is not just a precision issue, this is an actual point estimate issue as well. In the output, the adj- R^2 gives us the estimated asymptotic t-value and the same measure-of-fit. The R^2 that is given at SLM and SEM is not a regular R^2. Because it is a very crude indicator of fit, we would rather rely on the Log-Like or AIC if the models are in very different specifications. The Pseudo- R^2 values are in general not directly comparable due to how they are calculated, represent the proportion of the variation of hedge fund returns that is accounted for by each model and are remarkably improved especially for the spatial autoregressive models, for instance, ELS (going from 0.84 in OLS to 0.867 in SLM), PREQ (from 0.659 to 0.718 in SEM), AED (from 0.779 to 0.811 in SLM), ACNV (from 0.583 to 0.689 in SLM), AFI (from 0.485 to 0.574 in SLM), Dist (from 0.631 to 0.731 in SLM), MS (from 0.207 to 0.314 in SLM), and CTA (from 0.289 to 0.4 in SLM). The standard error of the estimate indicates the extent to which the estimated sale prices vary from their actual values, and the values are remarkably similar across three models, but slightly improved for the SLM. In a similar vein, the SLM and SEM have a lower AIC value compare to the OLS model suggesting the superiority of spatial models in terms of the trade-off between its goodness-of-fit and complexity.

However, this does not end the treatment of pure spatial correlation, pure spatial dependence, and spatial heterogeneity in the hedge fund data set. The value for the Breusch-Pagan (B.P.) test for spatial models are 44.8 (BHFI), 22.16 (ELS), 27.44 (ELB), 29.9/31.98 (PREQ), 30.26 (AED), 208.04 (ACNV), 361.54 (AFI), 20.99 (Dist), and 154.55 (MS), all of them are still highly significant for a χ^2 variate with 11 degrees of freedom, strongly indicating the presence of remaining heteroscedasticity even after the treatment to correct the spatial dependence.

When some of the explanatory variables are correlated with the error term, we refer to them as endogenous variables. This simultaneous equation bias violates one of the basic assumptions of underlying OLS, namely $E[X'u] = 0$, which may suggest the need to use heteroscedasticity and autocorrelation consistent (HAC) standard error estimate for inference. The monthly log-returns of implied volatility

Exhibit 6. Estimation by Spatial Weighted Two-Stage Least Squares-Equity Strategies

Variable	BHFI	ELS	ELB	PREQ (L)	PREQ (E)	PREQ (LE)	Tech
Intercept	0.0008	0.00214***	0.0009	0.002	0.0013	0.002	0.0055***
xMKT	0.3057***	0.1967***	0.5113***	0.2189***	0.3381***	0.2206***	0.2868***
SMB	0.2109***	0.1732***	0.2790***	0.2327***	0.16	0.2308**	0.2023**
HML	0.0655	0.0539	-0.0097	0.1457	0.0656	0.1461*	-0.1537
RMW	-0.0195	-0.1400**	-0.1886***	-0.1535	-0.1494	-0.1523	-0.2694**
CMA	-0.2799***	-0.2356***	-0.2332***	-0.2341*	-0.1216	-0.2337*	-0.4841***
WML	0.0428***	0.0847***	0.0502**	0.0864**	0.0790**	0.0869***	0.1599***
PTFSBD	-0.0052	-0.0053	-0.0068	-0.0055	-0.0065	-0.0055	-0.0173**
PTFSFX	0.0046	0.0059*	0.0081**	0.0041	0.0035	0.0041	0.0203***
PTFSCOM	-0.0071**	-0.0067*	-0.0091*	-0.0022	-0.0063	-0.0024	-0.0173
PTFSIR	-0.0028	0.0004	-0.0007	-0.0004	0.0003	-0.0004	0.0083
PTFSSTK	0.0034	0.0136**	0.0075	0.0144	-0.0032	0.0142	0.0182
VIX (end)	-0.0044	-0.0162*	-0.005	-0.0235	0.0068	-0.0231	-0.0308**
Lambda (λ)					0.4260***	0.0139	
Rho (ρ)	0.1324**	0.1405**	0.1375***	0.2588**		0.2576**	0.002
S Pseudo-R^2	0.929	0.845	0.942	0.664	0.688	0.665	0.735
A.K. (d.f. 1)	0.107	0.924	0.008	0.000			1.558

*Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Exhibit 7. Estimation by Spatial Weighted Two-Stage Least Squares-Arbitrage Strategies

Variable	AED	ACNV	AFI	DistS	MS	BTOP50	Bar CTA
Intercept	0.0025***	-0.0008	-0.0023*	0.0001	0.0005	0.0007	0.0012
xMKT	0.1767***	0.1759***	0.1565**	0.1848***	0.1454***	0.0772	0.0751
SMB	0.2891***	0.1178	0.0316	0.1388	0.1828***	-0.072	0.0365
HML	0.1298*	0.0607	-0.0739	0.1705	0.0208	0.0434	0.0726
RMW	-0.0598	0.0137	0.05	-0.132	0.0288	0.2817*	0.1907*
CMA	-0.3012***	-0.3278**	-0.0939	-0.2734**	-0.2653***	-0.1008	-0.1373
WML	0.0306	-0.0505	-0.0224	0.02	0.0227	0.1289**	0.1127***
PTFSBD	-0.0142**	0.0111	0.0054	-0.0147*	-0.0004	0.0102	0.0084
PTFSFX	0.0043	-0.0129*	-0.0130**	0.0008	0.0017	0.0194**	0.0228***
PTFSCOM	-0.0071	-0.0005	-0.0018	-0.0135*	-0.0011	0.0165	0.0233***
PTFSIR	0.0003	-0.0166***	-0.0197***	-0.0079	-0.0091***	0.0057	0.0052
PTFSSTK	0.0187**	-0.0098	-0.0156	0.0022	0.0012	0.017	0.0147
VIX (end)	-0.0190*	0.0256*	0.0379**	0.0118	0.0017	-0.0181	-0.0179
Lambda (λ)							
Rho (ρ)	0.2616***	0.6631***	0.7450***	0.5624***	0.4155***	1.0493***	0.8698***
S Pseudo-R^2	0.796	0.562	0.492	0.717	0.716	0.309	0.237
A.K. (d.f. 1)	0.252	0.964	3.610*	2.407	0.000	5.29**	1.972

*Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

(VIX) are introduced as endogenous, with instruments such as one-month prior excess market return (LxMKT), one-month prior momentum premium (LWML), and one-month prior look-back straddle returns of short-term interest rate futures (LPTFSIR). Since the analytical results for the asymptotic variance of the ML estimator are based on stylized asymptotic setting, ignoring any potential sources of misspecification, such as non-normality or heteroscedasticity, the GMM-based spatial two-stage least square (GMM-S2SLS) estimators tend to be robust and remain valid in the presence of this misspecification. The remainder of this section provides an analysis of the same hedge fund data with the GMM-S2SLS estimation as shown in Exhibit 6 and 7.

The major impact from accounting for endogeneity of the VIX is in the estimate of ρ, which increased from 0.1207 to 0.1324 (BHFI), 0.1274 to 0.1375 (ELB), 0.2132 to 0.2616 (AED), 0.4702 to 0.6631 (ACNV), 0.3567 to 0.7450 (AFI), 0.4297 to 0.5624 (Dist), 0.3346 to 0.4155 (MS), 0.3086 to 1.0493 (BTOP50), and 0.3577 to 0.8698 (CTA), remaining highly significant. The estimates for the endogenous specifications are lower than those do not correct for the endogeneity. For instance, the GMM-S2SLS estimates of xMKT are much smaller at 0.3057 (for BHFI vs. SLM's 0.3217), 0.1967 (for ELS vs. SLM's 0.2521), 0.5113 (for ELB vs. SLM's 0.5299), 0.2189 (for PREQ(L) vs. SLM's 0.2979, 0.3229 in OLS), 0.2868 (for Tech vs. SLM's 0.3902), 0.1767 (for AED vs. SLM's 0.2484), 0.1848 (for Dist vs. SLM's 0.1692), and become insignificant at BTOP50 and CTA (vs. SLM's 0.1341 for both) in the GMM-S2SLS case. The larger the effect of the spatial multiplier, the smaller the estimates of coefficients for the lag. Therefore, the impact of excess market risk premium has been overestimated in the classical risk factor approaches in these hedge fund indices. On the other hand, the GMM-S2SLS estimates of xMKT become larger at 0.1759 (for ACNV vs. SLM's 0.1069 vs. 0.1485 in OLS), 0.1565 (for AFI vs. SLM's 0.0657 vs. 0.099 in OLS), and 0.1848 (for Dist vs. SLM's 0.1692 vs. 0.2468 in OLS) than the SLM counterparts because of the bigger coefficient for the lag. One way to interpret these differences is that the introduction of the neighboring historical observations of hedge fund returns adjusts the estimates for the other coefficients such that the impact of heterogeneity was partially removed. Therefore, in these three hedge fund indices, the correction for endogeneity better reveals the true impact of excess market risk premiums to the contemporaneous hedge fund returns, which has been underestimated at the classical approaches.

With a first-order spatial lag for the instruments, the Anselin-Kelegian (A.K.) statistic yields a value of 0.107 (BHFI), 0.924 (ELS), 0.008 (ELB), 0.000 (PREQ), 1.558 (Tech), 0.252 (AED), 0.964 (ACNV), 2.407 (Dist), 0.000 (MS), and 1.972 (CTA) are not significant with $p > 0.10$. This suggests the inclusion of the spatial lag term (or spatial error term for ELS) has corrected for the spatial autocorrelation that the spatial lag specification in GMM-S2SLS estimation is likely sufficient to address the evidence of spatial dependence in the estimations in Exhibit 4 and 5. In the meanwhile, the A.K. test for residual spatial autocorrelation statistic of BTOP50 is 5.29, which is significant ($p < 0.05$), suggesting the presence of residual spatial autocorrelation in this model. If we take a very conservative p-value of 0.10, we would conclude that at the AFI (3.610) estimation has remaining spatial autocorrelation. However, the test does not provide a suggestion as to whether the alternative is a lag or an error specification.

Overall, the standard errors are smaller when the spatial lags of the instruments are used, in line with greater efficiency that may be expected when more instruments are used. Similarly, the fit, as suggested by the Spatial pseudo- R^2 is slightly worse for the lagged instruments included relative to when the lags are not included. Also, the smallest standard errors are systematically given by the OLS results, which may be misleading since they ignore the presence of heteroscedasticity and spatial autocorrelation. The spatial estimation of PREQ (LE) in Exhibit 6 is based on Eq. 7 of the Error-Lag combined model, where the value of λ is insignificant 0.0139, which bears no relation to the highly significant estimate found in the pure SEM of PREQ(E) (0.4260). In contrast, the estimate for the autoregressive coefficient ρ is 0.2576, highly significant, and very close to the value obtained in the pure SLM of PREQ(L)'s 0.2588. This would suggest the benefit of including a fully specified spatial autoregressive error term is quite marginal in this hedge fund index analysis. The most drastic difference pertains to the coefficient of SMB which goes from a non-significant 0.16 to a significant 0.2308 ($p < 0.05$). While the estimated value of λ coefficient of Eq. 7 in PREQ(LE) does not necessarily point to a meaningful pattern of spatial spillover, it corrects for remaining spatial heterogeneity.

