Developing More Accurate Reasonable Rent Estimates in the U.S. Housing Choice Voucher Program

David P. Varady, Xinhao Wang and Jay Mittal*
University of Cincinnati

If the U.S. Department of Housing and Urban Development's (HUD's) Housing Choice Voucher program (HCV) is to be successful, it is important that the rents charged by property owners are reasonable; otherwise the effect could be to inflate costs in the housing market as a whole. This paper describes a cost-effective approach that local public housing authorities in the United States could use to develop reasonable rent estimates. The methodology involves (1) cluster analysis to combine similar census block groups, and (2) regression analysis to predict market rents based on cluster location, number of bedrooms, building age, and building type. In the paper we demonstrate the superiority of this methodology to the tabular housing market region methodology currently being used by the Cincinnati Metropolitan Housing Authority. Our clusters—which contain block groups not necessarily spatially linked—are relatively homogeneous and therefore correspond relatively closely to the housing submarkets utilized by householders in making their mobility and locational choices.

Keywords: *Housing, reasonable rent, voucher, regression*

The U.S. Department of Housing and Urban Development (HUD) requires public housing authorities (PHAs) to make rent reasonableness determinations before the PHA enters into a housing assistance payment (HAP) agreement with an owner or before the PHA pays any rent increase to the owner. HUD does not specify the methodology PHAs should follow in determining rent reasonableness but provides detailed suggestions in its *2001 Voucher Program Guidebook* (HUD, 2001). For example, HUD recommends that the local authorities conduct surveys of the rental housing market and suggests the types of information they should gather on individual rental units, i.e., location, building type, unit sizes and rents, utility costs, amenities provided by the owners. However, HUD does not offer guidance on the geographical level that should be the basis for the study, that is, census tract block

* School of Planning, P.O. Box 210016, University of Cincinnati, Cincinnati, OH 45221-0016, USA. Email: David.varady@uc.edu; xinhao.wang@uc.edu; mittalj@email.uc.edu.

groups, census tracts, or groupings of census tracts ("communities").

This paper summarizes the methodology we have developed for the Cincinnati Metropolitan Housing Authority (CMHA) and the Hamilton County Department of Community Development (HCDCD) to estimate reasonable rents based on building age, building type, number of bedrooms, and location. We demonstrate how our technique provides more accurate estimates of reasonable rents than the technique currently used by CMHA and HCDCD. The City of Cincinnati is a mid-sized city of 314,154 located in Hamilton County (population 845,303 in 2000) and is located in the southwestern corner of the State of Ohio.

In 2000, and then in 2001, we conducted rental surveys for CMHA and HCDCD; each resulted in data sets containing information on rents, utilities, building type, and apartment size. CMHA and HCDCD used the survey results to divide Hamilton County into ten "markets" based on common rents among the census tracts. Staff compared the asked-for rent with the calculated reasonable rent, defined as the average rent in the market the property fell into. The main weakness of this approach is that these large market areas are in fact quite heterogeneous and contain submarkets that should have been separated out. The impetus for this study was the need of CMHA and HCDCD to assure that the data and methodology used to assess rents in their respective Housing Choice Vouchers (HCV) programs still met the requirements of 24 CFR 982.507 and Chapter Nine of the *Voucher Program Guidebook* (HUD, 2001). For example, HUD recommends that the local authorities conduct surveys of the rental housing market and suggests the types of information they should gather on individual rental units, i.e., location, building type, unit sizes and rents, utility costs, amenities provided by the owners. However, HUD does not offer guidance on the geographical level that should be the basis for the study such as census tract block groups, census tracts, groupings of census tracts ("communities").

Additionally, the initial internal analyses within both agencies indicated a recent softening of overall rental market values in Hamilton County and marked market disparities within some census tracts. The preceding implied the need for a reasonable rent estimation technique based on small geographical units (i.e., census block groups) rather than larger agglomerations of census tracts.

Consequently, in 2004, CMHA and HCDCD asked us to provide them with more up-to-date rental data and to develop a methodology to estimate reasonable rents for geographical areas smaller than census tracts. We applied cluster analysis using socio-economic characteristics from the census to classify census block groups in Hamilton County into six clusters, representing the geographical factor. We then developed a regression model to predict gross rent, using the clusters, rental unit size, building age and building type as input variables.

