Diversification in Commercial Real Estate: Realizing Continuous Spatial Correlation

Darren K. Hayunga and R. Kelley Pace *

February 9, 2004

Abstract

In a mean-variance state investment managers must thoroughly specify portfolio risk to obtain full diversification and select investment options on the efficient frontier. However, diversification within real estate is not as costless as other financial assets such as stocks or bonds. The local nature of real estate implies that investment managers may need to specialize for simplicity or by portfolio objective. To address this trade-off between diversification and specialization this study quantifies the relation between commercial property returns and distance. The returns of four types of commercial properties are modeled as a function of the distance of separation by measuring the covariation between each commercial return. Results are obtained for the effects of diversifying based upon distance and also upon property type, size and distance. The results indicate that correlation for the same type of commercial properties located within the same or adjacent zip codes can be as high as 0.93. After controlling for property type and size, the correlation reduces significantly. However, when residential data are matched by location to the commercial returns and added to the controls for property type and size, the spatial autocorrelation between different property types in the same or adjacent zip codes is 0.89. In all instances, the distance at which the spatial autocorrelation decays to zero is approximately 60 miles.

JEL classification: G11, R12
Keywords: Real estate; portfolio diversification; spatial statistics

---

*E.J. Ourso College of Business Administration, Louisiana State University, 2164 CEBA, Baton Rouge, LA 70803. Corresponding author can be reached at dhayun1@lsu.edu or 225) 578-6238. Other spatial information can be viewed at http//:www.spatial-statistics.com. Research supported by a grant from the Real Estate Research Institute, Connecticut. We thank C.F. Sirmans and Jeff Fisher for help with the data, Tony Sanders for insight into the commercial real estate diversification issue, and Chris White for technical guidance.

Modern portfolio theory (MPT) as pioneered by Markowitz (1952) demonstrates that diversification can increase return for a given risk or reduce risk for a given return. Therefore, in application within perfect capital markets, individual investors can diversify their own portfolios to generate the desired return for a preferred variance. Portfolio managers have no advantage in providing diversification in the absence of restrictions and bankruptcy costs.

In commercial real estate, however, diversification is not free. The local aspect of real estate implies that commercial investment managers may wish to specialize by geography or property type. Also, a perfect capital market may not exist due to high transactions costs and restrictions (e.g.,"prudent man rule").[1] Thus, commercial investment managers may need to trade-off diversification against specialization or other considerations.

To address the trade-off, this study quantifies the correlation among commercial properties based upon distance to aid in the construction of optimal real estate portfolios. In particular, this analysis can uncover the minimum separation distance among properties that eliminates the geographic correlation for either the same or different property types. This result will potentially allow commercial real estate investors managers to fully diversify based upon geography or distance between properties. That is to say that an investment professional can specialize somewhat locally even in one property type and thereby potentially reduce the cost of diversification.

Quantifying the correlation among commercial properties in a portfolio over a measure of distance or space is a key in the construction of optimal portfolios. While previous literature has identified discrete spatial dependence for commercial property with the development of geographic locations advancing from large national regions

[1]Roulac (1976) notes that imperfect capital markets may exist in real estate due to the indivisibility of properties, the critical role of financing in the investment process, and the relatively small number of transactions.

to Metropolitan Statistical Areas (MSAs) to neighborhoods within a city, real estate policy makers and investors still do not have the tools to assess spatial dependence among commercial properties at a varying distance. Intuitively, we expect adjacent properties to exhibit some degree of spatial dependence, but research questions remain as to 1) how much initial spatial correlation is present, especially for the same property type since diversification based on type may not be desired or feasible, 2) how quickly does the spatial correlation decay, and 3) at what distance does the spatial correlation decay to zero.

To answer these questions we employ techniques from geostatistics. Commercial real estate returns are modeled as a function of the distance of separation among properties by measuring the covariation between each combination of two commercial returns. Subsequently, the empirical spatial covariation results are fit with a theoretical model, which best describes the empirical findings based upon spatial statistical standards and goodness of fit. It is this theoretical model that can provide commercial real estate investment managers with the spatial correlation over a continuous range.

While the focal point of this study is diversification in commercial real estate portfolios, spatial statistics can also be applied to residential real estate returns. Indeed, because of voluminous housing data, spatial statistics are found to be well-specified and may assist in determining the continuous spatial correlation for commercial real estate. At a minimum the use of residential data matched to commercial returns at the specific locations of all the observed commercial returns provides another avenue of spatial correlation analysis.

In general, by theoretically modeling the various empirical spatial autocorrelation findings, this study provides portfolio managers and investors with a more fully specified measure of risk. This, in turn, raises the possibility of revising existing

MPT-based portfolio strategies. Ultimately, this study can provide real estate portfolio managers and investors with the knowledge to reduce the cost of diversification.

1 Previous Studies

As might be expected, prior research regarding real estate portfolio diversification concentrates either on property types or geographic categories. Beginning in the early 1980s, Miles and McCue (1982) found that diversification by property type generated better characteristics than a strategy based upon geographic regions. Subsequent studies examined different mixtures of property type and geographic regions using a variety of databases. The real estate diversification literature by Hartzell, Hekman, and Miles (1986), Hartzell, Shulman, and Wurtzebach (1987), Mueller and Ziering (1992), Mueller (1993), Goetzmann and Wachter (1995), Williams (1996), Wolverton, Cheng and Hardin (1998), and Cheng and Black (1998) redefined the geographic categories either into more minute areas such as MSAs and neighborhoods or based upon economic strategies such as SIC groups. Overall, the research has established that: (1) geographic grouping based on economic characteristics is dominant over geographic division based upon political boundaries (i.e., state borders) and (2) smaller regions such as MSAs or neighborhoods are more appropriate for diversification than four or eight national regions.

Explicit or implied in these studies is a searching for homogeneous regions such that correlations among regions can be computed to diversify away unsystematic portfolio risk. The issue with establishing a region *ex ante* is that real estate is a local market and clearly, within a small market such as an MSA, multiple markets may exist that covary to some degree.

Real estate literature is presently addressing the search for homogeneous regions using spatial statistics. Dubin, Pace, and Thibodeau (1999) discuss applying spatial autoregression techniques to real estate data. More specific to residential data are a number of recent studies. Dubin (1998) applies geostatistics to housing data from Baltimore and finds advantages in utilizing spatial statistics. Basu and Thibodeau (1998) examine spatial autocorrelation in Dallas house prices and find spatial techniques generally improve OLS. Thibodeau (2003) finds that spatially adjusted predicted values of single-family home in Dallas are better than without spatial techniques. But although spatial statistics are employed to flesh out a more fully specified residential hedonic model, commercial real estate has yet to be addressed in the literature.

2 Specifications

To obtain the measures of spatial correlation over a continuous range a number of spatial and regression specifications must be established. The primary spatial statistical instrument is termed the variogram. Multiple types of variograms are utilized in examining commercial real estate returns.

The mathematical justification of the variogram is provided for three types of variograms in this section. The first type, the experimental variogram, measures covariation based upon the location of observed values of one type of data (e.g., commercial real estate returns). Another type of variogram is the experimental cross-variogram. This specification examines the covariation of two types of data such as commercial and residential real estate returns. Both the variogram and cross-variogram are built utilizing the actual values of the covariation of real estate returns. Subsequent to establishing the experimental models a theoretical variogram is fit to the data, which can fill in any missing values on the continuous range of distance.

Also included in this section is the hedonic return specification, which models commercial returns based upon previously well-established variables that explain real estate returns. Subsequently, the residuals from the regression models are examined using the various variograms to obtain spatial correlation results after removing the common factors that can explain a portion of spatial correlation.

2.1 The Experimental Variogram

The portion of geostatistics employed for this study focuses on the continuous nature of a series of return observations y_i, i =1,...,N, over a national study region R, the contiguous United States. The returns are assumed to be observations on a spatial stochastic process $\{Y(s),\ s \in \mathbb{R}\}$, which varies in a spatially continuous manner over $\mathbb{R}$ and has been sampled at fixed points.

The main objective of this spatial data analysis is to infer the nature of spatial variation. The covariogram and correlogram are the statistics that compute the spatial dependence of deviations in return values from their mean. The covariogram measures the way in which the deviations of observations from their mean at different locations "co-vary", hence covariogram. Similarly, the correlogram measures the correlation between the observation deviations at different locations.

Bailey and Gatrell (1995) build a succinct mathematical foundation for the covariogram, usually termed just variogram, and correlogram. If there exists a spatially stochastic process $\{Y(s),\ s \in \mathbb{R}\}$ where the expected value of *Y(s)* is $\mu(s)$ and *VAR(Y(s))* is $\sigma^2(s)$ then the covariance of this process at any two points s_i and s_j is defined as

$$C(s_i, s_j = E\left[(Y(s_i) - \mu(s_i))(Y(s_j) - \mu(s_j))\right]$$

with the corresponding correlation defined as

$$\rho(s_i, s_j) = \frac{C(s_i, s_j)}{\sigma(s_i)\sigma(s_j)}.$$

A spatially stochastic process is considered stationary if $\mu(s) = \mu$ and $\sigma^2(s) = \sigma^2$. Stationarity implies that the distribution of the mean or variance is invariant under translation. Thus, for a spatially stochastic process, the stationary random function is homogeneous in space. And for any increment of distance, the distributions of the mean and variance, or any other moments, are independent of location and constant throughout R.

The distance between the two points s_i and s_j is the simple Euclidean geographic distance, denoted as h, a (N x 1) column vector. Typically, the location of any one point, s_i, is described by x_i and y_i coordinates as expressed in latitude and longitude measurements. Thus, two or more locations can be described by vectors of latitude and longitude values. Total unique combinations are a factor of $n(n-1)/2$.

Since stationarity is assumed, the covariance of the spatially stochastic process can be reduced to:

$$C(s_i, s_j) = C(s_i - s_j) = C(h).$$

Therefore, $C(s_i,s_j)$ depends only on the distance difference between s_i and s_j and not on the absolute locations. $C(h)$ is referred to as the covariance function or the covariogram and $\rho(h)$ as the corresponding correlation function or correlogram.