While both pseudo- R^2 and spatial pseudo- R^2 measures across the SLM/SEM are computed as squared correlations between the observed and predicted values for the hedge fund returns, none of these measures is true R^2 because they do not correspond to the share of variance explained by the model. The usual interpretation of adj- R^2 based on the decomposition of the variance into an explained and a residual component therefore no longer holds. The pseudo- R^2 is only an approximate measure of fit and should be interpreted as such. It can be used as a rough guideline in model selection but is not an indication of the proportion of explained variance. In the absence of an alternative, the spatial pseudo- R^2 is the preferred indicator[9] to assess the relative model fit in an *ad hoc* fashion.

A couple of more things for further elaboration; one is the measure of fit. If we have different weights, how we can tell one model from the other one, it all ties back to the objective function. But one possible objective function is the measure of fit. In the spatial model, because of the correlation, the standard R^2, which is a sum of squared residuals treating each hedge fund returns as equal is inappropriate. Because of the correlation, we have to treat them as unequal: Using spatial correlation tells us how differentially weight each observation. If we think of the heteroscedastic case, often is tackled through Weighted Least Squares and the weights are inverse to the variance: If we have observations with large error variance, they count less. If they have a smaller variance, they count more. If we apply in our adj- R^2 criterion, it counts everybody equally, that's not appropriate. What we really need is some criterion that doesn't count each observation equally and that is provided by the likelihood. In that sense, as long as the dependent variable is the same, then we can compare the log-likelihoods ("Log-Like") to maximize Log-Like if it's between different specifications. In our case, our particular interest will be, how does this affect the weights matrix? We have used a triangular kernel function with an adaptive bandwidth based on the 4-nearest neighbors. If we have two models saying basically the same specification in terms of X's with two different weights matrices, the one with the highest maximized Log-Like is the one that has the best fit and that would be the preferred. This is not a test since there is no p-value. So, there is no way of telling how much better this fit is. The only thing we know that it is better but in the context of significance, they may be the same. We cannot really tell.

Then the discussion in practice is if there are some slight differences in which variables we can count, which of the spatial variables do we count? Do we count what we call nuisance parameters, which are the parameters of error variance, or do we not count? Some people count them, some people do not. That's going to give slight differences between the AICs. The AIC criterion is used to find the model with the lowest value. So, the higher the like-

9) Since the residual u in the spatial lag model is not used for the error variance under the homoscedasticity, spatial pseudo- R^2 measure of fit is based on the predicted values from the reduced form expressed as $\hat{y}_R = (I - \hat{\rho}W)^{-1}(X\hat{\beta} + Y\hat{\gamma})$, which gives a vector of predicted errors as $e_R = y - \hat{y}_R$. The predicted values $\hat{y}_R$ are used to compute a spatial pseudo- R^2.

lihood, the lower the AIC. Often, they will point to the same model but when there are huge differences between the number of explanatory variables, we might end up with a difference between the AIC as a criterion and maximized Log-Like. But in our work, we are interested in the weights matrix and this is not going to make a big difference. In that case, what do we use an AIC or Log-Like doesn't matter because K as in $AIC = -2L - 2K$ is the same. Any time K is the same, there's not going to be a difference between these two. When we compare vastly different models, the AIC is more appropriate guidance because it corrects for that and it corrects in favor of parsimony, i.e., in favor of models with fewer parameters in them. In this particular study, the correction for heteroscedasticity and endogeneity seems warranted, at least in terms of substantive interpretation of the excess market risk premiums.

Another question in our mind might be, is there a feasible solution to extract spatial lag from the errors? The real issue is, even after we have a satisfactory model, whether the model is turned out to be misspecified in some way or not. Then what we really want to look at that point is the residuals of that model. However, in very specific, the residuals of interest are not necessarily the difference between the predicted value and observed value because these are transformed residuals. The ones that enter into the likelihood function are not the raw residuals, they are in fact spatially-filtered residuals. Then we have to look at patterns in the spatially-filtered residuals, not in the raw residuals because if we just estimate by feasible Generalized Least Squares and we take these residuals, they still have the spatial effect in them. To get the spatial effect out of them, we have to spatially filter them as well. If we measure the Moran's I to the raw residuals, it might still look to be correlated but once we spatially filter them and the Moran Scatter Plot falls flat, which means we essentially have removed spatial autocorrelation from the financial time series dataset. It's always important to know where the line stops and so to deal with the residuals that are compatible with our original intention.

[4] Case Study: No Evidence of a 'meant-to-be' Spatial Dependence in a failed credit hedge fund

"International Investment Group (IIG), which specializes in trade finance lending to small and medium-sized companies in Latin America, has had its license revoked by the U.S. Securities and Exchange Commission (SEC) following what the regulator calls "a string of frauds". As investment advisor to the Trade Opportunities Fund, the Global Trade Finance Fund, and the Structured Trade Finance Fund, IIG offers institutions and other investors the opportunity to invest in diversified trade finance portfolios, originally through fund products and subsequently through other types of investment vehicles, such as collateralized loan obligations. (…) The amount sold in fake loan assets to clients by IIG was around US$60mn, the case illustrates the risks inherent to alternative trade finance. On its website and in investor prospectuses, IIG touts its risk control strategies, which include portfolio concentration limits at the borrower, country, and commodity level, as well as a "robust" credit review process for borrowers. However, a series of defaults, including on a US$30mn loan by Trade Opportunities Fund to a South American coffee producer and a US$30mn loan to an unnamed seafood producer, soon saw the firm getting into trouble." **"IIG trade finance fund caught in alleged Ponzi scheme," *Global Trade Review*, December 9, 2019.**

