Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

MULTIVARIATE ANALYSIS OF REAL ESTATE PRICES*

BY DELORES A. CONWAY AND DAVID DALE-JOHNSON

University of Southern California

Relatively little is known about the dynamics of prices and returns for local housing markets within metropolitan areas. Most previous economic research has concentrated on developing overall price indexes at the metropolitan or national level. In this paper, we formulate a general equilibrium model of real estate prices for communities within a metropolitan area. The model uses a multivariate, time series framework to generate expected price and return series for different communities. It generalizes earlier models proposed by Case and Shiller (1989) and permits empirical tests of market efficiency within metropolitan areas.

We also give an overview of the multivariate data base under development to estimate the model. The data provide detailed information on 300,000 real estate transactions for twenty cities in the Los Angeles metropolitan area from 1971–1991. Preliminary analysis shows substantial variation in prices, volume, and housing characteristics over time and across the different cities. The results have implications for the construction of housing price indexes to reflect this variation.

1. Introduction. Research into housing prices, rates of appreciation and volatility is fundamental to our understanding the economic behavior of real estate markets. Equity in the home dominates the wealth portfolio of the average household (in the USA, over twice the value of stocks and bonds). Even so, little is still known about housing market performance. Often quoted median or average price indexes may be a biased measure of true quality adjusted house price changes because the sample of houses that sell each period may have different characteristics. Indexing techniques have evolved to deal with the problem that arises because individual houses have different characteristics and are bought and sold at lengthy intervals. These include the repeat sales methodology of Bailey, Muth and Nourse (1963) (BMN) and the hedonic approach of Case and Quigley (1990).

---

* Research partially supported by the National Science Foundation and the School of Business Administration Research Fund at the University of Southern California.

AMS 1980 Subject Classifications. Primary 62M10; secondary 90A20, 62H99.

Key words and phrases. Financial markets, Los Angeles housing, multivariate time series, structural models.

Case and Shiller (1989) use the BMN repeat sales methodology. This approach creates an index by identifying repeat sales of the same houses and using simple regression techniques. The approach avoids the quality adjustment problem, but it dramatically reduces the number of useable transactions, introduces potential selection bias, and ignores sales volume. Case and Quigley (1990) use the same approach and a seemingly-unrelated-regression model (SURM) to show that augmenting repeat sales with hedonic data yields more efficient price estimates. The suitability of alternative indexing methods remains unresolved.

This paper presents an equilibrium model to analyze the dynamic behavior of housing prices and derive price indexes for submarkets within a metropolitan area. We also give an overview and preliminary results from the multivariate data base that is being developed to estimate the model. Price, volume and housing characteristics are examined for a subset of jurisdictions within Los Angeles County. Our objective is to illustrate some unique aspects of housing data that have implications for model construction and price indexing methodology.

2. An Equilibrium Model of Housing Prices and Returns. Conway and Dale-Johnson (1990) propose an equilibrium pricing model of existing single family houses for a metropolitan market. In our model, we consider K submarkets within an urbanized area. We assume that different houses within each submarket are priced in the same way. Let P_{ijt}, $\{j = 1, ..., n_i\}$ represent the observed log price of the jth individual house in submarket i, $\{i = 1, \cdots, K\}$ at time period t, $\{t = 1, \cdots, T\}$. Let the vector X_{ijt} represent individual physical characteristics of the house such as square feet of living space. An additional set of variables, V_{it}, includes measures of the fiscal condition of the jurisdiction and amenities. The pricing model is written,

$$\begin{aligned} P_{ijt} &= \alpha_{it} + \beta_{it}^{'} X_{ijt} + \gamma_i^{'} V_{it} + \varepsilon_{ijt}, \\ \varepsilon_{ijt} &= \varepsilon_{ijt-1} + (\delta_{it}^{'} Z_t + \delta_{it-1}^{'} Z_{t-1} + \cdots) + \nu_{ijt}. \end{aligned} \tag{1}$$

The errors ε_{ijt} are not independent over time, but have the dynamic structure specified by the second equation in (1). The vector Z_t refers to a set of exogenous macroeconomic variables which impact on the prices P_{ijt} from period to period. These include, for example, measures of employment, effective mortgage rates and inflation. The random error ν_{ijt} is assumed to be uncorrelated over time and independent across submarkets. It also has constant variance within a submarket and over time, so that $\mathrm{Var}(\nu_{ijt}) = \sigma_i^2$. For simplicity, we consider a linear relationship with no interaction effects as a first approximation to the functional form of the model.