On both theoretical and practical application it is acceptable to weaken the hypothesis of strict stationarity. Referred to in spatial statistics as intrinsic stationarity, Matheron (1963, 1965) developed the hypothesis that assumes that the increments of the spatially stochastic process are weakly stationary. Intrinsic stationarity implies that the mean and variance exist and are independent of the location. Thus:

$$E\left[Y(s+h) - Y(s)\right] = 0$$

and

$$VAR\left[Y(s+h) - Y(s)\right] = 2\gamma(h).$$

The function $2\gamma(h)$ is termed a semi-variogram. The variogram is $\frac{1}{2}$ of the semi-variogram. For stationary and intrinsic attributes, the mean of $Y(s+h) - Y(s) = 0$ thus let $\gamma(h)$ be the mean square difference as defined by

$$\gamma(h) = \frac{1}{2}E[Y(s+h) - Y(s)]^2$$

and the method of moment estimator of an experimental variogram is

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Y(s+h) - Y(s)]^2,$$

where the summation is over all pairs of observed data points with a vector of separation h and *N(h)* is the number of difference pairs. Note that when the distance difference vector is null, theoretically the sample estimator is equal to 0. Additionally, the variogram is symmetric in h.

Although theoretically $\hat{\gamma}(0) = 0$, small-scale variability may cause sample values with small separations to be disparate. This causes a discontinuity at the origin of the experimental variogram, which is termed the *nugget*. A real estate example of the nugget is the covariance of two residential properties adjacent to each other. After controlling for the size of home, the year built and other hedonic explanatory variables, the homes could still exhibit a high spatial covariance because of their close proximity. That is, the property returns are correlated because the houses are in the same neighborhood, city and region even though the spatial difference h is approximately 0.

In the earth sciences, the actual distances of h may be quite uniform. For example, the variogram of an underground mineral deposit may be computed using readings from holes drilled beginning at the most southern and western point of the sample plot and continued every five yards in both a northerly and easterly direction. This would result in a uniform plot of sample readings along with a separation vector h that would have variogram or correlogram values for h=5, h=10, etc.

In applying the variogram to national real estate returns, which consists of irregularly spaced sample points, there will rarely be observations with an exact vector separation of h. Therefore, intervals or bins, known as lags, are created such that

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{h_{ij} \approx h}^{N(h)} [Y(s+h) - Y(s)]^2.$$

The average of the values that fit within the lag is used as the variogram value at distance h. To account for observations that are on the edge of lags, a lag tolerance is often employed. The use of a lag tolerance results in values that lie outside the lag distance being included in the variogram calculation. This produces a smoothing effect and eliminates some of the arbitrary judgment of how large to establish the lag. Therefore, if correlation computation requires each lag to set at 25 miles with a tolerance of 5 miles, the first variogram value will be computed based upon properties located between 0 and 30 miles from one another. The second variogram value will be computed based upon locations between 20 and 55 miles apart. The number of lags is adjusted to ensure 30 to 50 observations in each lag. The maximum tolerance is $\frac{1}{2}$ of the lag distance.

To compare dissimilarly-scaled variograms, the experimental variogram is standardized such that

$$\hat{\gamma}_s(h) = \frac{\hat{\gamma}(h)}{(\sigma_{-h})(\sigma_{+h})},$$

where $\hat{\gamma}(h)$ is the variogram from the separation vector h, and

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{h_{ij} \approx h}^{N(h)} Y^2(si) - m_{-h}^2, \ \textit{where } m_{-h} = \frac{1}{N(h)} \sum_{h_{ij} \approx h}^{N(h)} Y(s_i)$$

and

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{h_{ij} \approx h}^{N(h)} Y^2(s_i) - m_{+h}^2, \ \textit{where } m_{+h} = \frac{1}{N(h)} \sum_{h_{ij} \approx h}^{N(h)} Y(s_i).$$

Division by the standard deviation of the s_i and s_j values within each lag rescales the variogram by the lag variance. Again, the standardized variogram is symmetric in h. For an omnidirectional variogram (i.e., variogram without regard for direction between observations), the standardized variogram and correlogram are linked as $\hat{\rho}(h) = 1 - \hat{\gamma}(h)$.

2.2 The Experimental Cross-Variogram

An extension of the variogram and correlogram is the cross-variogram and cross-correlogram. Spatial statistical methods provide the possibility of mutual estimation of multiple interconnected data. The cross-variogram and cross-correlogram are utilized for the purpose of establishing mutual correlation between the interconnected data of residential and commercial real estate returns.

Similar to the aforementioned variogram, if $\{Y(s), s \in \mathbb{R}\}$ is the process relating to the first variable and $\{X(s), s \in \mathbb{R}\}$ is the process relating to the second variable, and both of these processes are assumed to be at least intrinsically stationary, then the cross-variogram is defined as:

$$C_{YX}(h) = E\left[(Y(s+h) - \mu_Y)(X(s+h) - \mu_X)\right],$$

where h is an arbitrary vector separation in $\mathbb{R}$. We can extend the same mathematical derivation above to a cross-variogram defined as

$$2\gamma_{YX}(h) = E\left[(Y(s+h) - Y(s))(X(s+h) - X(s))\right].$$

The sample estimator of the cross-variogram, given n pairs of observations (y_i, x_i) at sample sites s_j is

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{s_i - s_j = h} (y_i - y_j)(x_i - x_j),$$

where the summation is over all pairs of observed data with a separation vector of h and $N(h)$ number of pairs. Again, as in the estimation of the variogram for common property type real estate returns, there will not be an exact separation vector h. Therefore, lag intervals are utilized with the possibility of also using lag tolerances. Lastly, once the cross-variogram is standardized similar to the variogram, the cross-correlogram is simply $\hat{\rho}(h) = 1 - \hat{\gamma}(h)$.

2.3 The Theoretical Variogram

Since the sample estimates of the variogram or correlogram are binned by lag intervals, the average lag value is attributed to the mean lag distance. If a dataset has many observations, the lag distance may be small thus creating a smooth variogram curve, which might resemble Figure 1. Rarely, however, is the experimental variogram smooth enough to provide a value at all the distances desired. A model that will enable a user to compute a variogram value for any possible separation vector is required; therefore the experimental variogram is fitted with the appropriate theoretical variogram.

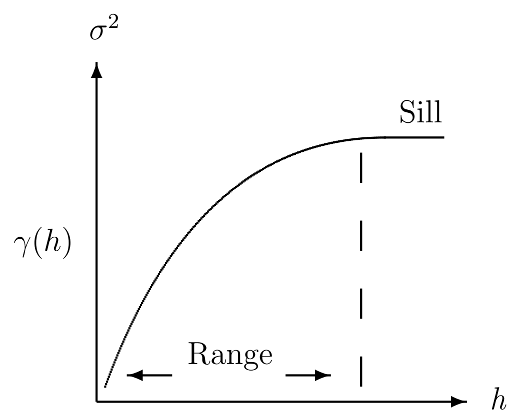

Figure 1: General Shape of a Bounded Variogram.

The three typical theoretical variograms are the spherical, exponential and Gaussian. The three models are similar in shape with differences as to how quickly the model reaches a plateau or *sill*. The distance at which the theoretical variogram reaches the sill is termed the *range*. These features are shown graphically in Figure 1. The spherical theoretical model has a finite range, while the exponential and Gaussian theoretical models asymptotically approach a limiting value. The spherical model is probably the most commonly used bounded theoretical variogram. Its equation is:

$$\gamma(h) = \begin{cases} 1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3, & \text{if } h \leq a; \\ 1, & \text{otherwise,} \end{cases}$$

where a is the range. It has a linear behavior at small separation distance and reaches the sill at a. The tangent at the origin reaches the sill at about $\frac{2}{3}$ of the range. Another commonly bounded theoretical model is the Gaussian. Its equations is:

$$\gamma(h) = 1 - exp\left(\frac{-3h^2}{a^2}\right),$$

where again a is the range. Unlike the spherical model, the Gaussian reaches its sill asymptotically thus the parameter a is defined as the practical range at which the variogram is 95% of the sill. The distinguishing feature of the gaussian model is its parabolic behavior near the origin.

After computing and potentially plotting the empirical variogram or cross-variogram models, it is necessary, then, to decide on a theoretical variogram. While there are methods of fitting semivariogram models, such as least squares and maximum likelihood as described by Cressie (1993), these techniques are not appropriate for data sets resulting in a small number of variogram points. Instead, a visual fit of the variogram points to a few standard models is often satisfactory. Even when there are sufficient variogram points, a visual check against a fitted theoretical model is appropriate as suggested by Hohn (1988).

Analysis shows that residential real estate results can be adequately modeled though a visual check. To account for standard error in the commercial data an Indicative Goodness of Fit (IGF) measure can be calculated to ensure the best theoretical variogram is modeled to the experimental variogram. The IGF is a number without units, which is standardized and is compared across diverse experimental variograms. An IGF value close to zero indicates a good fit. IFG is calculated by Pannatier (1996) as

$$IGF = \left(\frac{1}{N} \sum_{k=1}^{N} \sum_{i=0}^{n(k)} \frac{P(i)}{\sum_{j=0}^{n(k)} P(j)} \right) \left(\frac{D(k)}{d(i)} \right) \left(\frac{\gamma(i) - \hat{\gamma}(i)}{\sigma^2} \right)^2 ,$$

where

N	=	the number of directional variograms,
$n(k)$	=	the number of lags relative to variogram k,
$D(k)$	=	the maximum distance relative to variogram k,
$P(i)$	=	the number of pairs for lag i of variogram k,
$d(i)$	=	the mean pair distance for lag i of variogram k,
$\gamma(i)$	=	the experimental measure of spatial continuity for lag i,
$\hat{\gamma}(i)$	=	the modeled measure of spatial continuity for $d(i)$, and
σ^2	=	the variance of the data for the variogram.