"TCA is a registered investment adviser which has offices in New York, Las Vegas, London and Melbourne. TCA Global Credit Fund GP, Ltd., a Cayman firm , was also named as a defendant. That firm serves as the general partner of Master Fund and Feeder Fund. Over a nine-year period tracing to 2010, TCA has falsely inflated NAV and its compensation through two methods. First, fees paid to the Master Fund were prematurely recognized. When that Fund entered into a lending arrangement typically a term sheet was executed. Rather than waiting for the deal to close and the fees to be earned, TCA had Master Fund recognize the fees at the time the term sheet was signed. Second, over a three-year period, beginning in 2016, the adviser caused the Master Fund to prematurely recognize fees for investment banking services. Specifically, when an agreement for services was entered into TCA immediately recognized the fees as income despite the fact that little if any of the services had been provided. Collectively, these practices left Master and Feeder Funds in difficult financial conditions. NAV had been falsely inflated by almost $160 million. In 2017 and 2018 the auditors issued a qualified opinion with respect to 89% of the Master Fund's NAV. In January the Feeder Funds were forced to suspend redemptions. The complaint alleges violations of Securities Act Section 17(a), Exchange Act Section 10(b) and Advisers Act Sections 206(1), 206(2) and 206(4). The case is pending." **"SEC v. TCA Fund Management Group Corp., Civil Action No. 1:20-cv-21964 (S.D. Fla. Filed May 11, 2020).**

Since its launch in March 2010, the fund of the case study has been known as a predominately short-duration, absolute return private credit fund specializing in senior secured lending as well as bespoke investment banking and advisory services to small and medium-sized companies mainly in the US, Canada, Western Europe, and Australia. The fund was known to have some focused strategies such as; (a) the steady origination of $10~$25mn quarterly; (b) loans are generally receivables-driven, focusing on the most liquid part of the balance sheet; (c) small position sizes, typically less than 1% of the portfolio with average loan size $1~$2mn; (d) loans are generally senior secured, income-generating, and short-duration with targeted annual net returns to investors of 8~12%; (e) no investment style drift, no shorting[10]; and (f) utilize industry relationships to generate non-lending related investment banking services to the small and midcap markets.

Exhibit 8. Spatial Dependence Examinations in a failed credit hedge fund

Variable	OLS	LM-SLM	LM-SEM	GMM-S2SLS	GMM-2SLM	GMM-2SEM	GMM-SLEM
Intercept	0.0097***	0.0109***	0.0097***	0.0124***	0.0078*	0.0124***	0.0054*
xMKT	0.0282	0.0282	0.0275	-0.1812	-0.1231	-0.1808	-0.1212
SMB	0.0779	0.0806	0.0782	0.2426*	0.1885	0.2421	0.192
HML	-0.0995	-0.1007	-0.1053	0.0511	0.0132	0.0511	0.0213
RMW	-0.1878	-0.1723	-0.1726	-0.2325	-0.2693*	-0.2339	-0.1989
CMA	0.2329	0.2347	0.234	0.0253	0.0772	0.0256	0.0299
WML	0.0328	0.0287	0.0253	0.0986	0.0933	0.0989	-0.1989
PTFSBD	0.0149	0.0152	0.015	0.0066	0.008	0.0066	0.0058
PTFSFX	-0.0137	-0.0138*	-0.0138*	-0.0092	-0.0101	-0.0091	-0.0087
PTFSCOM	0.0089	0.0085	0.0084	0.0148	0.0146*	0.0148*	0.0126*
PTFSIR	-0.0046	-0.0044	-0.004	-0.0064	-0.0063	-0.0065	-0.0028
PTFSSTK	0.0032	0.0032	0.0034	0.0281	0.0212*	0.0281	0.0199*
VIX (end)				-0.0525*	-0.0379**	-0.0524	-0.0355*
Lambda (λ)			-0.0951			0.0208	0.6194**
Rho (ρ)		-0.1163			0.3698		-0.5376
Moran's I	-0.0296						
Adj-R^2	-0.0317	0.0842	0.0814	0.0991	0.096	0.0992	0.0805
Log-Like	317.003	317.266	317.15				
AIC	-610.007	-608.533	-610.301				
JB (d.f. 2)	932.02***						
B.P. (d.f. 11)	92.76***	93.02***	94.44***				
K.B. (d.f. 11)	11.48						
L.M.-Error	0.6905						
L.M.-Lag	0.4939						
L.R. (d.f. 1)		0.5263	0.2943				
A.K. (d.f. 1)				0.054	0.696		

*Note: * p < 0.10, ** p < 0.05, *** p < 0.01*

The offering circular of the fund explains to use two primary loan structures: The majority of the fund's loans are short-term, senior secured, revolving lines of credit where the primary direct collateral used for loan sizing is the borrowing firm's near-term receivables and vendor payments due to them. This represents generally over 98% of the transactions in the portfolio. The second structure is a short-term self-liquidating debenture (e.g., a revenue bond with a sinking-fund feature), which is also a senior or directly secured instrument that generally has an "estoppel" or direct payment from a much larger customer of the borrower. Measured from a fixed income perspective, the fund's loan portfolios are extremely short-term. The revolving lines of credit have a term (renewable at the fund's option) of 180 days, but are 'proved up' each week and are based on collateral that is generally not more than 90 days. The fund's few amortizing loan structures are 6- to 12-month