In 2005, CMHA bought a new software system for handling its housing voucher client data. According to a CMHA staff member, the new system precluded the use of the regression model that we had developed. At this point in time HCDCD had

not adopted the new software and consequently was interested in considering using the regression approach. However this possibility evaporated in early 2006 when the two voucher programs were merged to be operated solely by CMHA.

About the same time, CMHA and HCDCD hired a consultant to develop a rental market database to be used in conjunction with the new software to determine reasonable rent for units proposed for or assisted under the tenant-based Section 8 program. CMHA decided to go back to the previously used housing market approach. We believe that CMHA and HCDCD missed an important opportunity for developing accurate reasonable rent estimates when they dropped our methodology for bureaucratic reasons. Consequently, a major aim of this paper is to advocate for the use of a regression-based methodology for rent reasonableness estimates for U.S. metropolitan areas. While it appears that our methodology will not be used in Hamilton County, due to changes in voucher software used by CMHA, it is broadly suitable in other metropolitan areas. The remainder of this paper supports this assertion.

Section 2 describes the development of a regression model for estimating reasonable rents. In Section 3 we compare the rent estimation technique we developed, based on clusters of census block groups, with the technique recently chosen by CMHA which is based on the use of housing markets.

RENT REASONABLENESS ESTIMATION BASED ON CLUSTERS

The 2004 Rent Survey

In the survey we relied heavily on newspapers (*Cincinnati Enquirer, Cincinnati Post, Western Hills/Price Hill/Delhi Press*) and web sites (apartmentguide.com and rent.com) for lists of apartments for rent, but also obtained a smaller number of listings from community leaders and CMHA staff.

During the period, July to August 2004, a graduate research assistant called landlords and apartment managers seven days a week between 10:00 AM and 9:00 PM, inquiring about: (1) the exact street address; (2) the number of bedrooms in the apartment; (3) the rental cost; (4) the number of apartments in the building; and (5) the building type (e.g., house, garden apartment). The graduate assistant posed as a typical unassisted consumer who was seeking housing. She was able to obtain information on a total of 1004 properties.

Among the 1004 properties, 950 had a valid address in Hamilton County. The latter properties were geocoded with ArcGIS, Geographic Information System (GIS) software (ESRI, Redlands CA). Then the census block groups where those properties were located were identified and recorded. After the survey and the initial analysis were completed CMHA/HCDCD decided to change to gross rent as the basis for reasonable rent estimation, rather than the originally planned contract

rent. Gross rent is the cost of renting a property including utility costs. On the other hand, contract rent is the asked-for rent of a unit. If the property owner pays for all of the utilities then the gross rent equals to the contract rent. However, if the tenant is expected to pay some or all of the utilities, then it is necessary to identify this amount and add it to the contract rent to calculate the gross rent. At that time we realized that we had not gathered information on utility costs and as a result could not measure gross rents across our sample.

Adding Utilities to Compute Gross Rents

Between December 2004 and January 2005, two research assistants and two HCDCD staff re-contacted the property owners originally interviewed in Summer 2004 to obtain two types of utility data: (1) whether the owner or the tenant was expected to pay for gas, electricity, heating, or water; and (2) whether space heating, cooking, or water heating was provided by gas or electricity. We were able to obtain utility information for 532 properties at different levels of detail. Appendix 1 lists the steps for calculating utility costs.

Because some properties had multiple apartment types and varying numbers of bedrooms, our dataset included 805 unit types at 532 properties for the rent analysis. To estimate utility allowances for particular properties we took into account the rent survey information on the building and utilities as well as utility allowance standards from HUD's calculation sheet (CMHA, 2005). Table 1 displays the modified utility allowance standards using the assumptions stated above. The table lists the different types of allowances, such as space heating, taking into account the number of bedrooms in the home. For example, we calculated that for a two bedroom single-family house, where the tenant paid for all utilities, the allowance would be: \$64+\$10+\$21+\$59+\$26 = \$175. Next, we added, on a property-by-property basis, the calculated utility allowance to the contract rent to obtain the gross rent.

Table 1: Utility allowance.