2.4 Hedonic Return Specification

To address properties with different characteristics such as property type and size we need to adjust our returns. One means of doing this is to model commercial returns as a variable dependent upon other explanatory variables. The hedonic regression equation for real estate relates market return to a set of property and neighborhood value characteristics. Much research has been conducted as to the best explanatory variables for real estate price or return specifications. The return modeling for this study removes the main explanatory hedonic variables from the spatial consideration. The specification for commercial property is

$$R_{it} = \beta_{1t} ln(avesqft) + \beta_{1t} ln(numunits) + \beta_{2t} ln(mvlast) + \beta_{3t} apt$$
$$+ \beta_{4t} ind + \beta_{6t} office + \beta_{7t} retail + \beta_{8t} res + \beta_{9t-1} res$$

where

R_{it}	=	the rate of return on the ith property for the tth quarter,
avesqft	=	average square feet,
numunits	=	average number of units, which is used by some apartment complexes instead of the avesqft. measure,
mvlast	=	average market value from last $t-1$ quarter,
apt	=	dummy variable equal to 1 if the property is an apartment complex and 0 otherwise,
ind	=	dummy variable equal to 1 if the property is an industrial building and 0 otherwise,
office	=	dummy variable equal to 1 if the property is an office building and 0 otherwise,
retail	=	dummy variable equal to 1 if the property is a retail building and 0 otherwise, and
res	=	residential property return matched on the specific location of the commercial property.

Three points are worth noting regarding the commercial real estate return specification. First, *mvlast* is included in the regression specification as a potential proxy for size or quality. For example, lower priced properties might experience different returns than higher priced properties. Since the most immediate observation of *mvlast* is a lagged component of the dependent return variable, market value is averaged over the time period under study. This minimizes potential correlation between the dependent variable and the *mvlast* independent variable but maintains the identity of the independent variable. The second note regards the two other proxies (i.e., *avesqft* and *numunits*) for size. Both of these explanatory variables were also averaged over the study time period. The reasoning is that some size proxies changed significantly the last quarter of the study period. If the last quarter was the only measure utilized the proxy was not necessarily indicative of the value over the study period. Averaging smooths the specification for late or drastic changes. The third point concerns the fact each observations in the NCREIF data is one of four property types. As indicated, dichotomous dummy variables were utilized for the four types of property.

Because the four variables span all observations in the dataset the variance-covariance matrix is singular and will not invert. Therefore, the initial specification was executed including each dummy variable but without an intercept as indicated by the notation. This is to obtain clean coefficients of the four dummies and other independent variables without interpreting based upon one of the dummy variables. Subsequently, a second execution of the data was conducted to obtain the appropriate *adjusted* R^2 measure. During the second set of executions, the intercept was confirmed to the the same as the missing dummy variable.

3 Commercial Property

This section presents the findings of examining the spatial autocorrelation of commercial returns in isolation. Spatial results are provided for base commercial returns and residual returns after controlling for property type and proxies for size and quality. Also, a practical example is provided from a portfolio of commercial mortgage pass-through certificates.

3.1 Data

Variogram, correlogram, and regression models were computed from panel data utilizing contiguous U.S. zip code data for commercial property types. Commercial returns are based upon the National Council of Real Estate Investment Fiduciaries (NCREIF) continuous quarterly returns for apartment, industrial, office, and retail properties. Returns are computed from the second quarter of 2002 to the first quarter of 2003. Table 1 provides descriptive statistics of the data. Observations from Alaska and Hawaii were removed due to the extreme spatial discontinuity.

While the NCREIF database contains observations from 1978, there are few consecutive quarterly returns ending with the last quarter, 2003:1, which extend over

	Number of Obs.	Mean Return	Median Return	Standard Deviation
Apartment	17	5.38	4.04	8.76
Industrial	66	6.33	6.93	5.45
Office	60	-0.04	1.15	8.24
Retail	1	9.93	9.93	N/A

Table 1: Summary statistics of commercial returns calculated from NCREIF. Returns and standard deviation are in percent.

a lengthy time frame. Using the last sixteen quarters yields 46 observations, which is not enough to compute a quality variogram. Employing twelve quarters increases the observations to 72 but does not provide the detail needed for properties in close proximity to each other. Examining other time periods that do not end in the most recent quarter does not provide a greater number of observations because the latest observations are the most complete. Therefore, the time period that afforded the greatest benefit, especially the ability to examine nugget effects, is the last four quarters in the dataset. The 144 observations generate 10,296 unique combinations for variogram computation. The location of each of the 144 observations are mapped as Figure 7, which appears in the appendix.

3.2 Spatial Autocorrelation of Base Returns

The returns of each of the 144 commercial properties were contrasted against every other return using the correlogram measures and subsequently binned by common distances and fitted with a theoretical correlogram. Figure 2 plots the detail of the experimental and theoretical correlogram values for the pertinent range of the

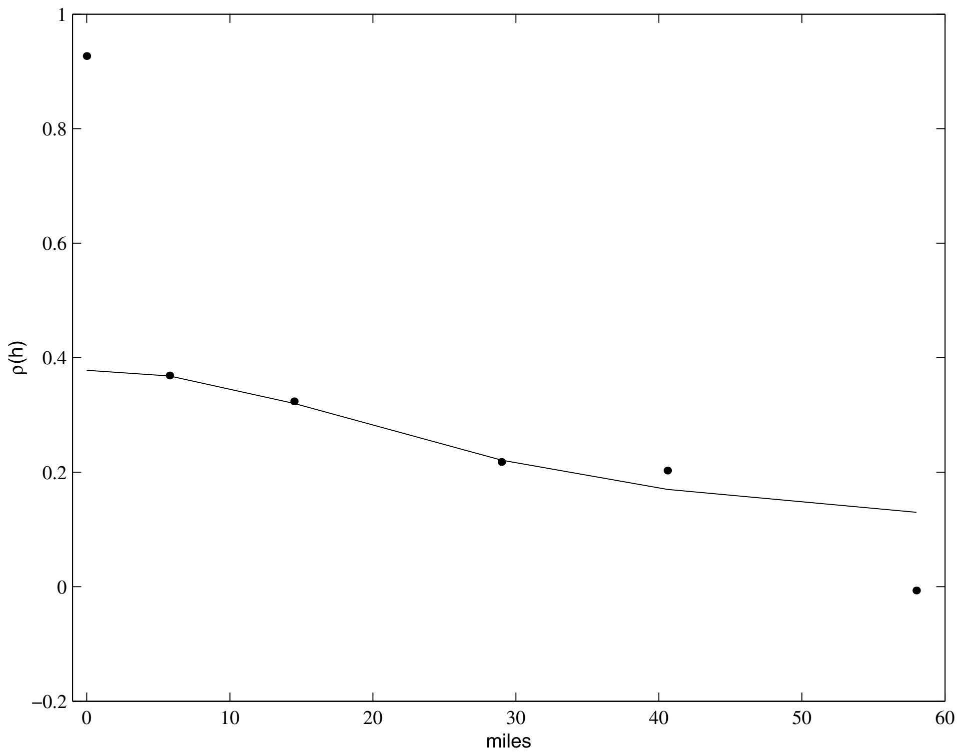

Figure 2: Correlogram of U.S. commercial property returns.

NCREIF commercial property returns.[2]

The first notable result shown by the diagram is that the maximum relevant distance difference for commercial property is approximately 58 miles. This is the point at which the experimental correlation between properties decays to zero. After 58 miles, the empirical correlogram values become divergent. The measure of 58 miles is a fundamental result of the commercial analysis as four out of the five commercial outcomes found 58 miles to be the point at which spatial autocorrelation of returns decays to zero.

Excluding the nugget factor for the moment, the second lag distance bin result indicates that the correlation is about 0.37 for a distance of 5.8 miles. Thereafter,

[2] The Guassian theoretical correlogram, which is the best fit of the three bounded theoretical models, is based upon a nugget = 0.64, sill = 0.90, and range = 58 miles. The IGF measure = .0017338. The lag distance is 14.5 miles with 50% lag tolerance.

Approximate Lag Distance	**Lag Count**	**Experimental Correlogram**	**Theoretical Correlogram**
0.0 miles	310	0.93	0.36
5.8 miles	114	0.37	0.36
14.5 miles	368	0.32	0.34
29.0 miles	310	0.22	0.30
40.6 miles	168	0.21	0.26
58.0 miles	94	-0.01	0.20

Table 2: Correlations calculated from spatial correlogram of 2002:2 – 2003:1 NCREIF commercial returns computed for conterminous U.S. zip codes.

the correlation reduces monotonically to about 0.22 at 14.5 miles and between 0.21 and 0.22 for distance from 14.5 to 40.6 miles. At this point the experimental correlogram value falls to zero. As shown by the theoretical correlogram line, the actual correlations are difficult to theoretically modeled. The distance is quite short and the Gaussian model, which is the best model based upon the IGF, does not come close to meeting the initial nugget effect and does not model the decay to zero correlation well. Table 2 presents the computed empirical and theoretical values at the various lag distances.

A significant feature of the commercial data is that there exist different types of property within some of the zip codes. Therefore, the nugget or correlation at $h \simeq 0$ can be computed using data that truly are at a zero distance difference. The notable result for these 310 correlogram values is that the experimental correlation is approximately 0.93. The theoretical model does not allow for this large correlation and still fit the subsequent correlogram values; it is computed as 0.36. Nevertheless, the data exhibit that within the same or nearby zip codes (lag tolerance of 50%)

that the spatially correlation is extremely high. Therefore, if a real estate investment manager is considering the purchase of multiple real estate properties in a centralized location, he or she should consider that the price of the properties move quite similarly.

A particular example of the implications is found in a $1.4 billion portfolio of commercial mortgage pass-through certificates offered by Banc of America Securities LLC and Bear, Stearns and Co. Inc. Dated October 2003, the offering held $32 million in six multifamily properties in two cities in South Carolina. All of the properties are within 30 miles of each other. Thus, that portion of the portfolio exhibits 20-30% correlation among the properties. Similarly, there are two office complexes in each of the same zip codes in Irving and Austin, Texas and Atlanta, Georgia. Additionally, three office complexes are located within two nearby Los Angeles zip codes. The spatial autocorrelation between the properties in each respective city may be as high as 0.93 because the properties are of the same type. In practical terms, these multiple properties act more like a single property than a diversified portion of the portfolio.