The first difference of the model in (1) yields the following measure of

returns for individual houses:

$$R_{ijt} = (\alpha_{it} - \alpha_{it-1}) + (\beta_{it}' X_{ijt} - \beta_{it-1}' X_{ijt-1}) + \gamma_i' \Delta V_{it} + (\delta_{it}' Z_t + \delta_{it-1}' Z_{t-1} + \cdots) + \nu_{ijt}. \quad (2)$$

Changes in prices of individual houses are a function of changes in the drift parameter α_{it}, changes in the characteristics of the house and the neighborhood, and exogenous macroeconomic variables.

3. Analysis of Data for Los Angeles County. We have developed a comprehensive data base which consists of all sales of single family houses for Los Angeles County from January 1971 to the present. The data base includes more than 1.5 million sales transactions and is updated annually. The three main components provide transactional, community, and exogenous measures. The transactional component includes the sale price and housing characteristics, such as living space and number of rooms, for all sales. The community component includes revenue and expenditure data from more than 80 individual jurisdictions. The exogenous component includes macroeconomic time series measures such as interest rates, inflation, and other measures of regional and national economic conditions.

This paper gives preliminary results from the analysis of 20 submarkets within the Los Angeles metropolitan area. These submarkets comprise 350 thousand of the 1.5 million housing transactions. Summary data for the 20 submarkets show that the compound annual rates of return based on the median price are substantial, ranging from 9.5 to 13.4 percent. The number of transactions falls abruptly during 1981 and 1991 reflecting the impact of the recessions. The median prices across the 20 submarkets also have substantial variation. For example, the median price of $1,150,000 in Beverly Hills for 1991 exceeds the median price of $113,000 in Compton by a factor of ten. Variation in returns, by submarket, is significant but far less dramatic than the variation in price levels.

Examination of the annual data from the 20 submarkets illustrates the dynamic behavior of prices within the Los Angeles area. Prices typically rise in nominal terms from period to period and remain relatively flat during recessions, reflecting downward stickiness. That is, prices seem to resist downward movements. However, market activity as measured by volume of transactions drops dramatically during economic slowdowns. For example, during the three recessions of 1974, 1981 and 1991, volumes drop roughly 20, 90 and 75 percent, respectively, from prior peaks. Figure 1 illustrates this behavior and graphs the median price and number of transactions by year for Beverly Hills.

We plan to use quarterly prices for specific areas to capture fundamental price dynamics when estimating model (1). The model should improve upon indexes based on quarterly median housing prices, which are highly volatile.

For example, housing prices in Beverly Hills during the last three quarters of 1982 drop over 75 percent. During the first two quarters of 1989, they drop over fifty percent. Many of these changes reflect quarterly differences in the characteristics of the houses sold. The quarterly time series on median living space for these transactions has the same volatility and the same patterns as the quarterly median price series. The two series closely track one another and are strongly correlated. Because the equilibrium model incorporates the characteristics of houses sold each quarter, the derived price indexes should more accurately reflect true prices.

