Geographical Analysis ISSN 0016-7363

Impact of Cliff and Ord on the Housing and Real Estate Literature

R. Kelley Pace[1], James LeSage[2], Shuang Zhu[1]

[1]Department of Finance, Louisiana State University, Baton Rouge, LA, [2]Texas State University, San Marcos, TX

The works of Cliff and Ord have had a major impact on empirical practices in real estate. Cliff and Ord proposed both techniques for detecting as well as modeling spatial dependence. Because the existence of spatial dependence is almost assured in real estate data, their most important contribution was feasible means of estimating spatial models. The full implications of these ideas and the numerous modeling techniques spawned by their seminal works have not been fully explored and provide numerous opportunities for future research.

Introduction

Spatial ideas have always been fundamental to real estate and housing. However, the need for simplicity in theory and statistical analysis in early work led to distilling the two dimensions of space into a single dimension distance (such as from each home location to the urban center). Ideally, a regression containing a distance variable would yield residuals that show no obvious spatial patterns.

Because often this did not occur in practice, researchers sometimes included distances to other points, regional or neighborhood indicator variables, polynomials in the locational coordinates, and other trend surfaces in an effort to reduce the obvious map patterns in the residuals. Even after controlling for space in this fashion, samples containing a large number of nearby homes usually exhibit spatial clusters of regression residuals with the same sign. This outcome is because pairs of nearby houses lie in the same neighborhoods so that neighborhood indicator variables do not treat these observations differently. In addition, trend surfaces change little over short distances so that these variables provide little gain in explanatory power, and distances to other locations are virtually the same for pairs of neighboring houses. We now know that these standard statistical techniques were based on assumed independence among sample observations, which real estate and housing data violate. In the face of sample data inconsistent with independence,

Correspondence: R. Kelley Pace, Department of Finance, Louisiana State University, LREC Endowed Chair of Real Estate, Baton Rouge, LA 70803-6308
e-mail: kelley@pace.am

conventional independent statistical methods can at best lead to inefficient model estimates and invalid inference about these parameters.

Against this background, Cliff and Ord (1969) devised a parsimonious specification for the structure of spatial dependence among observations that could be used to quantify the problem of spatial interdependence. Moreover, they proceeded in the corpus of their work (Cliff and Ord 1973, 1981; Ord 1975) to further develop these ideas, and to propose spatial autoregressions that could properly account for spatial dependence in sample data. From a regression perspective, the work by Cliff and Ord led to two strategies. The first involves use of diagnostics to identify spatial dependence in a current model, followed by an associated increase in complexity in the revised model to reduce the dependence. We refer to this as the ''spatial detection strategy.'' As Ripley (1981, p. 98) states,

> Indeed, the philosophy adopted seems to have been that if ''spatial autocorrelation'' is found more explanatory variables should be introduced until it disappears!

The second was to incorporate dependence into the estimation model using a spatial autoregression. We refer to this as the ''spatial estimation strategy.''

In terms of housing and real estate journals, these ideas have been growing in influence over time. An examination of six of the leading journals in this area (*Journal of Housing Economics*, *Journal of Real Estate Finance and Economics*, *Journal of Real Estate Research*, *Journal of Urban Economics*, *Real Estate Economics*, and *Regional Science and Urban Economics*) shows a clear pattern. A search of the terms ''Cliff and Ord,'' ''spatial autocorrelation,'' ''spatial dependence,'' and ''spatial autoregression'' turns up 62 articles published in these journals from 1977 to 2008. Over the 1977–1999 period, 22 of these articles appeared, representing slightly >1 per year. During the 2000–2004 period, 17 of these works were published (or 3.4 per year), and during the 2005–2008 period, 22 articles were published (or 5.5 per year). As another indication of influence, we calculate that 19 of the 25 most cited spatial articles in housing and real estate journals cited at least one work by Cliff and Ord (or Ord), while five of the remaining articles cited an article that directly cited a work by Cliff and Ord (or Ord). References to the ideas of Cliff and Ord often occur indirectly, through well-known books such as Anselin (1988), Griffith (1988), Haining (1990), and various literature reviews in special issues.