3.3 Regression Models

Computed from the original base returns, commercial property can be highly spatially correlated over a somewhat short distance. Intuitively, there should exist factors that assist in explaining a portion of spatial autocorrelation. The spatial autocorrelation witnessed in the base return correlogram may be due to factors such as property type, size, or quality. To address the issue of the effect of other variables on commercial returns, we estimate multiple hedonic regression models, which are presented in Table 3.

The results of the commercial property regression models indicate that size as proxied by average square feet is not significant, however size proxied by average number of units, which is an apartment metric, has more explanatory power. Average

	Model					
Variable	**A**	**B**	**C**	**D**	**E**	**F**
Average Mkt. Value Last Quarter	1.84 (2.37)	1.88 (2.45)		2.07 (2.68)	2.06 (2.69)	
Average Square Feet	0.15 (0.41)		0.27 (0.73)	-0.03 (-0.09)		
Average Number of Units	1.22 (1.66)	1.13 (1.62)	1.52 (2.07)			
Apartment	-35.00 (-2.44)	-33.90 (-2.42)	-3.97 (-0.68)	-31.88 (-2.23)	-32.07 (-2.28)	5.58 (3.25)
Industrial	-29.34 (-2.06)	-28.30 (-2.02)	2.57 (0.55)	-31.23 (-2.18)	-31.49 (-2.26)	5.94 (6.57)
Office	-36.86 (-2.45)	-35.87 (-2.43)	-3.03 (-0.65)	-38.98 (-2.59)	-39.24 (-2.67)	0.32 (0.34)
Retail	-27.03 (-1.64)	-26.01 (-1.60)	6.52 (0.76)	-29.09 (-1.76)	-29.35 (-1.81)	9.93 (1.36)
Adjusted R^2	0.15	0.16	0.13	0.14	0.15	0.11

Table 3: Hedonic regression models of commercial property returns. Returns computed for U.S. zip codes as provided by NCREIF. *T*-statistics are in parentheses.

market value last quarter may be another factor for size or could be a proxy for quality. In every model utilizing average market value last quarter finds that the variable has significant power in explaining the dependent variable of commercial return. Thus, the residuals from the regression using average market value, average number of units, and the dummy variables that control for type of property are used to recreate new correlogram values, which is Model B.

Approximate Lag Distance	**Lag Count**	**Experimental Correlogram**	**Theoretical Correlogram**
0.0 miles	42	0.42	0.31
2.9 miles	136	0.31	0.31
14.5 miles	368	0.26	0.29
29.0 miles	310	0.17	0.25
40.6 miles	168	0.15	0.21
58.0 miles	94	0.00	0.11

Table 4: Correlations calculated from spatial correlogram of 2002:2 – 2003:1 NCREIF commercial returns computed for U.S. zip codes.

3.4 Spatial Autocorrelation of Residual Returns

The correlation values were recalculated utilizing the residual returns. The lag distance and tolerance are the same as the base commercial returns with the exception of the nugget, which is computed with a lag distance of 5.8 miles to facilitate a closer examination of characteristics of commercial properties within the same or nearest zip code.

The results presented in Table 4 show that the residual correlation values are reduced, although the distance values remain unchanged. The distance at which full correlation distance is obtained is the same 58 mile measure as with the base returns. Beyond the initial nugget, yet within the first 6 miles, the correlation reduces from 0.37 for the base returns to 0.31 using the residual returns. Similar results are seen at subsequent lag distances. The correlations are reduced by 20% to 25% for the same distances of the base returns. This is depicted graphically in Figure 3.

Again, the notable uniqueness of the nugget effect persists with the residual returns. Whereas the experimental correlation was 0.93 for the base returns, the resid-

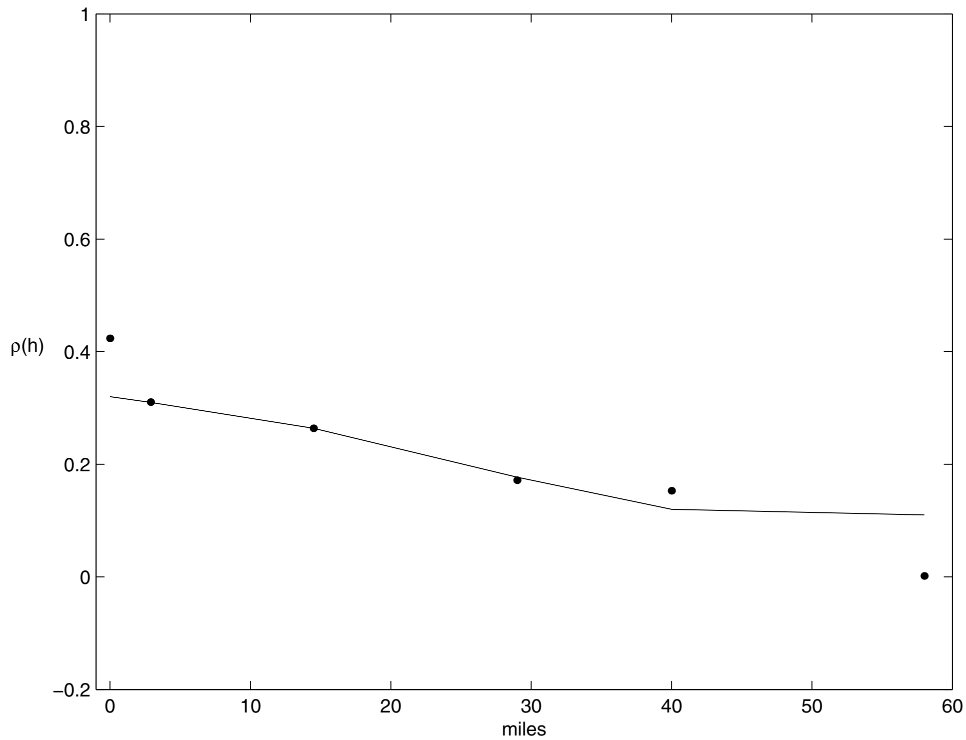

Figure 3: Correlogram of NCREIF residual returns. Experimental and theoretical correlogram values of U.S. commercial property residual returns from 2002:2 to 2003:1.

ual nugget is approximately 0.42. The Guassian theoretical correlogram displays a better fit than the original model computed from the base returns.[3] However, the theoretical correlogram still cannot fit the distinct values displayed by the actual commercial return residuals.

The residual results indicate that an investment manager can significantly reduce the correlation within a portfolio by diversifying among different types and sizes of properties within the same zip code or neighboring zip codes. However, after diversifying by property size, quality and type, a portfolio may still realize 42% correlation between returns within the same or adjacent zip code. Likewise, an investment professional could hold a office building in a northeast suburb of Washington, D.C. and

[3] Theoretical correlogram based upon nugget = 0.69, sill = 0.94, and range = 58 miles. IGF = 0.00083467.

an apartment complex in downtown Baltimore, Maryland and still incur a 20% correlation between the two types of property just based on proximity. Thus, to remove spatial correlation means to move beyond the 58 mile measure. The residual results indicate that 58 miles is effective in diversifying the spatial correlation between different types of properties. Moreover, if an investment professional, who may desire to specialize in one type of property due to portfolio objective, simplicity, or expertise can still obtain diversification by concentrating on properties that are at least 58 miles apart.

4 Residential Property

This section presents the results of examining residential property for spatial autocorrelation over the conterminous United States. The outcomes of this section are utilized in the next section to further explain commercial spatial autocorrelation.

4.1 Data

Residential returns are based upon the aggregate housing values obtained from the Neighborhood Change Database (NCDB), which is based upon the U.S. Bureau of the Census decennial censuses for 1970, 1980, and 1990 (2000 data did not became available until late 2003). Each year is standardized to the base year of 2000. Figure 8 in the appendix maps the locations of the residential property sample points. The paramount feature of the NCDB is that there exists data for 9,699 national zip codes. Since each sample point is compared spatially to all other sample points, the various residential models were computed using over 40 million combinations each.

4.2 Base Results

Conterminous U.S. residential property offers a great number of observations and a particular well-specified example of a bounded variogram. This feature can be utilized in combination with commercial property but first some findings concerning residential only. Figure 4 plots the standardized experimental and theoretical variograms for contiguous U.S. residential property for the first decade of study from 1970 to 1980. Since the variogram in Figure 4 is omnidirectional and standardized, the correlogram is simply *1 - variogram*[4]. Since the variogram is well-specified, lag tolerances are not employed.

[4]The theoretical variogram is spherical and based upon a nugget = 0.57, sill = 1.14, and range = 1,767 miles.

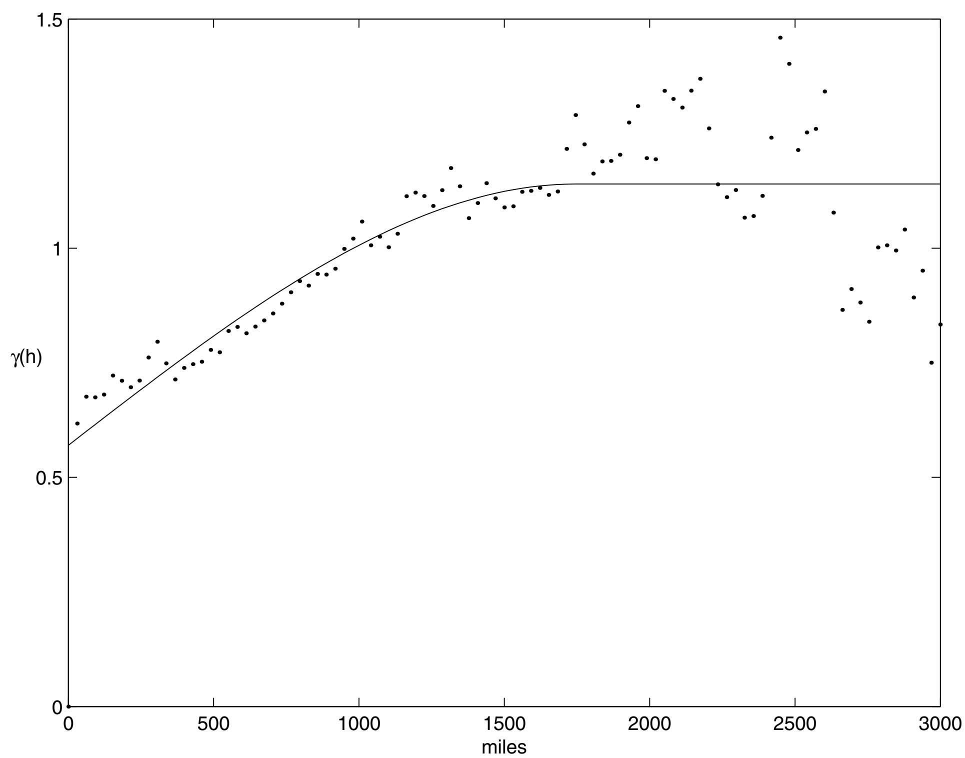

Figure 4: Standardized experimental and theoretical variograms of conterminous U.S. residential property from 1970 to 1980.

