A SPATIAL ANALYSIS OF PRICE CHANGE IN THE HOUSING COMPONENT OF THE CONSUMER PRICE INDEX   December 2006

William E. Larson

U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue, N.E., Room 3655,
Washington, D.C. 20212
larson_W@bls.gov

Opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Key Words: CPI, spatial statistics, housing

Historically there has been an assumption that geographic proximity is one of the most important predictors of price change behavior in the housing component of the Consumer Price Index (CPI). However there has been limited research examining at what level of geography rent changes are correlated. This spatial data analysis explores the strength of the geographic relationship.

1. CPI Shelter Indexes and Price-Change Variances

For a full discussion of the CPI the reader is referred to Chapter 17 of the *BLS Handbook of Methods*. The rent estimates (also called economic rent) used in the CPI are "contract rents", which include all services, such as utilities, included in the rent payment. The REQ estimates (also called pure rent) are contract rent payments adjusted so that only that the shelter is included.

The CPI includes 31 self-representing PSUs consisting of large Metropolitan Statistical Areas MSAs), and 56 non-self-representing PSUs selected to represent medium to small-size MSAs in 4 Census regions and urban non-MSAs in 3 regions.

The 1999 housing sample is a stratified cluster sample that represents housing built before 1990. Census blocks were grouped into segments based on minimum numbers of housing units, and the PSU was divided into 6 geographic strata. Segments were ordered by shelter expenditure within county within a stratum, and chosen by a systematic sample. The weight for a segment (W_s) was assigned based on segment shelter expenditure as a proportion of total PSU shelter expenditure. The shelter weight is divided between rent and REQ. The rent and REQ indexes are estimated separately using a 6-month chained estimator. The PRC aggregates the weighted rents for the units (i) in the Index Area (a) for the current period (t) and for the 6-months previous (t-6) and then computes a 6-month price relative:

$$REL_{a,t,t-6} = \frac{\sum_{i \in a} W_s * R_{i,t}}{\sum_{i \in a} W_s * R_{i,t-6}}, \text{ where}$$

R=rent.

The rent and REQ indexes need a 1-month price relative, so the 6th root of the $REL_{a,t,t-6}$ is derived:

$$REL_{a,t,t-1} = \sqrt[6]{REL_{a,t,t-6}}$$

and passed to the Index Estimation system where:

$$IX_{a,t} = IX_{a,t-1} * REL_{a,t,t-1}.$$

2. Use of Spatial Information within PSUs in the Current CPI

Each PSU is divided into six geographic strata. The two central strata were formed by taking the smallest square in the central portion of the PSU that includes one-third of the housing expenditure. This area was then split in two to form stratum 1 and stratum 2. The remaining, or suburban, portion was split in two in a way that maximized the mean rent level difference, and then each of those two areas was split to form strata 3 and 4, and 5 and 6. In the current price relative calculation of the housing component of the CPI, if a housing unit fails to return a quote for a given collection period then its weight is distributed according to a defined scheme based on the stratification.

3. Study of Spatial Effect on Rent Relatives

All of the housing units that reported usable quotes in the CPI from July 1998 through December 2005 were geocoded using Tele Atlas North America Inc. EZ-Locate Client©.

The units' addresses were edited and standardized and then passed through the program. The program returns longitude and latitude in seven categories, including an exact match, a near match (same block), ZIP+2, 5-digit ZIP, 3-digit ZIP, multiple street segments and no match. Exact matches, near matches, and ZIP+2 matches were deemed successful. Further research was done on the unsuccessful matches using the U.S. Postal Service website and websites such as Mapquest and Google maps. In some instances the original listing of the unit was used: when a block was chosen for the housing sample, all of the housing units in that block were listed. If the address of the selected housing unit could not be matched, a close neighbor was substituted for processing in the program.

The units actually used in the study were those that had usable quotes in the period between January 2003 and December 2005. In that period there were 32,334 housing units reporting 146,170 usable quotes. The dependent variable was the log of the normalized six-month price relative. This is a relative that is unadjusted except to bring rents that are collected for non-monthly periods, e.g. weekly rents, to a monthly level.

Previous work has shown that rent level has a remarkable predictive value for price change. Accordingly, a regression was performed on the data by categorizing rents as low (below the 25th percentile), high (above the 75th percentile) and medium Rent level turned out to be significant at $p<0.0001$. Accordingly the spatial analysis was run on the residuals of this regression. The distances that were examined were relatively small so PSU was not used as an independent variable in the regression. Units that are a short distance from each other but in adjacent PSUs were included in the analysis.