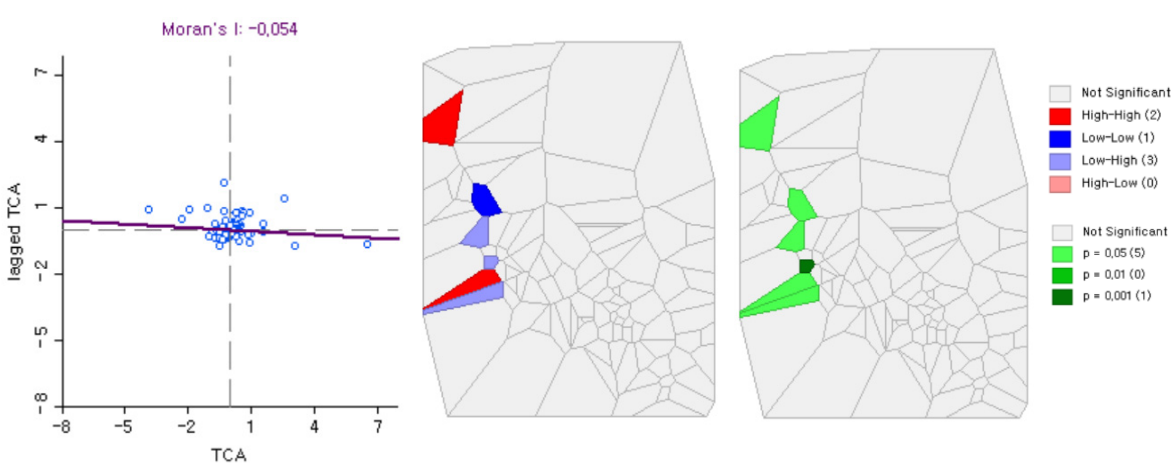
10) After all, there is no feasible way available to the managers of private debt funds to short the loans of the non-rated, small and medium-sized companies even in the highly sophisticated U.S. markets.

self-liquidating instruments. The fund takes a "blanket" lien or a complete fixed and floating charge over all of the borrower's assets. One indicator that is most important to the fund was the 'Debt Service Coverage Ratio,' which is a measurement of the cash flow available to pay current debt obligations. By having an average loan-to-value (LTV) of 70%, the fund is at least 130% overcollateralized.

By assuming the independence of the observations, the OLS ignores the presence of spatial autocorrelation. Based on monthly net-of-the fee returns data from April 2010 until December 2018, we run seven different factor risk premium models from OLS to the GMM-S2SLS of both lag and error models with the summarized estimation result in Exhibit 8. Before running the spatial autoregressive models, spatial autocorrelation tests via the global Moran's $\boldsymbol{I}$ statistic were performed. From Exhibit 9, the diagnostic statistic of the Moran's $\boldsymbol{I}$ on the fund return data shows -0.054, which implies no evidence of smoothing whatsoever in the fund's NAV. Likewise, Moran's $\boldsymbol{I}$ statistic for the errors ε of the OLS model shown in Exhibit 8 was also estimated to the value of -0.0296 but insignificant, which is also indicative of no-spatial autocorrelation in the obtained OLS residuals.

The significance of the spatial parameters ρ and λ in the SLM and SEM models are indicative of the effects of spatial autocorrelation in the fund returns. However, the examination of the coefficients indicates that the use of the spatial lag term $\rho \boldsymbol{W} \boldsymbol{y}$ in the SLM was simply insignificant. In the presence of strong spatial autocorrelation, likely, the spatial regression model will significantly outperform the OLS model, which is not the case as shown in Exhibit 8. Furthermore, the large difference between B.P. and K.B. test statistic confirms potential error non-normality, as already confirmed in its highly significant Jarque-Bera statistic, since under normality, the value for both should converge each other. The GMM-SLEM estimation in the last column of Exhibit 8 is based on Eq. 7 of Spatial Lag and Error specification. Eq. 5.3 shows $y = (\lambda + \rho)Wy - \lambda\rho W^2 y + X\beta - \lambda W X\beta + \varepsilon$.

Exhibit 9. Moran's $\boldsymbol{I}$, LISA Cluster, and Significance Maps of a failed credit hedge fund



With λ estimate of 0.6194 and ρ estimate of -0.5376, the first coefficient of $(\lambda + \rho)$ term would end up +0.082, which is near to zero and negating any significant influence from the first-order neighbors. The impact of second-order neighbors ($-\lambda\rho$) becomes positive since the opposite sign of the estimated coefficients in the second term.

The Local Indicators of Spatial Association (LISA) significance map shows the locations with a significant local statistic, with the degree of significance reflected in increasingly darker shades of green. The map starts with $p < 0.05$ and shows all the categories of significance that are meaningful for the given number of permutations. In the third panel of Exhibit 9, since there were 999 permutations, the smallest pseudo p-value is 0.001, with the darkest shade of green locations. The second panel of Exhibit 9 shows the cluster map of the fund returns which augments the significant locations with an indication of the type of spatial association based on the location of the value and its spatial lag in the Moran Scatter Plot. With all four categories; dark red for the high-high clusters, dark blue for the low-low clusters, and light blue for the low-high spatial outliers, the fund demonstrates both Hot (red) and Cold (blue) Spots below the reversed 45-degree line of the Voronoi map of Thiessen Polygons symbolized by the entropy of variation between neighboring returns, which is a very rare phenomenon in the world of short-volatility trades. As the positive correlation of the Hot Spot implies the locations with high fund returns with a high level of similarity with its neighboring historical returns, the Cold Spot implies the locations with low hedge fund returns with a high level of similarity with its neighboring returns. Here, the Hot Spots can be observed when the equity market index level is relatively low but VIX may be either low and high but no signs of significance in between, while the small evidence of the Cold Spot stays in the mid-level of VIX.