The experimental variogram increases with distance up to an approximate difference distance of 1,745 miles. After h reaches the sill, the variogram becomes divergent. This coincides with the variogram criterion that the maximum distance of reliability is $h <$ total distance/2.

The specific values for decade combinations of 1970 to 1980, 1980 to 1990, and 1970 to 1990 at various distances are described in Table 5. All decades of residential data are best fit by a spherical theoretic model and it is the theoretic models that provide correlation factors on a continuous spectrum. Specific increments are listed in Table 5, however, any distance difference can be computed based upon the spherical correlogram for the specific nugget, range, and sill.

The results indicate a distinct difference between the decades of 1970-80 and 1980-90. The first decade has a lower correlation among residential properties within the same zip code (i.e., approximately 0.40) while the second decade has a higher correlation value of approximately 0.60. Additionally, the rate of correlation decay differs with the first decade decaying to zero at approximately 950 miles while the second decade declines quickly to zero at approximately 300 miles.

The combination of the two decades presented in Panel C indicates that similar to the first decade, the nugget effect is approximately 0.40 and the rate of decay extends to zero at approximately 1000 miles. Interestingly, based upon over half a million observations, residential property extending as far as 150 miles exhibited 0.35 correlation over the 20 year period. This implies that a homeowner in Cleveland, Ohio should consider not just her own local market but one as far away as Pittsburgh, Pennsylvania as that market will have spatial impact on her home price.

Approximate Lag Distance	Lag Count	Experimental Correlogram	Theoretical Correlogram
Panel A: 1970 - 1980			
30.61 miles	345,901	0.38	0.42
153.05 miles	513,825	0.28	0.36
306.10 miles	709,321	0.20	0.28
459.15 miles	832,363	0.24	0.21
612.20 miles	816,794	0.19	0.14
765.25 miles	670,650	0.10	0.08
948.91 miles	671,954	0.00	0.01
Panel B: 1980 - 1990			
15.31 miles	130,541	0.57	0.62
61.22 miles	219,652	0.54	0.51
122.44 miles	239,948	0.42	0.38
183.66 miles	274,317	0.37	0.25
244.88 miles	332,306	0.20	0.14
275.49 miles	338,173	0.13	0.09
306.10 miles	360,056	0.01	0.04
Panel C: 1970 - 1990			
30.61 miles	345,901	0.35	0.40
153.05 miles	513,825	0.35	0.33
306.10 miles	709,321	0.19	0.26
459.15 miles	832,363	0.14	0.19
612.20 miles	816,794	0.12	0.12
765.25 miles	670,650	0.09	0.07
918.30 miles	652,836	0.06	0.02
1,010.13 miles	722,171	0.01	0.00

Table 5: Correlogram of residential returns.

4.3 Regression Models

Similar to commercial returns, it is reasonable to expect that there exist factors that explain a portion of the spatial autocorrelation. For example, higher incomes in a neighborhood can provide more amenities, which in turn could produce higher, correlated returns. Thus, the base returns for each of the three combination of decades were regressed by independent variables that explain a portion of residential returns. The specification for the residential property is

$$\begin{aligned} R_i = \beta_0 \; &+ \; \beta_1 ln(educ12) \; + \; \beta_2 ln(educ16) \; + \; \beta_3 ln(educpp) \\ &+ \; \beta_4 ln(income) \; + \; \beta_5 ln(size) \; + \; \beta_6 ln(bltoc70) \\ &+ \; \beta_7 ln(bltoc59) \; + \; \beta_8 ln(bltoc49) \; + \; \beta_9 ln(bltoc39), \end{aligned}$$

where

R_i	=	the rate of return on the ith house,
educ12	=	persons 25 years old or older who completed high school,
educ16	=	persons 25 years old or older who completed college,
educpp	=	number of persons 25 years old or older,
income	=	average family income,
size	=	aggregate number of rooms in the home, and
bltoc	=	total occupied housing units built up to the year specified in the variable.

Pace, Barry and Sirmans (1998) show that the number of variables needed to remove all local variation can grow quickly. Therefore, the main variables of education, family income, proxy for house size, number of persons of age to buy a home, and age of home are modeled *ex ante* and *ex post* relative to the decade modeled. That is, the decade from 1970 to 1980 is modeled using 1970 and 1980 explanatory variables and likewise for 1980 to 1990 and 1970 to 1990. As expected, almost all the variables are significant at the 1% level. (The coefficients and t-statistics are provided in the appendix as Tables 10, 11, and 12.)

4.4 Residential Residual Autocorrelation

The spatial autocorrelation values for each residential time frame were recomputed using the residual returns. The residuals are not averaged across zip codes but left within census tracts since each observation is the remaining portion not explained by the hedonic model and thus is unique information. This produces slightly less than 1 billion observations for each time frame. (As a particular example, Figure 9 in the appendix depicts graphically the full-range results of removing conspicuous explanatory variables for the long-term time frame of 1970 to 1990. The shorter, relevant range of correlograms for other time frames are provided in Figures 10 - 13 in the appendix.)

Table 6 provides the details for each of the three residential combinations. The residual returns from 1970 to 1980 yield basically the same results utilizing either 1970 or 1980 explanatory variables. The results demonstrate that the initial correlation is reduced from approximately 0.40 based upon base returns to 0.15 for the residuals. Additional reduction occurs in the lag distance. Whereas previously the correlation was approximately .020-0.28 at 300 miles, the residual returns decay to insignificance by 300 miles.

Similar results are seen for the decade from 1980 to 1990 but two differences are worth noting. First, there is a major shift in the nugget effect. Instead of approximately 0.60 correlation at $h \simeq 0$, the nugget is 0.20 based upon the Gaussian theoretical values, which is the best fitting theoretical correlogram for all the residential residual return experimental correlograms. This value might even be slightly less based upon the 0.13 experimental value. Second, there is not much reduction in the lag difference between the two correlograms. The base returns exhibit a zero correlation at approximately 300 miles. The residual returns indicate zero correlation

Approximate Lag Distance	Lag Count	Experimental Correlogram	Theoretical Correlogram
Panel A: 1970 - 1980 using 1970 RHS variables			
15.31 miles	4,777,264	0.14	0.15
30.61 miles	6,192,038	0.13	0.15
76.53 miles	4,175,299	0.13	0.13
153.05 miles	4,883,237	0.09	0.09
229.58 miles	5,898,320	0.08	0.04
306.10 miles	6,478,281	0.04	-0.01
Panel B: 1980 - 1990 using 1990 RHS variables			
30.61 miles	7,156,825	0.13	0.20
61.22 miles	5,528,380	0.20	0.19
153.05 miles	6,429,092	0.22	0.13
229.58 miles	7,637,638	0.22	0.07
244.88 miles	8,104,028	0.11	0.06
260.19 miles	9,058,917	-0.03	0.05
Panel C: 1970 - 1990 using 1990 RHS variables			
15.31 miles	4,783,788	0.19	0.25
30.61 miles	6,198,538	0.17	0.25
76.53 miles	4,178,956	0.22	0.23
153.05 miles	4,888,290	0.22	0.19
229.58 miles	5,898,176	0.20	0.13
306.10 miles	6,483,273	0.07	0.08
382.63 miles	7,302,486	0.09	0.03
428.54 miles	7,112,267	0.03	0.01

Table 6: Residential correlogram values of residual returns.

at approximately 260 miles. Whereas the prior decade witnessed a reduction of about 66% in lag distance, the decade from 1980 to 1990 saw a reduction of 15%. This is probably in part due to the steep decline in correlation within the decade from 1980 to 1990.

After combining the two decades, the residual returns exhibit a nugget effect of approximately 0.20. The initial correlogram values are less than values at slightly longer lag distances starting at 0.19 at 15.31 miles, moving down to 0.17 at 30.61 miles and then up to approximately 0.22 at 76.53 and 153.05 miles and 0.20 at 229.58 miles. After this distance the correlation drops expeditiously.

These results imply that the residential markets in Detroit, Toledo, Columbus, Cleveland, and Pittsburgh have a degree of spatial autocorrelation. Each of these cities is within 230 miles of one another, therefore, based upon almost 26 million residual observations, the long-term results from 1970 to 1990 indicate a spatial correlation of approximately 20% after controlling for conspicuous hedonic factors.

5 Combining Commercial with Residential

The spatial autocorrelation of residential returns may be able to add detail or confirmation to the commercial-only results. We examine two methods for applying residential findings to commercial returns. The initial analysis examines the effects of including residential returns as an explanatory variable in a regression model of commercial returns. Subsequently, spatial correlation values for commercial returns are computed using the residuals that include residential returns. The second analysis is to compute a cross-correlogram utilizing both sets of data. The cross-correlogram is well suited for this specific analysis.

5.1 Data

Under both methods of analysis, the commercial data are matched with residential returns at the same specific location. Since census data (i.e., residential returns) are not available for the quarters required for commercial property analysis, the residential data for this portion of the study are based upon Metropolitan Statistical Areas (MSAs) as provided by the Office of Federal Housing Enterprise Oversight (OFHEO). The OFHEO data set provides quarter residential housing price data by MSAs, which were matched to the 144 zip code observations for the NCREIF data.

5.2 Commercial Autocorrelation with Residential Residuals

Clearly, commercial and residential markets possess similar economic considerations at common locations. Unemployment and the scarcity of land suitable for development are examples of factors that can affect all types of real estate within a sufficiently small geographic boundary. To understand the combined affect of both general property types, lagged residential returns are included as an independent variable in the commercial return regression specification.