The residuals were analyzed by constructing experimental variograms, which are empirical estimates of the covariance that quantify spatial correlation The empirical variogram is computed by measuring the mean-squared difference of the value of interest **z** (in this case the residual of the log of the normalized price relative) evaluated at two points **x** and **x+h**. This mean-squared difference is the semivariance and is assigned to the value **h**, which is known as the lag. The variogram is the plot of the semivariance against **h**. The notation for this computation is

$$2g(h) = \frac{1}{n(h)} \sum_{n(h)} [z(u) - z(u + h)]^2$$ for all

possible locations **u** and where **n(h)** is the number of pairs for lag **h**. Isotropy, that is directional invariance, was assumed. An experimental variogram that demonstrates a distance relationship will show a relatively low value at **h=0**, which is known as the nugget and is purported to be measurement error. The sill is the asymptotic value at which there is no apparent spatial correlation. The partial sill is the difference between the nugget and the sill. The range is the distance at which the variogram reaches the sill. If there is a spatial correlation evident in the data we should be able to see a consistent pattern in the variograms.

The results fail to show any consistency or spatial correlation. Figure 1 shows the variogram for all data from January 2003 through December 2005. We see a nugget at lag distance h=0, and then a rise up to an apparent sill at lag distance h=6. However, after that point the variance actually decreases, suggesting an increase in spatial correlation as distance between two points as distance increases. The data were examined more closely. Variograms were charted for each collection period. Figures 2-4 show the experimental variograms for December 2003, December 2004, and December 2005 respectively. These charts are representative of the variograms computed for all the collection periods, in that there is no consistent spatial correlation shown. Figure 2 shows a high initial value at lag distance=0, and then a decreasing variance (increasing correlation) to lag distance=1.0, at which point the variance increases. Figure 3 shows a change in direction of variance at virtually every lag point, as does Figure 4.

4. Conclusions

On the basis of the variograms calculated in this study, we must conclude that there is not a relationship between rent relatives and the distance between housing units that is consistent and therefore usable for prediction, imputation, or weight assignments in the price relative calculation of the CPI.

5. References

Chapter 17, *BLS Handbook of Methods*, on the
Internet at
http://www.bls.gov/opub/hom/homch17_a.htm

Figure 1. Variogram Plotting Semivariance of Distance between Housing Units versus Residuals of Log Rent Relatives, January 2003 to December 2005.

Distance * Residuals of Log Rent Relatives
January 2003 to December 2005
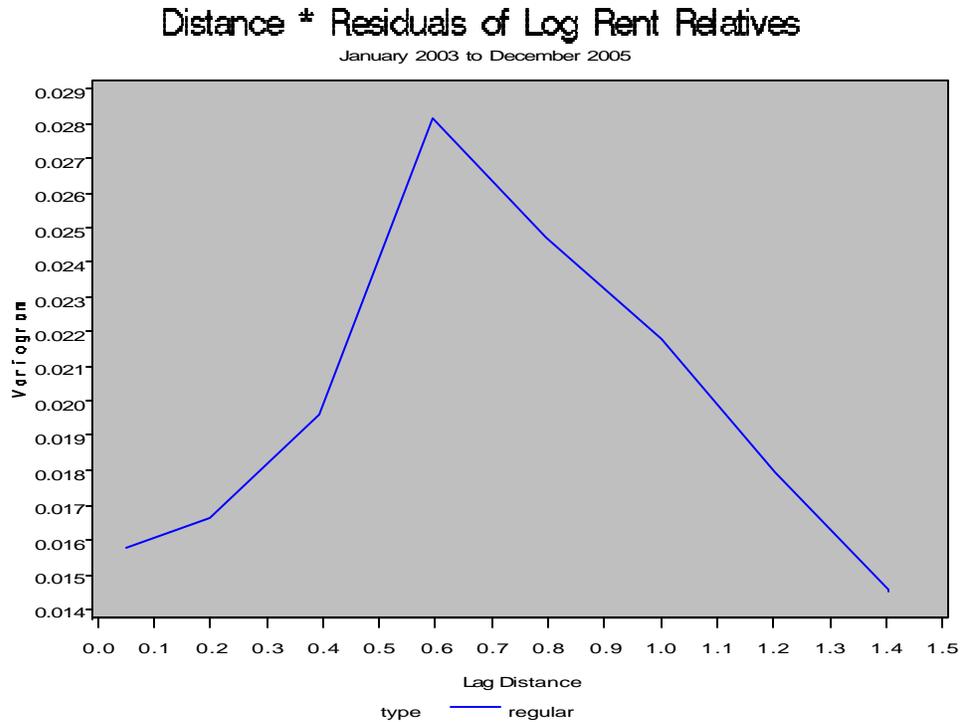
Figure 2. Variogram Plotting Semivariance of Distance between Points versus Residuals of Log Rent Relatives, December 2003.