We have a few qualitative points from the spatial analytical perspective: Firstly, the adjusted- R^2 from the OLS model was negative and none of the rest of spatial pseudo- R^2 value was over the threshold of 0.1, which implies that any typical risk factor premiums are no viable exogenous explanatory variable, which is quite exceptional. Secondly, our separate analysis shows that the share of variance explained by the S&P U.S. High Yield Corporate Bond Index ("HYI") to the variations of the S&P/LSTA U.S. Leveraged Loan 100 B/BB Rating Index ("LSTA") marks 69.8% during the period, both HYI and LSTA do not have any significant explanatory power to this failed credit hedge fund. With R^2 value of mere 0.012 and non-significant negative estimated coefficients, the failed credit hedge fund might be a wholly different kind of animal to deal with. In fact, as with the usual case of analyzing leveraged loans and high-yield indices, those loans with external credit ratings of B/BB should come along with highly significant exposures to the excess market risk premiums (xMKT), positive size (SMB), and negative CMA (thus aggressive capital structure) risk premium characteristics. However, none of the spatial or non-spatial specifications show a slight bit of significance in xMKT, SMB, CMA, and even PTFSIR premiums. Because the fund's maximum duration of its disbursed loan was stated as less than 180 days during the initial conference calls with us, the pricing and revenue structure of the fund should have been highly sensitive to the fluctuation of short-term money market futures and the relevant derivatives. Since the size of the intercept, which can be marketed as manager's alpha, ranges from 97 to 124 bps per month, the fund manager should have proudly announced the uniqueness of their absolute return strategy even though the fund specializes in small and medium-sized lenders with no external credit ratings, i.e., junk-loans. Thirdly, the fund did not value the loan portfolio according to the viable marking-to-market practice, which means the fund manager price its portfolio to the 'best of the manager's knowledge and experience.' Certainly, the quarterly reports of operation review on the validity of the pricing of the fund's net asset value (NAV) were available after subscription of the fund but any prior review of the report as part of investor due diligence was not typically allowed. This might be a classic case of spatial autocorrelation on its returns. In summary, the spatial variation of the credit hedge fund returns failed to be within the spectrum of conventional knowledge and wisdom. Then, some thorough due diligence question and answer sessions should have accompanied before any further commitment to this failed credit manager with 'overly outlying spatial out-performance.' Once the track record of a manager seems too good to be true, it probably is not, at least within the credit hedge fund domains.

[5] Conclusion

Unlike traditional multi-factor performance attribution analysis, which usually attempts to explain hedge fund returns in terms of classical and alternative factor risk premiums, considering the neighboring historical return characteristics through the 'imaginary' geographic location grids, spatial econometric models in general explicitly account for two major spatial effects in hedge fund returns typically ignored in global models: Spatial dependency and spatial heterogeneity. The former refers to the similarity commonly observed in the values of nearby historical returns whilst the latter indicates that the processes generating hedge fund returns might vary over space, usually reflecting fluctuations in risk appetites. Parameter estimates from traditional OLS models, which represent the relationships between hedge fund returns and the associated time-varying risk premium characteristics, can be biased in the presence of these spatial effects. With spatial independence, the OLS estimator is the best linear unbiased estimator (BLUE). However, the spatial dependence in the hedge fund returns might be present because neighboring historical hedge fund returns are more alike than long-distance returns in the grid surfaces. Furthermore, the OLS estimator is inefficient due to the presence of a correlation in the disturbances. Since the spatial dependence and spatial heterogeneity frequently coexist in many spatial processes, factor risk premium models capable of addressing both spatial effects become a desirable option. Given the importance of understanding spatial variation in hedge fund returns and obtaining the grid surfaces of these variations, the methodologies discussed in this paper enable the management of the heterogeneity of financial time series and it may be one of the main contributions of our work. To this end, of primary interest here is to understand both spatial effects in the processes of hedge fund returns through the application of the GMM-S2SLS augmented by the endogeneity of implied volatilities.

The major findings and implications of this study are summarized follows: First, a signifi cant spatial dependence in the return process was found among the hedge fund indices of BHFI, ELS, ELB, PREQ, Tech, AED, ACNV, AFI, Dist, MS, BTOP50, and CTA. This implies that a conventional factor risk premium model based on OLS estimation requires further augmentation. The spatial performance attribution used to address the spatial dependence problem outperformed the conventional performance attribution analysis in terms of goodness-of-fit measures. The coefficient estimates of the excess market risk premium of the spatial model were mostly smaller than those of the classic OLS model, except for ACNV, AFI, and Dist, indicating that the effects of market risk premium would have been overestimated in the classical attribution analysis. Besides, the spatial model revealed that the process of hedge fund returns had a positive impact on those of adjacent units. For instance, the monthly coefficient estimate for xMKT yields 0.099% (AFI) in the non-spatial OLS model but consistently gives 0.0657% in the SLM, which is the direct effect. Under the total effect of a change in xMKT can be computed as 0.1021% by $\hat{\boldsymbol{\beta}}/(\mathbf{1}-\hat{\boldsymbol{\rho}})$, over 55% increase in the magnitude of the original estimate. Of the total effect, 0.0364% per month (35.7% of the total effect) is due to the spatial multiplier. The estimated total effect of 0.1021% is also higher than the estimate of 0.099% obtained in the non-spatial OLS model. This illustrates the extent to which the spatial multiplier changes the interpretation of marginal effects in the hedge fund performance attributions.

There is no strong evidence for the superiority of combining both nuisance and substantive spatial dependencies in a single model. The major implication is to keep the specification simple. In particular, the value of the λ coefficient in the spatial error and lag combined model does not necessarily point to a meaningful pattern of spatial spillover but ends up correcting for remaining spatial heteroscedasticity. The standard errors show the usual pattern, with values increasing in order from the classic OLS result to White and HAC. The latter consistently yields the larger values in most standard error estimates, which suggests we should use caution in the interpretation of the standard approaches because the classic OLS results for the coefficient standard error may provide a false sense of precision. In applications where precision plays a crucial role, such as the calculation of economic costs and benefits, this should be properly taken into account.