In terms of the two strategies implied by their work, the spatial detection strategy has had less influence on housing and real estate research than the spatial estimation strategy. Several factors have limited the influence of the spatial detection strategy. First, the mantra of ''location, location, and location'' suggests an awareness of strong spatial dependence, which ironically may have lessened the need for formal diagnostics of the type set forth in Cliff and Ord (1969). For real estate data that have not been thinned through random sampling or other means to reduce the role of space, a maintained hypothesis should be spatial dependence, because encountering independence in such data would be a shock. Second, practitioners also may have discovered that the strategy of adding additional explanatory variables in an effort to eliminate spatial dependence in disturbances often was not very successful. For example, Pace, Barry, Gilley, and Sirmans (2000) contrast conventional regression models using a sample of 5243 house values and 199 spatial indicator variables for time and space versus a spatio-temporal autoregression (a model that includes spatial and temporal lags of the dependent variable), and report that the conventional regression approach with the large number of additional explanatory variables is not fully successful in addressing the presence of spatial dependence in the disturbances. One might conclude from this that high levels of spatial dependence found in real estate data make use of the spatial detection strategy difficult to execute. We note that the spatial detection strategy may be useful in circumstances where data have low density over space (whether naturally or through random sampling). In this situation, the diagnostics of Cliff and Ord (1969) might reveal residuals that are approximately independent, allowing standard statistical methods to be used.

However, the spatial estimation strategy is becoming more important in real estate modeling because advances over time have resolved many of the computational issues. Ord (1975) proposes modeling strategies for cases involving spatial dependence in both the disturbances (spatial error model) as well as the dependent variable (spatial autoregressive model), and these new methods allow for a parsimonious modeling of spatial dependence in basic regression models that avoids the need for a large number of spatial indicator variables or other complications.

Despite the progress of spatial methods, many practitioners still use ordinary least squares (OLS). Although OLS produces biased and inconsistent estimates in models involving the spatial lag of the dependent variable, in situations where spatial dependence resides in the disturbance process, these estimates should be asymptotically equivalent to those from a spatial error model (although the standard error estimates may well be biased). For housing research that focuses on coefficient estimates, large sample sizes allow OLS to produce regression parameter estimates that should be close to estimates from the spatial error model. Of course, the sample sizes needed to ensure this result are larger, sometimes substantially larger, than for nonspatial data because spatial dependence reduces the effective degrees-of-freedom (Griffith 2005).

We now know that in applied practice, OLS regression parameter estimates often differ materially from spatial error model estimates. This finding is indicative of potential model misspecification. Interestingly, Ord (1975) provides an example of this, which can be seen in his table 2. The OLS estimate reported for the single explanatory variable is 5.27, while the spatial error model estimate is 3.87, with a standard error of 0.66, a gap of 2.12 standard errors away from the spatial error estimate. This statistically significant difference in estimates from the two model specifications is inconsistent with theoretical assumptions about these two models. (Pace and LeSage (2008) propose a spatial Hausman test to formally explore these types of differences.) A possible explanation for this is that disturbances are correlated with the independent variable, which can arise from an omitted variable. Of course, in the presence of omitted variables, OLS produces biased estimates of model parameters and their standard errors.

As a more recent example, Clauretie and Daneshvary (2009) estimate the reduction in price foreclosed properties receive in the market (foreclosure discount) using Las Vegas data. Comparing their nonspatial regression specification 4, in table 3, to the most comparable spatial regression specification 6, shows that the parameter for the foreclosure discount moved from a 10.3 percentage discount to a 8.2 percentage discount, which is a move of 2.23 nonspatial standard errors or 5.10 spatial regression standard errors (note, these are in terms of parameters and not marginal effects). In addition, the time on the market variable went from a *t*-statistic of 2.62 using the nonspatial regression, to 1.12 using the spatial regression.

LeSage and Pace (2009) show that a simple modification of the Ord (1975) explanatory variable specification, which includes a spatial lag of the explanatory variable [the spatial Durbin model (SDM)], provides valid estimates and inferences in the presence of spatially dependent omitted variables. They also discuss the richer spatial spillover interpretation of this model as well as other motivations (spatial heterogeneity, model uncertainty, and spatiotemporal equilibrium) all of which lead to the SDM specification.