Lagged residential returns are used because the specification for commercial returns explained in part by residential returns in the same quarter may suffer from statistical simultaneity. Furthermore, from a practical standpoint, a commercial real estate investment professional cannot utilize an *ex post* observation of residential returns unless the observation is *ex ante* to the next quarterly commercial return. To address both of these issues, the regression models were executed with the residential returns lagged one quarter.

When residential returns are matched by location and included as an explanatory variable, the regression models remove noise as measured by *Adjusted* R^2. Table 7 presents specific coefficient values with t-statistics in parentheses. The adjusted coef-

Variable	Model A	B	C	D	E
Lagged Residential Return	0.69	0.67	0.70	0.69	0.78
	(3.02)	(2.98)	(3.05)	(3.06)	(3.48)
Average Market Value Last Quarter	1.26	1.33	1.47	1.48	
	(1.68)	(1.80)	(1.97)	(2.01)	
Average Square Feet	0.22		0.05		
	(0.64)		(0.15)		
Average Number of Units	1.17	1.04			
	(1.69)	(1.58)			
Apartment	-28.78	-27.22	-25.73	-25.42	1.14
	(-2.12)	(-2.04)	(-1.89)	(-1.90)	(0.56)
Industrial	-23.21	-21.72	-24.98	-24.55	1.90
	(-1.72)	(-1.63)	(-1.84)	(-1.85)	(1.32)
Office	-29.65	-28.25	-31.64	-31.23	-3.22
	(-2.08)	(-2.01)	(-2.21)	(-2.23)	(-2.25)
Retail	-19.09	-17.66	-21.00	-20.59	7.37
	(-1.22)	(-1.14)	(-1.33)	(-1.33)	(1.08)
Adjusted R^2	0.19	0.20	0.18	0.19	0.17

Table 7: Hedonic regression models of commercial property returns. Regression of theoretical variables that explain commercial real estate returns from 2002:2 to 2003:1, including lagged residential returns specific to location. *T*-statistics are in parentheses.

ficient of determination of each model increases from 0.11-0.15 to 0.17-0.20 over the commercial models that do not include residential returns. Interestingly, the coefficient of residential returns is fairly consistent in models A-D at between 0.67 and 0.70. The residential return coefficient for model E is higher at 0.78 however, this model is inconsistent with the other models with regards to the sign and magnitude of the dummy variables for property type.

Approximate Lag Distance	Lag Count	Experimental Correlogram	Theoretical Correlogram
0.0 miles	308	0.89	0.30
5.8 miles	114	0.30	0.28
14.5 miles	364	0.12	0.19
29.0 miles	308	0.07	0.02
40.6 miles	168	0.12	-0.03
58.0 miles	92	-0.11	-0.05

Table 8: Correlations of commercial returns computed for conterminous U.S. based upon residuals after controlling for lagged residential returns matched on specific location.

Maintaining the same model as previously utilized for the commercial-only correlogram,[5] Table 8 reports the results of modeling the remaining spatial autocorrelation. The geographic dimension is zip codes for commercial property and MSAs for residential property. Consistent with prior commercial results, the distance where the spatial autocorrelation decays to zero is reached at 58 miles. This range is plotted in Figure 5. In general, the experimental and theoretical correlograms exhibit much the same shape and values as the commercial-only correlogram[6].

One striking result is the correlation of property with the same or adjacent zip code. Recall that the correlation decayed from 0.93 for the base commercial returns to 0.41 for residual returns. When residential property is added to the regression model the experimental correlogram value for commercial property within the same or neighboring zip code is 0.89. This is quite similar to the correlation of base commercial returns without accounting for size, quality and type of property. Thus, after controlling for these variables and reducing noise in the model, the level of spatial

[5] Table 3, model B

[6] Theoretical correlogram is Guassian based upon nugget = 0.70, sill = 1.05, and range = 23.2 miles.

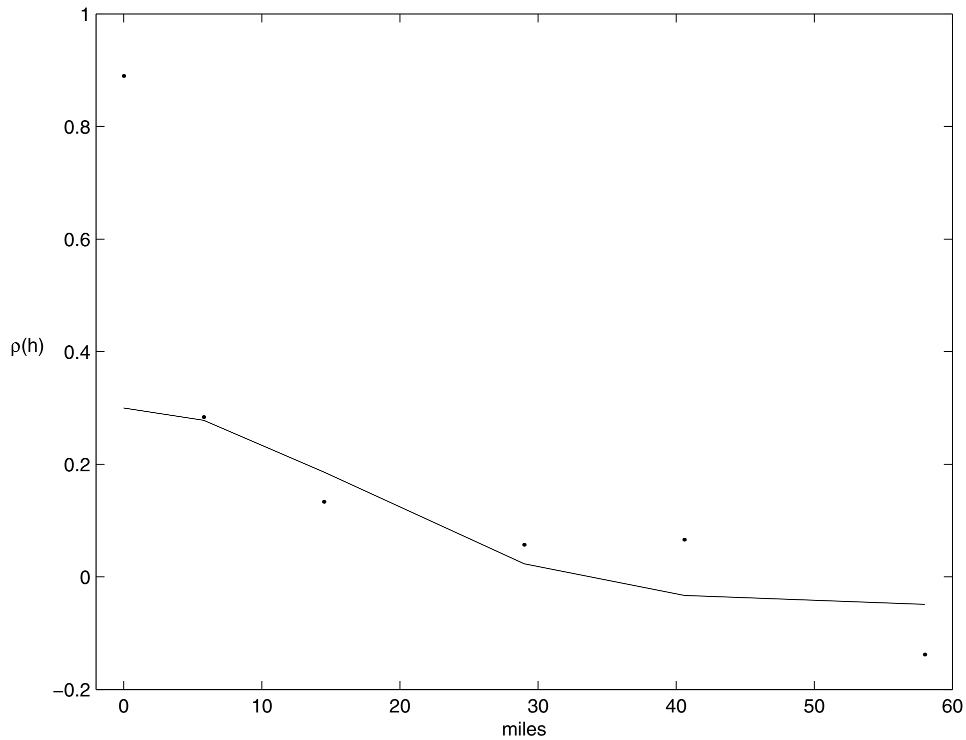

Figure 5: Correlogram values of conterminous U.S. commercial property residual returns from 2002:2 to 2003:1 after controlling for lagged residential returns matched to specific location.

autocorrelation within the same or adjacent zip code is higher than without including residential returns. One justification of this result might be that each residential return is based upon an MSA, which could span many miles. However, MSA could easily encompass the entire lag distance of 58 miles. Therefore, it seems peculiar that all the results are not affected by the MSA distance factor. So a question remains as to why, after controlling for property type and other potential property factors, does the initial nugget exhibit such a high degree of spatial autocorrelation. Since including residential return increases the power of the test and thus removes some of the specification noise, the result implies that controlling for property type and proxies for size and quality is not as beneficial as the initial residuals may indicate.

Revisiting the commercial mortgage pass-through certificate portfolio example we see that two office complexes in Atlanta are in the same zip code, a hotel is in a neighboring zip code and another office building is within ten miles. The results imply that the best case scenario is that these properties exhibit 0.12 correlation based upon the full specification, which controls for lagged residential property and the commercial explanatory variables. However, again based upon the specification with lagged residential property, the acctual correlation may be much higher - up to the 0.89 finding. Additional spatial problems can be noted in the portfolio between different property types in the Tidewater area of Virginia and in Irving, Texas.

5.3 Cross-Correlogram

The second examination of the residential and commercial return combination is with the use of the cross-correlogram. This statistical tool can offer additional insight into the spatial autocorrelation because it is comparing both types of real estate returns based upon distance separation. Therefore, it can be utilized as a robustness check for all types of property in this study. The returns are computed from 2002:2 to 2003:1 for both types of property, which were matched by specific location. Again, commercial property is based upon zip codes and residential property is based upon MSAs. The theoretical correlogram is Guassian.

Table 9 reports the findings of matching the 144 returns for commercial and residential property. It appears that the cross-correlogram exhibits a portion of both of the residential- and commercial-only correlograms. First, the nugget effect is not as pronounced as the commercial-only correlograms but slightly more than the residential-only outcomes. Second, the distance of reliability is increased, which could be the residential influence of autocorrelation at hundreds of miles. All previous commercial tests found 58 miles to the be distance where correlation decayed to zero. Last, the rate of decay appears to be a compromise between the fast rate of commer-

Approximate Lag Distance	**Lag Count**	**Experimental Correlogram**	**Theoretical Correlogram**
0.0 miles	136	0.50	0.50
14.5 miles	368	0.50	0.48
29.0 miles	312	0.41	0.42
43.5 miles	166	0.38	0.33
58.0 miles	94	0.37	0.24
72.5 miles	44	0.42	0.15
87.0 miles	20	0.41	0.08
101.5 miles	50	-0.02	0.03

Table 9: Cross-correlogram values of commercial and residential returns.

cial property and the long distances of the residential property.

The final note is the precipitous drop of the cross-correlogram values at approximately 100 miles. This can be seen in Figure 6. From 0 to 100 miles the experimental correlogram values calculate between 0.50 and 0.37. However, after approximately 87.0 miles the experimental correlogram value decays from 0.41 to zero rapidly. The theoretical value is much less between distances of approximately 58 to 87 miles[7]. As with all the comparison in this study between the experimental and theoretical values, the theoretical model errs on the side of being conservative.

6 Summary and Policy Implications

Intuitively, it seems reasonable that adjacent commercial property will demonstrate returns that move in the same direction and quite possibly express a common variance or correlation. Prior literature informs us that type and geographical or economic re-

[7]Theoretical correlogram is Guassian based upon nugget = 0.50, sill = 1.05, and range = 72.5 miles.