The parameters and their standard errors are also stable and, in this way, the results from the spatial lag and error models seem to be robust. Estimating the lag model gives a significant value of the spatial dependence parameter ρ for one type of neighboring structure of queen contiguity. This indicates that spatial dependence or the adjacency effect to some extent play a role in determining the return generation process of some hedge funds. The secondary feature of the data is spillover effects in the residuals, which may be accounted for by using the SEM. Although these global spatial models represent a substantial improvement over classic performance attribution analysis, a major issue is that the hedge fund return processes are assumed to be stationary over space in the longer term, which is not necessarily the case in practice.

Lastly, the spatial modeling might be applicable to an *ad hoc* screening and help to create a more focused due diligence questionnaire. Based on our examinations and findings through the spatial exploratory framework applied to the rare time series for possible irregularity detection, it is basically impossible to elicit any meaningful information from the monthly returns of a recently failed credit hedge fund. The manager might claim and position itself as an absolute return manager with 100% short-duration, long-only exposures to low-profile credits. However, at the onset of an economic downturn, earnings of those small and medium-sized low credibility firms tend to fall sharply. Most forms of risk premium connected with credit markets have short volatility characteristics because the fundamental idea behind a risk premium is to gain a higher return in exchange for accepting systematic risk. When risk aversion rises, the prices of many risky assets fall including the implied price of low or no credit loans. Furthermore, many risk premia are correlated because they rely on healthy economic activity, cash flow, and profits. The fund would demand that risk premia generate excess returns because they subject the fund to additional risk. The fund has a special need for robust returns from its investments when the economy is weak to supplement diminished incomes from labor, corporate cash flow, or tax revenues. But this is also precisely the time when most of the portfolio companies (i.e., the borrowers) of the fund would perform worst. Therefore, without any concrete evidence of spatial autocorrelations in the monthly return profile, it is highly unlikely to accept the manager's track records as it is. The further detailed investors' due diligence processes by highly experienced professional calibers considered to be a viable preventive measure from lines of the victims of any hedge fund Ponzi scheme going forward.

Acknowledgment

We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. We also thank Mr. Sebastian Schäfer, Head of Systematic Strategies of QCAM Currency Asset Management AG, Mr. Hans Kim, Institutional Advisory of QCAM Currency Asset Management AG, Mr. Seon Goo Lee, C.E.O. of Sellymon.com, the most comprehensive online algorithmic tax information platform, and Ms. Karen Jo from

Landmark Asset Management in Seoul for their discussion and support for the development of the alternative investment solutions for the qualified Korean institutional and private investors. Some part of the elaboration is based upon various insightful inputs from these gentlemen, but any error remains solely responsible to the author of this article. This research was supported by Seokyeong University in 2020.

References

Agarwal, V. and Naik, N.Y. "Risk and Portfolio Decisions Involving Hedge Funds." *Review of Financial Studies,* Vol. 17, No. 1 (2004), pp. 63-98.

Anselin, Luc and Rey, Sergio J. "Modern Spatial Econometrics in Practice." *GeoDa Press LLC* (2014).

Cho, J.K. and Kim, G.W. "Making a Bet at a Right Time: Style and Volatility Timing Abilities of Korean Equity Hedge Funds." *Asset Management Review*, Vol. 5, Issue 2, December 2017.

Cho, J.K. (a) "Implied Equity Volatility and Liquid Alternative Hedge Fund Investing by Qualified Korean Investors." *Information Journal*, Vol. 20, No.10(B), October 2017.

Cho, J.K. (b) "Information Content of the VKOSPI Entropy Indicator - A Dynamic Asset Allocation Approach." *Information Journal*, Vol. 20, No.7(B), July 2017.

Cho, J.K. (c) "Qualified Korean Investors' Liquid Alternative Global Hedge Fund Investing under Autocorrelation Effects." *Journal of Economic Studies (Korean)*, Vol. 35, Issue 2, May 2017.

Cho, J.K. "A Conditional Volatility Feedback Moderation Modeling to the Market Timing Behavior per Hedge Funds' Investment Styles." *Asset Management Review*, Vol. 7, Issue 1, June 2019.

Cho, J.K. "Cross-border Tax Implications in the U.S. CLO Equity Investing by the Qualified Korean Investors." *Asset Management Review*, Vol. 6, Issue 2, December 2018.

Drukker, D. M., Egger, P., and Prucha, I. "On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors." *Econometric Reviews*, Vol. 32, pp. 686-733.

Getmansky, M., Lo, A. W., and Marakov, I. "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns." *Journal of Financial Economics,* Vol. 74 (2004), pp. 529-609.

Inoue, Ryo, Shimizu, Eihan, and Kigoshi, Naoyuki. "Visualization of Spatial Distribution and Temporal Change of Land Prices for Residential Use in Tokyo 23 Wards using Spatio-Temporal Kriging." In: *Proceedings of 10th International Conference on Computers in Urban Planning and Urban Management*, Iguassu Falls, Brazil: Paper 63, (2007), pp. 1-10.

Haas, T. C. "Lognormal and Moving Window Methods of Estimating Acid Deposition." *Journal of American Statistical Association*, Vol. 85, (1990), pp. 950-963.

Harris, P., Charlton, M., Fotheringham, A.S. "Moving window kriging with geographically weighted variograms." *Stochastic Environmental Research & Risk Assessment*, Vol. 24, Issue 8, (2010), pp. 1193

Harris, P., Fotheringham, A.S., Crespo, R., and Charlton, M. "The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets." *Mathematical Geoscience*, Vol. 42, (2010), pp. 657-680.