Although the housing literature tends to focus on coefficient estimates and associated standard errors, valuation of real estate focuses on predictions. Both assessment for property taxes and automated valuation models used in lending require accurate predictions. A standard method used by appraisers is termed the "adjustment grid method" (Colwell, Cannaday, and Wu 1983). This method requires an appraiser to select "comparable properties" and most appraisal forms allow at least three comparables. The appraisers can decide about weights assigned to each of the comparable properties, and comparables are typically nearby properties with similar characteristics. Pace and Gilley (1998) show that one could interpret the adjustment grid method as a restricted spatial regression model. Although it originates from a different source, the practice of assigning varying weights to nearby observations such as in Cliff and Ord (1969), has a long and successful history in real estate valuation, and the intellectual base associated with the Ord (1975) spatial regression models can improve real estate valuation. In particular, the use of best linear unbiased prediction (BLUP) such as in LeSage and Pace (2004) exploits spatial dependence among observations to improve prediction. Kelejian and Prucha (2007) illustrate sizable gains from using BLUP, and LeSage and Pace (2009) discuss how to rapidly calculate this type of prediction.

Rapid and accurate predictions also allow imputation of unknown values, which has a number of uses. First, imputation allows extension of spatial regression methods using, for example, Bayesian Markov Chain Monte Carlo methods that treat binary, count, or multinomial dependent variable observations as additional parameters to be sampled from their predictive distributions. Second, imputing values for all observations allows researchers to work with a complete sample or population, avoiding problems caused by sample selection. Real estate research frequently relies on homes that actually sell, which are known to differ from a truly random sample of houses. For example, if individuals in better financial circumstances do not place homes on the market during a recession, observed sales might reflect only distress sales. A selective sample of distress sales that does not control for differences in characteristics of distressed and normal properties may yield an artificially bleak picture of an overall market. Repeat sales indices are based on houses that sell multiple times, which also are not a random sample of all houses. Because many policy decisions depend on perceived changes in real estate values over time, controlling for selection biases should improve decision making.

Changes over time in the availability of large spatial datasets along with better hardware and software have made some of the tradeoffs required to use various modeling strategies in earlier times less relevant. For example, Cliff and Ord (1973) and Ord (1975) largely focused on the spatial error and autoregressive models, while a SDM nests these two. Of course, the additional parameters required by the SDM may have seemed problematic given the 26 observations in the Irish dataset used by Ord (1975). In samples involving thousands of observations, the additional parameters in the SDM do not lead to any real efficiency losses, but may materially reduce biases.

The early literature often focuses on choice of a spatial weight matrix. Because all error models (including OLS) yield the same estimates asymptotically under a correct specification of the independent variables, different weight matrices do not yield materially different coefficient estimates in the presence of spatially dependent disturbances and large samples (barring other forms of misspecification). Therefore, larger sample sizes often make concerns over the exact form of weight matrix moot in the case of the error model. For other models, often the marginal effects (partial derivatives) are not sensitive to the choice of the weight matrix in large samples (LeSage and Pace 2009).

The use of eigenvalues proposed in Cliff and Ord (1973) and Ord (1975) to compute maximum likelihood estimates was an advance in the early 1970s. Early computational problems accounted for a substantial amount of research aimed at various ways to avoid maximum likelihood estimation. Over time, alternatives involving approximate and exact methods have arisen that work very well for large data sets (e.g., Pace and Barry 1997; Griffith 2000; Smirnov and Anselin 2001; LeSage and Pace 2009).