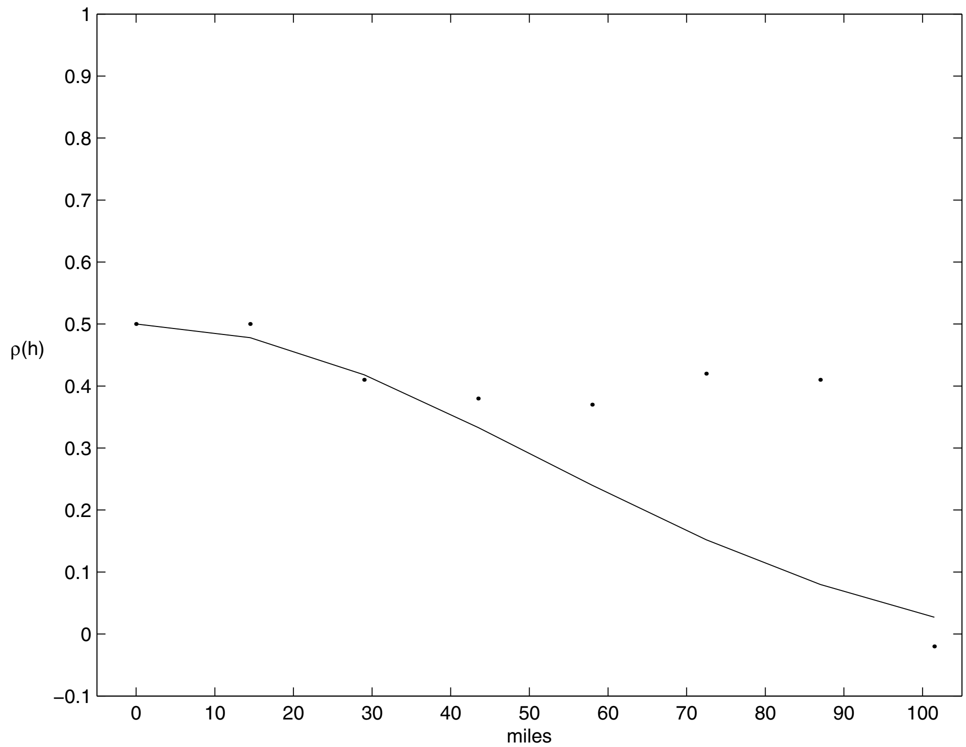

Figure 6: Experimental and theoretical cross-correlogram values of conterminous U.S. commercial and residential property returns from 2002:2 to 2003:1. Returns based upon commercial zip codes and residential MSAs.

gions influence prices of commercial property. Similarly, it has been established that residential property is influenced by proximity to amenities such as central business districts, quality schools and mass transportation stations. However, this top-down approach has not formulated an uninterrupted scale by which we can predict spatial correlation at arbitrary distances. Likewise, the quantitative relation between distance and correlation has not been modeled so that real estate investors can determine the separation distance required to effectively remove spatial correlation among properties. In addition to identifying the distance of zero correlation, this study creates a framework for assessing correlation among properties within a small geography boundary as well as measuring the rate of correlation decay over the continuous range.

Commercial property exhibits a relatively short distance scale. Almost all commercial results show that correlation decays to zero at approximately 58 miles. After controlling for type of property and proxies for size and quality, the residual results indicate that property beyond the same or adjacent zip codes exhibits about 0.31 correlation. This degenerates to approximately 0.26 at 15 miles and about 0.15 for distances between 30-40 miles.

A foremost result is the correlation between different property types located within the same or adjacent zip codes. The results show that a commercial property portfolio manager will incur 0.42 correlation between, for example, an industrial building that is within a couple of miles of an apartment complex. The implication is that although of different type and thereby diversified in one respect, further spatial diversification is required because the market is not going to reward a real estate manager for the additional variance under MPT.

The long-term residential results indicate that after controlling for common hedonic factors the initial correlation is approximately 0.20 for distances up to 230 miles. After which distance the spatial autocorrelation of housing returns decays to zero at approximately 425 miles. The approximately 0.20 correlation translates into homes between Boston and New York, Chicago and Indianapolis or Dallas and Oklahoma City experiencing a meaningful correlation with each other, respectively.

The suggestion of the residential-only results is that policy makers and real estate professionals in one city must be aware of decisions and economic changes in neighboring cities, and the definition of a neighbor may be greater than realized. For example in Washington D.C. base residential returns will be influenced up to 30-40% by policy decisions or economic conditions in New York, Philadelphia, Dover, Annapolis, Baltimore, Richmond, Virginia Beach, Hampton and/or Norfolk. And while this study does not classify all hedonic factors that affect spatial autocorrelation, a change in property taxes, for example, in Richmond, Baltimore, Dover, and/or Annapolis will have an approximate impact of 0.20 on D.C. residential returns since property taxes have an impact on household income, which was controlled for in this study.

The combined results of residential and commercial returns provide two manner of advancements. As a robustness check the cross-correlogram confirms significant spatial autocorrelation over a lag distance of approximately 0.87 miles. Autocorrelation ranges from 0.37 to 0.50 over that distance utilizing all types of real estate considered in this study. Moreover, modeling commercial returns with the inclusion of residential returns implies that spatial autocorrelation is higher than exhibited by residuals that do not include residential returns. Therefore, as noise is removed more spatial autocorrelation is realized. Commercial property correlation within the same or adjacent zip codes is a factor similar to the correlation of commercial property without controlling for property type or proxies of size and quality. This result is based upon lagged residential returns to account for simultaneity and to provide the added benefit of offering a commercial real estate professional with a potential forecast of commercial property returns for the next quarter.

One final implication of the entire study is based upon the lag distance required for the spatial autocorrelation to decay to zero. The commercial property returns are metropolitan based whereas the residential returns are nationwide, comprising both urban and rural property. The various forms of the commercial returns uniformly resulted in a distance of decay of 58 miles. However, the residential returns exhibited a much longer distance of decay. Coincidentally, the combined returns provided by the cross-correlogram expanded the commercial distance of decay. The explanation might be that the residential property results are measuring the large expanses of land in rural areas. The implication is that the commercial distance of decay is of some magnitude greater than the 58 miles. This is a further empirical question that may be addressed by combining all the residential zip codes returns against the commercial returns (i.e., pseudo cross-correlogram) or may be better analyzed as a greater quantity of commercial data becomes available.

References

Bailey, T.C. and A.C. Gatrell, 1995, *Interactive Spatial Data Analysis,* Longman Group, Essex.

Cheng, P. and R.T. Black, 1998, "Geographic Diversification and Economic Fundamentals in Apartment Markets: A Demand Perspective," *Journal of Real Estate Portfolio Management* 4:2, 93-105.

Cheng, P. and Y. Liang, 2000, "Optimal Diversification: Is It Really Worthwhile?" *Journal of Real Estate Portfolio Management* 6:1, 7-16.

Cressie, N.A., 1993, *Statistics for Spatial Data*, John Wiley and Sons, New York.

Dubin, R.A., R.K. Pace and T.G. Thibodeau, 1999, "Spatial Autoregression Techniques for Real Estate Data," *Journal of Real Estate Literature* 7, 79-95.

Fik, T.J., D.C. Ling, and G.F. Mulligan, 2003, "Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach," *Real Estate Economics* 31:4, 623-646.

Goetzmann, W.N. and S.M. Wachter, 1995, "Clustering Methods for Real Estate Portfolios," *Real Estate Economics* 23:3, 271-310.

Hartzell, D.J., J. Hekman and M. Miles, 1986, "Diversification Categories in Investment Real Estate," *AREUEA Journal* 14:2, 230-54.

Hartzell, D.J., D.G. Shulman and C.H. Wurtzebach, 1987, "Refining the Analysis of Regional Diversification for Income-Producing Real Estate," *Journal of Real Estate Research* 2:2, 85-95.

Hohn, M.E., 1988, *Geostatistics and Petroleum Geology*, Van Nostrand Reinhold, New York.

Markowitz, H., 1952, "Portfolio Selection," *Journal of Finance* 7:1, 77-91.

Matheron, G., 1963, "Principles of geostatistics," *Economic Geology* 58, 1246-66.

Matheron, G., 1965, *Les variables régionalisées et leur estimation,* Masson, Paris.

Miles, M. and T. McCue, 1982, "Historic Returns and Institutional Real Estate Portfolios," *AREUEA Journal* 10:2, 184-98.

Mueller, G.R., 1993, "Refining Economic Diversification Strategies for Real Estate Portfolios," *Journal of Real Estate Research* 8:1, 55-68.

Mueller, G.R. and B.A. Ziering, 1992, "Real Estate Portfolio Diversification Using Economic Diversification," *Journal of Real Estate Research* 7:4, 375-86.

Pagliari, J.L. Jr., J.R. Webb and J.J. Del Casino, 1995, “Applying MPT to Institutional Real Estate Portfolios: The Good, The Bad and the Uncertain,” *Journal of Real Estate Portfolio Management* 1:1, 67-87.

Pannatier, Y., 1996, *Variowin* Springer-Verlag, New York.

Pace, R.K., R. Barry, and C.F. Sirmans, 1998, “Spatial Statistics and Real Estate,” *Journal of Real Estate Finance and Economics* 17:1, 5-13.

Roulac, S.E., 1976, “Can Real Estate Returns Outperform Common Stocks?,” *Journal of Portfolio Management*, 26-43.

Thibodeau, T.G., 2003, “Marking Single-Family Property Values to Market,” *Real Estate Economics* 31:1, 1-22.

Williams, J.E., 1996, “Real Estate Portfolio Diversification and Performance of the Twenty Largest MSAs,” *Journal of Real Estate Portfolio Management* 2:1, 19-30.

Wolverton, M.L., P. Cheng and W.G. Hardin, III, 1998, “Real Estate Risk Reduction through Intracity Diversification,” *Journal of Real Estate Portfolio Management* 4:1, 35-41.