Distance * Residuals of Log Rent Relatives
December 2003
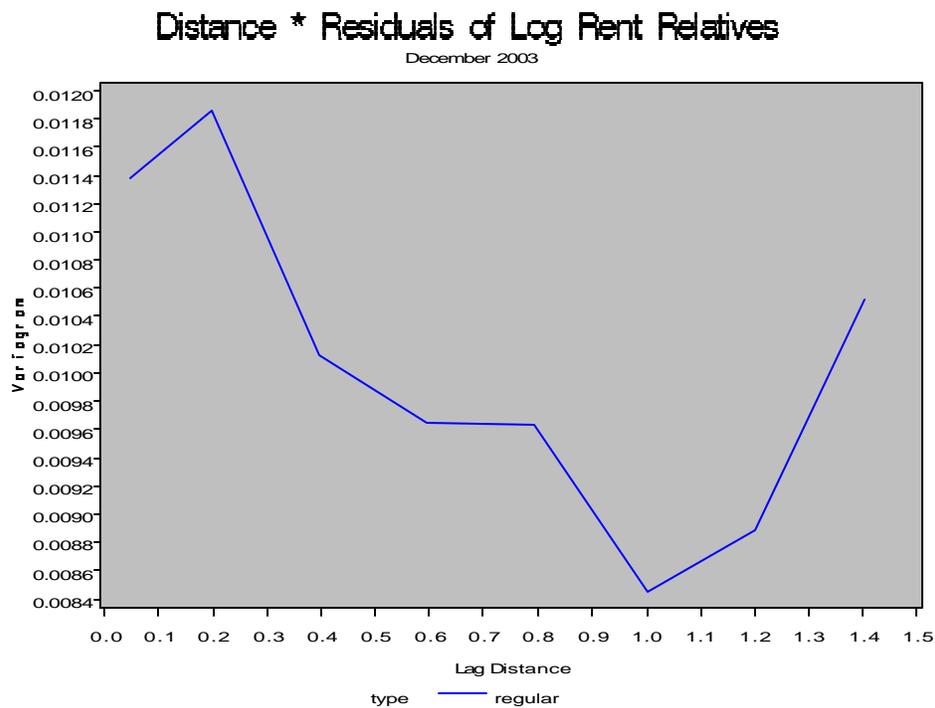
Figure 3. Variogram Plotting Semivariance of Distance between Points versus Residuals of Log Rent Relatives, December 2004.

Distance * Residuals of Log Rent Relatives
December 2004
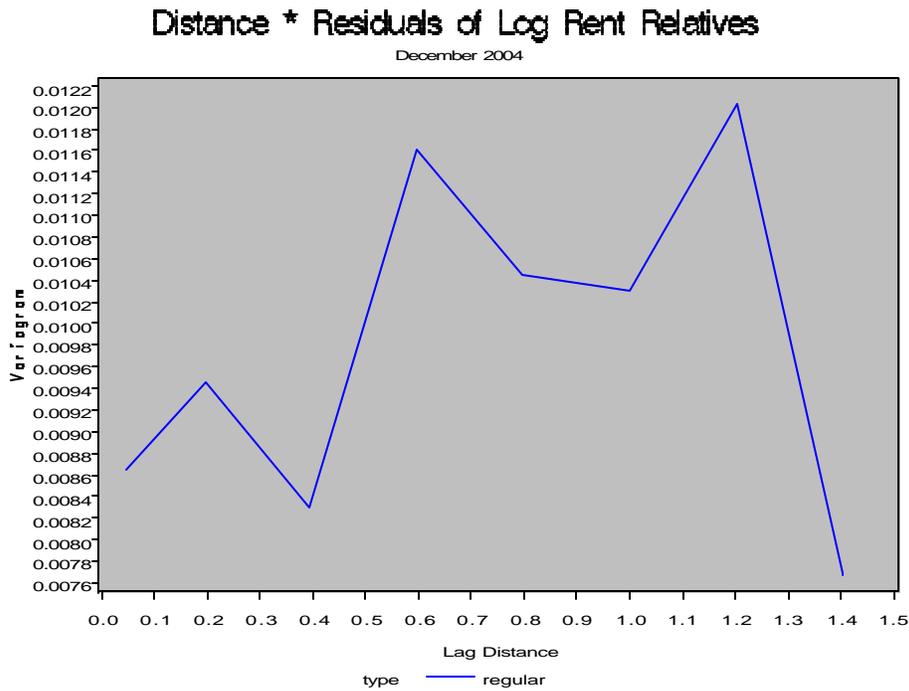
Figure 4. Variogram Plotting Semivariance of Distance between Points versus Residuals of Log Rent Relatives, December 2005.

Distance * Residuals of Log Rent Relatives
December 2005