In summary, the original Cliff and Ord (1969) work focusing on detecting the presence of spatial dependence had limited influence on real estate modeling practice. However, the corpus of work it motivated about spatial dependence in regression modeling that culminated in Ord (1975) provided the basis for spatially aware housing research. Forty years later, in 2009, the full implications of these ideas and the numerous modeling techniques spawned by their seminal works have not been fully explored. In terms of the intensive margin of spatial econometrics, advances will likely come in areas involving theoretical and econometric motivations for these models, interpretation of estimates from these models, and a greater understanding of the statistical foundation for spatio-temporal processes that underlie these methods (Elhorst 2001; Anselin 2003; Brueckner 2003; Ertur and Koch 2007; LeSage and Pace 2009). On the extensive margin, we expect to see these methods handle more and more econometric problems that arise in practice. For example, currently to estimate a spatial bivariate probit model would be tedious (but possible). Progress with these methods eventually will make such a task rather ordinary. Combinations of data types (spatial, irregular spatio-temporal, panel, flows), dependent variable types (binary, categorical, continuous, duration), and standard econometric problems (simultaneity across variables, selection biases, omitted variables) along with the better modeling of time, space, or multivariate dependence among observations will provide employment for researchers for some time to come.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. (2003). ''Spatial Externalities, Spatial Multipliers and Spatial Econometrics.'' *International Regional Science Review* 26, 153–66.

Brueckner, J. K. (2003). ''Strategic Interaction among Governments.'' *International Regional Science Review* 26, 175–88.

Clauretie, T. M., and N. Daneshvary. (2009). ''Estimating the House Foreclosure Discount Corrected For Spatial Price Interdependence and Endogeneity of Marketing Time.'' *Real Estate Economics* 37, 43–67.

Cliff, A. D., and J. K. Ord. (1969). ''The Problem of Spatial Autocorrelation.'' In *London Papers in Regional Science*, 25–55, edited by A. Scott. London: Pion.

Cliff, A. D., and J. K. Ord. (1973). *Spatial Autocorrelation*. London: Pion Ltd.

Cliff, A. D., and J. K. Ord. (1981). *Spatial Processes Models and Applications*. London: Pion Ltd.

Colwell, P. F., R. E. Cannaday, and C. Wu. (1983). ''The Analytical Foundations of Adjustment Grid Methods.'' *Journal of the American Real Estate and Urban Economics Association* 11, 11–29.

Elhorst, J. P. (2001). ''Dynamic Models in Space and Time.'' *Geographical Analysis* 33, 119–40.

Ertur, C., and W. Koch. (2007). ''Convergence, Human Capital and International Spillovers.'' *Journal of Applied Econometrics* 22, 1033–62.

Griffith, D. (1988). *Advanced Spatial Statistics*. Dordrecht, The Netherlands: Martinus Nijhoff.

Griffith, D. A. (2000). ''Eigenfunction Properties and Approximations of Selected Incidence Matrices Employed in Spatial Analysis.'' *Linear Algebra and its Applications* 321, 95–112.

Griffith, D. A. (2005). ''Effective Geographic Sample Size in the Presence of Spatial Autocorrelation.'' *Annals of the Association of American Geographers* 95, 740–60.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge, UK: Cambridge University Press.
Kelejian, H., and I. Prucha. (2007). ''The Relative Efficiencies of Various Predictors in Spatial Econometric Models Containing Spatial Lags.'' *Regional Science and Urban Economics* 37, 363–74.
LeSage, J. P., and R. K. Pace. (2004). ''Models for Spatially Dependent Missing Data.'' *The Journal of Real Estate Finance and Economics* 29, 233–54.
LeSage, J., and R. K. Pace. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: Taylor and Francis/CRC.
Ord, J. K. (1975). ''Estimation Methods for Models of Spatial Interaction.'' *Journal of the American Statistical Association* 70, 120–26.
Pace, R. K., and R. Barry. (1997). ''Quick Computation of Spatial Autoregressive Estimators.'' *Geographical Analysis* 29, 232–46.
Pace, R. K., R. Barry, O. W. Gilley, and C. F. Sirmans. (2000). ''A Method for Spatial-Temporal Forecasting with an Application to Real Estate Prices.'' *International Journal of Forecasting* 16, 229–46.
Pace, R. K., and O. W. Gilley. (1998). ''Optimally Combining OLS and the Grid Estimator.'' *Real Estate Economics* 26, 331–47.
Pace, R. K., and J. LeSage. (2008). ''A Spatial Hausman Test.'' *Economics Letters* 101, 282–84.
Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.
Smirnov, O., and L. Anselin. (2001). ''Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach.'' *Computational Statistics and Data Analysis* 35, 301–19.