Appendix

Figure 7: 144 commercial return sample points for the conterminous U.S. Source: National Council of Real Estate Investment Fiduciaries

Figure 8: 9699 residential return sample points for the conterminous U.S. Source:U.S. Bureau of Census

	Log. Return using 1970 RHS variables	**Log. Return using 1980 RHS variables**
Intercept	5.925	-1.373
	(46.97)	(-10.44)
High School Education *(educ12)*	0.094	0.126
	(9.90)	(13.70)
College Education *(educ16)*	0.045	0.038
	(10.76)	(7.70)
Num. of Persons 25 yrs. & older *(edupp)*	-0.579	-0.020
	(-26.84)	(-0.92)
Family Income *(income)*	-0.021	0.365
	(1.73)	(31.22)
Aggregate Number of Rooms *(size)*	-0.169	-0.099
	(-8.60)	(-5.27)
Year built 1970-1980 *(bltoc80)*		0.167
		(93.68)
Year built 1960-1969 *(bltoc70)*	0.080	-0.070
	(44.01)	(-21.67)
Year built 1950-1959 *(bltoc59)*	0.023	-0.152
	(7.12)	(-41.03)
Year built 1940-1949 *(bltoc49)*	0.009	0.011
	(2.61)	(3.00)
Year built pre-1939 *(bltoc39)*	0.004	-0.079
	(1.51)	(-33.09)
R^2	0.4996	0.4667

Table 10: Hedonic regression model of 1970-1980 residential returns. *T*-statistics are in parentheses.

	Log. Return using 1980 RHS variables	**Log. Return using 1990 RHS variables**
Intercept	1.897	-3.639
	(15.64)	(-37.67)
High School Education *(educ12)*	-0.061	-0.057
	(-6.93)	(-8.17)
College Education *(educ16)*	0.011	0.030
	(24.73)	(6.63)
Num. of Persons 25 yrs & older *(edupp)*	0.501	0.614
	(26.24)	(44.02)
Family Income *(income)*	0.183	0.456
	(17.13)	(58.06)
Aggregate Number of Rooms *(size)*	-0.772	-0.467
	(-47.31)	(-39.48)
Year built 1980-1990 *(bltoc90)*		0.085
		(59.75)
Year built 1970-1979 *(bltoc79)*	0.027	-0.065
	(16.07)	(-28.82)
Year built 1960-1969 *(bltoc69)*	-0.019	-0.058
	(-6.27)	(-19.85)
Year built 1950-1959 *(bltoc59)*	-0.036	-0.067
	(-10.52)	(-23.64)
Year built 1940-1949 *(bltoc49)*	-0.044	-0.027
	(12.72)	(-10.41)
Year built pre-1939 *(bltoc39)*	0.041	-0.016
	(18.79)	(9.27)
R^2	0.1746	0.3189

Table 11: Hedonic regression model of 1980-1990 residential returns. *T*-statistics are in parentheses.

	Log. Return using 1970 RHS variables	**Log. Return using 1990 RHS variables**
Intercept	5.394	-6.959
	(31.34)	(-49.66)
High School Education *(educ12)*	-0.034	0.026
	(-2.61)	(2.65)
College Education *(educ16)*	0.079	-0.009
	(13.93)	(-1.34)
Num of Persons 25 yrs & older *(edupp)*	-0.049	0.710
	(-1.66)	(34.50)
Family Income *(income)*	0.347	0.840
	(20.69)	(74.10)
Aggregate Number of Rooms *(size)*	-0.817	-0.433
	(-30.58)	(-24.56)
Year built 1980-1990 *(bltoc90)*		0.127
		(62.07)
Year built 1970-1979 *(bltoc79)*		-0.095
		(29.57)
Year built 1960-1969 *(bltoc69)*	0.071	-0.209
	(28.48)	(-50.43)
Year built 1950-1959 *(bltoc59)*	0.022	-0.188
	(5.02)	(-46.45)
Year built 1940-1949 *(bltoc49)*	-0.028	-0.040
	(-6.16)	(-10.79)
Year built pre-1939 *(bltoc39)*	0.054	-0.057
	(15.84)	(-23.25)
R^2	0.4817	0.5712

Table 12: Hedonic regression model of 1970-1990 residential returns. *T*-statistics are in parentheses.

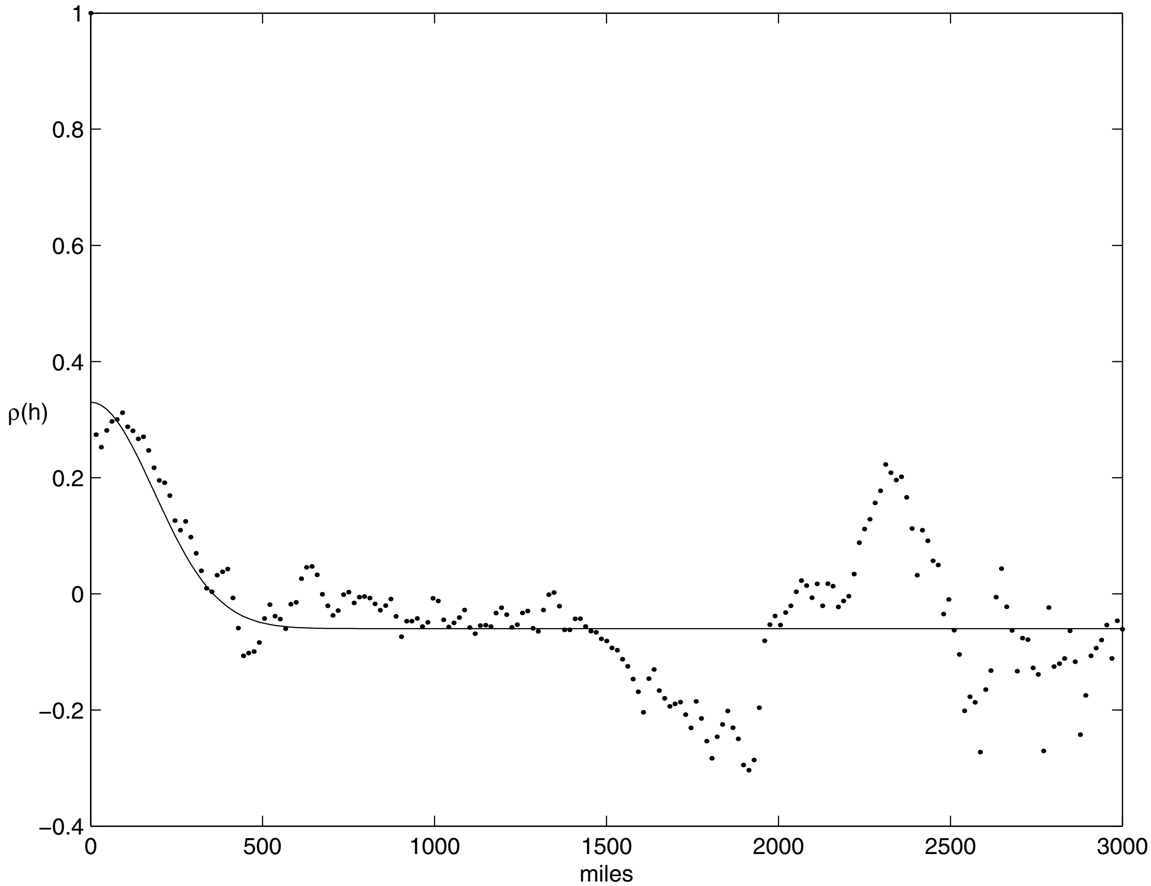

Figure 9: Standardized experimental and theoretical correlograms of residual returns for continental U.S. residential property from 1970 to 1990. Residuals remaining after modeling explanatory variables from 1970. Guassian theoretical variogram based upon a nugget = 0.67, sill = 1.06, and range = 263.5 miles. No lag tolerance.

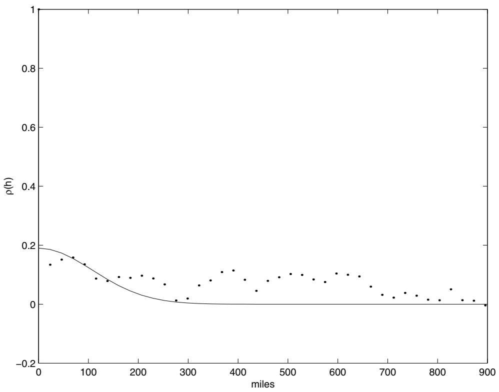

Figure 10: Correlogram of residential residual returns from 1970 to 1980 computed with 1970 RHS variables.

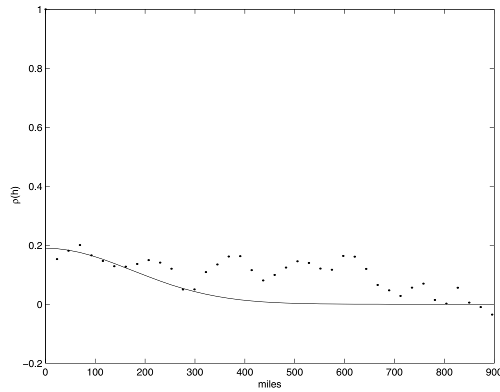

Figure 11: Correlogram of residential residual returns from 1970 to 1980 computed with 1980 RHS variables.

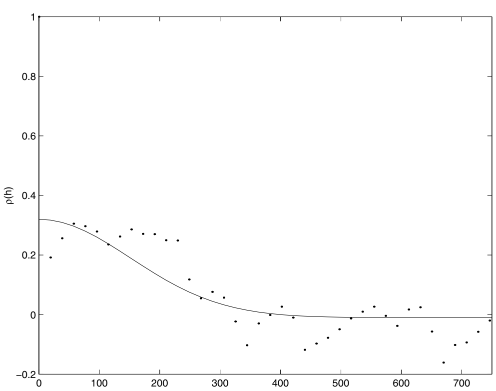

Figure 12: Correlogram of residential residual returns from 1980 to 1990 computed with 1980 RHS variables.

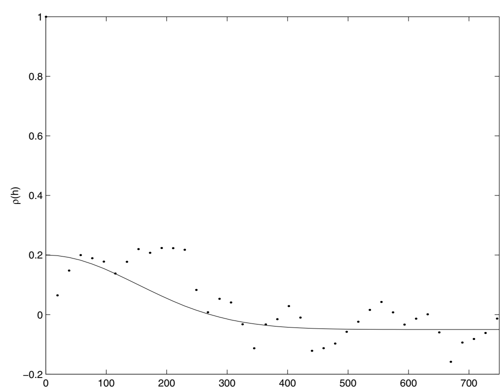

Figure 13: Correlogram of residential residual returns from 1980 to 1990 computed with 1990 RHS variables.