A Spatial Model of Housing Returns and Neighborhood Substitutability

WILLIAM N. GOETZMANN
Yale School of Management, Box 208200, New Haven, CT 06520-8200

MATTHEW SPIEGEL
University of California at Berkeley, Haas School of Business

Abstract

This article provides a method for estimating housing indices at the local level. It develops a “distance-weighted repeat-sales” procedure to exploit the factor structure of the error-covariance matrix in the repeat-sales model. A distance function defined in characteristic and geographical space provides weights for the generalized least-squares model, and allows the use of all of the repeated sales in a metropolitan area to measure returns for the specific neighborhood of interest. We use distance-weighted repeat sales to estimate return indices for all zip codes in the San Francisco Bay area over the period 1980–1994.

When distance is defined in terms of socioeconomic characteristics, we find that median household income is the salient variable explaining covariance of neighborhood housing returns. Racial composition and educational attainment, while significant, are much less influential. Zip-code level indices often deviate dramatically from the citywide index, depending upon income levels. This has implications for investors and lenders. Our results indicate that rates of return may vary considerably within a metropolitan area. Thus, simply using broad metropolitan area indices as a proxy for capital appreciation within a specific neighborhood may not be justified.

Key Words: housing returns, distance-weighted repeat-sales method, neighborhood substitutability

Repeat-sales data within small geographic areas are typically sparse, and this naturally impedes efforts to construct accurate local housing indices. The smaller the definition of the geographical unit, the fewer repeat sales can be used to construct an index, and the less accurate the index becomes. This article provides a method for estimating housing indices for arbitrarily small neighborhoods. Our solution exploits the intuition that the returns to investment in nearby homes should be highly correlated. We model the covariance matrix of repeat-sales errors, using a distance function defined in characteristic and geographical space. This model provides weights for a GLS repeated-sale estimator, and allows us to use all of the repeated sales in the citywide sample to measure returns for the specific neighborhood of interest. We apply our “Distance-Weighted Repeat-Sales” (DWRS) procedure to estimate return indices for all zip codes in the San Francisco Bay area over the period 1980–1994.

The DWRS procedure allows one to estimate the influence of particular geographic and socioeconomic factors as determinants of the return covariance across homes. Further, one can interpret the estimated parameters as indicating the salient factors that make different neighborhoods good or poor substitutes for each other. To see why, consider two neighborhoods, A and B, which are identical in all ways except for factor X, and further suppose...
that home buyers do not care about X’s value. In a competitive market, A and B must have highly correlated capital appreciation returns. If, for example, A appreciates relative to B, then potential buyers will flock to B (the relatively cheap area) and avoid A (the relatively over priced area). Conversely, if buyers regard X as crucial to their choice, there is no reason to believe that the prices in neighborhoods A and B will move together.

The model estimated in this article defines distance in terms of geographical proximity, median household income, average educational attainment, and racial composition. We find that median household income is the salient variable explaining the covariance among neighborhood housing returns. Racial composition and educational attainment, while significant, are much less influential. Geographical proximity is nearly meaningless as a determinant of the covariance among returns.

Our results have immediate implications for lenders, equity holders, and tax authorities. We find that rates of return may vary considerably within a metropolitan area. Consequently, there are opportunities for within-city diversification. In addition, our results suggest that simply using broad metropolitan area indices as a proxy for capital appreciation within a specific neighborhood may not be justified. Finally, the “Distance-Weighted Repeat-Sale” procedure allows accurate estimates of the covariance of housing returns within metropolitan areas, and suggests that low covariances imply gains to diversification for lenders, equity holders, and tax authorities.

The DWRS methodology also has applications to a range of assets, including bonds, commercial real estate, and collectibles, or any other asset whose heterogeneity can be described within an econometrically meaningful characteristic space. While we apply it here to housing, it has natural application to any market that is characterized by infrequent trading. This includes the intra-day stock market, in which minute-by-minute prices are unobserved.

1. Background

Two techniques are commonly used to construct housing price indices. The first, repeat-sales, was first described by Bailey, Muth, and Nourse (1963) and subsequently extended by Case and Shiller (1987), who developed a three-stage estimator called weighted-repeat-sales (WRS). This methodology uses matched purchase and sale price–date pairs for homes within a defined geographical area. The total capital appreciation return from each repeated sale is the dependent variable in a weighted least-squares regression that “explains” these returns by the time periods over which the asset was held.

The second method estimates the value of a “representative” house in the market for each period via a set of priced characteristics. This uses individual home prices rather than matched sale pairs, and in addition makes use of attribute information such as the number of bedrooms, baths and other amenities. Research on the relative merits of the “hedonic” and repeat-sales methods includes Halvorsen and Pollakowski (1981), and Meese and Wallace (1991, 1995), Case, Pollakowski, and Wachter (1991), and Clapp and Giacotto (1992). While no method clearly emerges the winner, the literature thus far makes the trade-offs between the two methods clear. The repeat-sales approach throws away potentially useful information from unique sales, while hedonic indices are not invariate to the choice of