Hengl, Tomislav, Heuvelink, Gerard B.M., and Rossiter, David G. "About Regression-Kriging: From Equations to Case Studies."

Computers and Geosciences, Vol. 33 (2007), pp. 1301-1315.

Kelegian, H.H. and Prucha, I.R. "A Generalized Spatial Two-Stage Least Squares Procedures for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics*, Vol. 17, (1998), pp. 99-121.

Kelegian, H.H. and Prucha, I.R. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." *Journal of Econometrics*, Vol. 157, (2010), pp. 53-67.

Kim, C.-W., Phipps, T. T., and Anselin, L. (2003). "Measuring the Benefit of Air Quality Improvement: A Spatial Hedonic Approach." *Journal of Environmental Economics and Management*." Vol. 45, pp. 24-39.

Meng, Qingmin. "Regression Kriging versus Geographically Weighted Regression for Spatial Interpolation." *International Journal of Advanced Remote Sensing and GIS*, Vol. 3, Issue 1, (2014), pp. 606-615.

Selby, Brent and Kockelman, Kara M. "Spatial Prediction of Traffic Levels in Unmeasured Locations: Applications of Universal Kriging and Geographically Weighted Regression." *Journal of Transport Geography,* Vol. 29, (2013), pp. 34-32.

Shimizu, Chihiro, Nishimura, Kiyohiko G., and Watanabe Tsutomu. "Biases in Commercial Appraisal-Based Property Price Indexes in Tokyo – Lessons from Japanese Experience in Bubble Period." *International Conference on Commercial Property Price Indicators*, (2012), pp. 1-33.

글로벌 헤지 펀드 수익률의 공간의존성 연구

조 정 근*

Abstract

Barclays Hedge Fund 스타일 지수의 2008년1월부터 2018년 12월까지 월간 수익률 자료 분석을 통해, 당월 헤지펀드 수익률에 대한 과거 수익률의 공간적 의존성을 확인하였다. 기존 최소자승법 (OLS)과 Arbitrage Pricing Theory에 기초한 전통적 헤지펀드 성과분석방법론에서는 다루지 않은 시계열데이터의 공간의존성이란 주가지수를 경도로, 내재변동성지수를 위도로 두고 가상의 투자수익-위험의 지도를 작성한다면, 현재 시점의 투자자가 당월에 예측할 수 있는 특정한 스타일의 헤지펀드 수익률은 당월에 실현된 투자수익-위험의 위경도상 지표에서 가까운 거리에서 실현된 역사적 수익률과는 매우 높은 자기상관을 가지는 현상을 설명함으로써 W. Tobler(1970)의 지리학 제1법칙, "All things are related, but closer things are more related (모든 것은 다른 모든 것과 관련되어 있으나, 가까운 것은 먼 것보다 더 관계가 깊다)"의 적용을 가능하도록 한다. 분석을 통해 Equity Long-Short, Equity Long-Bias, Event-Driven Arbitrage, Convertible Arbitrage, Fixed-Income Arbitrage, Distressed Securities, Multi-Strategies 및 Commodity Trading Advisors 등 Barclays 헤지펀드 스타일지수의 월간 수익률 데이터는 이러한 공간의존성을 내재한 것으로 확인되었다.

금융 시계열 데이터의 공간의존성은 월간 데이터의 공간적 자기상관 (Spatial Autocorrelation), 공간적 이분산성 (Spatial Heteroscedasticity) 및 이들의 상호적인 영향을 의미한다. 본 연구는 시계열 공간의존성을 해소할 수 있는 방법으로 1차적으로 Spatial Lag Model (SLM)과 Spatial Error Model (SEM)을 활용하였으며, 특히 공간변수의 도입만으로 해소되지 않는 금융시계열 데이터에 잔존하는 공간의존성은 Generalized Method of Moment (GMM) 방법론에 따른 Spatial Two Stage Least Squares (S2SLS) 모형을 통해 상당부분 해소가 가능함을 확인했다. 공간계량경제학적 방법론을 최근에 국내외에 물의를 빚으며 운용을 중단하고 현재 미국 증권감독국 (SEC)으로부터 기소된 한 사모대출펀드의 수익률 데이터에 적용하여, 사전 Due Diligence를 통해 궁극적으로 문제성이 내재한 펀드를 가려낼 수 있는 가능성에 대한 질적 및 계량적인 분석을 적용하였다. 근거는 해당 사모대출펀드는 대출자산의 시가평가 (Marking-to-Market)를 할 수 없기 때문에 헤지펀드 운용사의 주관적인 판단에 따른 인위적인 수익률 Smoothing으로 월간 수익률 데이터의 공간적 자기상관과 이분산성이 잘 설명될 수 있으리라는 판단에 따른 것이었다. 하지만 "Once the track record seems too good to be true, it probably is not, at least within the hedge fund domains" 이라는 잠정적 결론에 도달할 정도로 기존의 정형화된 방법론 및 공간의존성을 고려한 방법론으로도 해석이 되지 않았다. 결국, 공간의존성이 필연적으로 존재할 수밖에 없는 운용전략에 공간의존성을 설명하는 방법론이 적용될 수 없다면

해당 매니저의 운용수익률 데이터는 운용전략의 질적 분석부터 다시 시작해야 한다는 판단이다.

주제어: 공간의존성, 공간시차, 공간오차, 헤지펀드성과분석
Article history : Receive 26 March 2020, Revised 21 April 2020, Accepted 23 April 2020

JEL 분류기호 : G23, G11
* 조정근 (Institutional Advisory to QCAM Currency Asset Management AG, U.S. Tax Advisory to Sellymon.com, 서경대학교 경영학부 조교수)