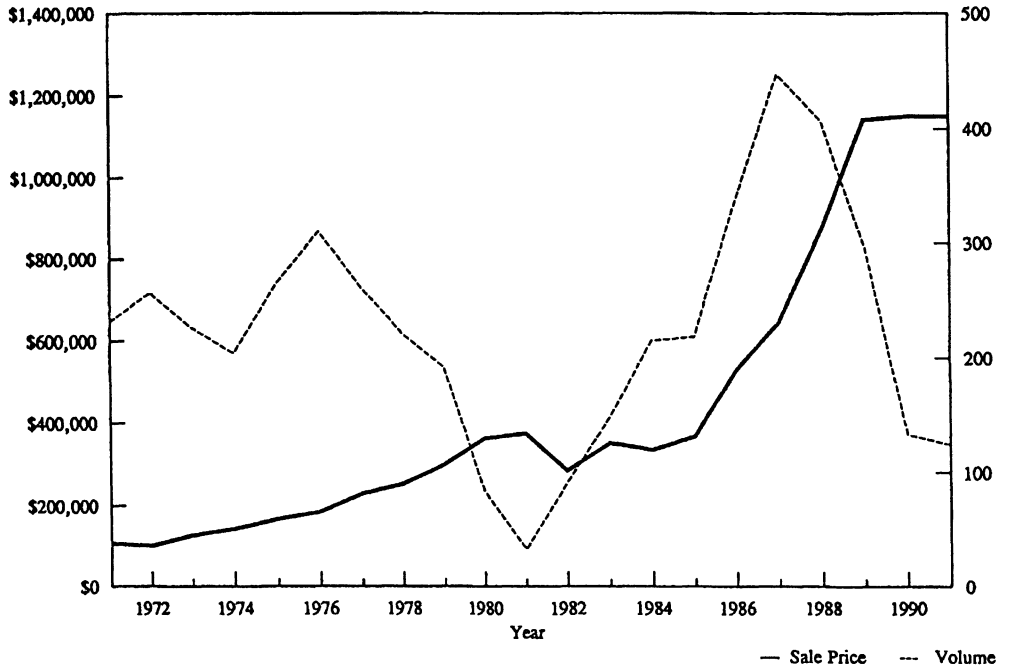

Figure 1. Median Sale Price and Volume by Year for All Houses Sold in Beverly Hills (1971-1991)

We find that price variation is highly correlated with building area within submarkets that account for the location effects in housing price levels. Our preliminary results suggest that most of the cross-sectional price variation within a submarket is explained by living space (about 80 percent). While most hedonic regression models include five or six explanatory variables, careful definition of homogeneous submarkets can yield more precise estimates of price with fewer explanatory variables.

4. Conclusions. The preliminary analysis of the Los Angeles data in

this paper is an important step in the development of a multivariate, time series model of housing prices for a metropolitan area. Our data suggest great diversity in property values and housing returns within Los Angeles County. The dynamic movements of housing prices, attributes, and volume require careful analysis so that price indexes capture real changes.

Hedonic price methods are likely to be greatly enhanced through careful definition of smaller market areas to account for important location differences. Consequently, a greater number of price indexes may be necessary to capture the richness of a metropolitan market. Furthermore, simple models with fewer explanatory variables may explain the variation in housing prices within local areas if the submarkets are defined carefully.

Empirical measures of housing prices and returns at the local level are important decision-making tools for investors, lenders and policy-makers. When these measures are based on metropolitan median or average prices, they may be misleading. For example, between 1981 and 1986, the median house price in Los Angeles County increased by 21 percent. However, the median price in Beverly Hills increased by 41 percent, whereas it declined by 9 and 19 percent in Monterey Park and Palos Verdes, respectively. A multivariate, time series model that stratifies on location is likely to yield improved and more accurate measures of price variation within local submarkets.

Acknowledgements. The authors would like to thank Ingram Olkin, two anonymous referees, and the editors for their helpful comments on a previous version of this paper.

REFERENCES

BAILEY, M. J., MUTH R. F. and NOURSE H. O. (1963). A regression method for real estate price index construction. *J. Amer. Statist. Assoc.* **58** 933–942.

CASE, B. and QUIGLEY J. M. (1990). The dynamics of real estate prices. *R. Econ. Statist.* **73** 50–58.

CASE, K. E. and SHILLER R. J. (1989). The efficiency of the market for single family homes. *Amer. Econ. R.* **79** 125–137.

CONWAY, D. A. and DALE-JOHNSON D. (1990). Real estate asset pricing and market efficiency. *Proceedings of the Business and Economic Statistics Section* 10–16. Amer. Statist. Assoc. Washington D.C.

School of Business Administration
University of Southern California
Los Angeles, CA 90089-1421
USA