Number of Bedrooms	Space heating (single house)	Space heating (others)	Cooking	Water heating	Water, sewer and trash collection	Electric
0	\$36	\$27	\$7	\$12	\$53	\$16
1	\$50	\$38	\$9	\$16	\$56	\$21
2	\$64	\$49	\$10	\$21	\$59	\$26
3	\$79	\$60	\$12	\$26	\$62	\$31
4	\$100	\$77	\$14	\$33	\$67	\$38
5	\$114	\$88	\$16	\$37	\$71	\$43
6	\$129	\$99	\$18	\$42	\$73	\$48

Source : CMHA, 2005

Cluster Analysis

Census block group level data for Hamilton County were downloaded from the U.S. Census Bureau's website (www.census.gov) using census block boundaries provided by the Cincinnati Area Geographic Information Systems (CAGIS). The following census variables were used for this study: median household income in 1999 (P053001), household income in 1999 below poverty level (P092002), total households (P092001), total population (P006001), Black or African American population (P006003), owner occupied housing units (H007002), total occupied housing units (H007001), median value of owner-occupied housing units (H085001), median gross rents (H063001), education level for population 25 years and over (P037001- P037035), and median year structure built (H035001). From the census data, we derived the following variables: proportion of households in poverty, proportion of African Americans, proportion of homeowners, and proportion of population that attended college.

Cluster analysis, part of the SPSS® software package (SPSS Inc. Chicago, IL), was used to combine the 731 census block groups into six clusters based on five census variables: percent African American, percent attending college, median household income, proportion of households in poverty, and proportion of homeowners. The variables that we used—measures of socio-economic status, life style, and ethnicity—are the ones sociologists and geographers have used to study urban residential differentiation (Knox and McCarthy, 2005). Our decision to stop the clustering process at six clusters was based on changes in the agglomeration statistic. The agglomeration coefficients are used for guidance in deciding how many clusters are needed to represent the data. It is best to stop agglomeration as soon as the increase between two adjacent steps becomes large. In our cluster analysis this occurred between stage 724, seven clusters (coefficient=2.218E+10) and stage 725, six clusters (coefficient= 2.979E+10) leading to our decision to use six clusters. Table 2 lists the six clusters and the median household income (Detailed results are available from the authors).

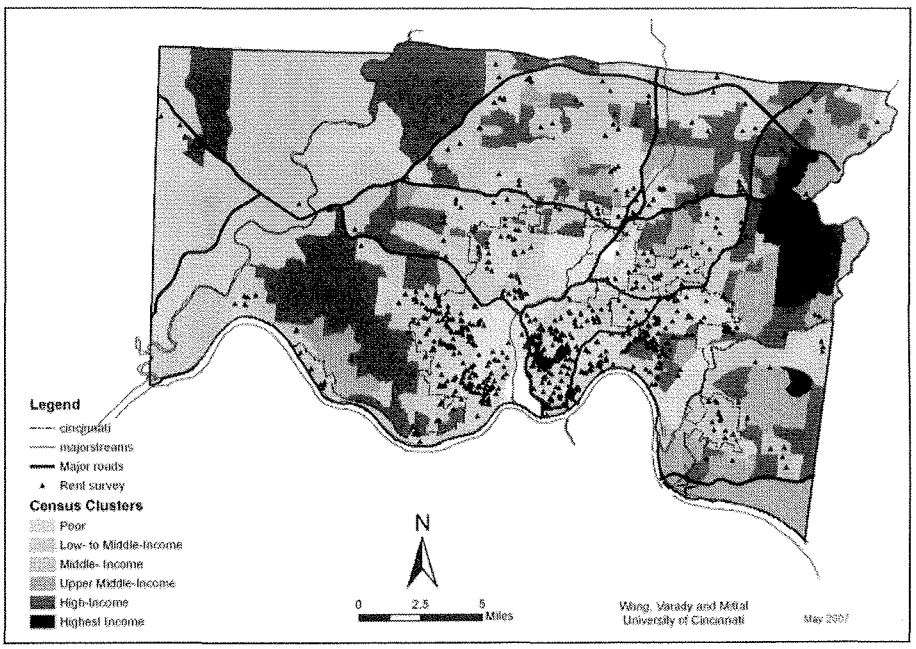
Table 2: Clusters of census block groups.

Cluster ID	Cluster Name	Average Median Household Income ($)
1	Poor Income	$20,170
2	Low- to Middle-Income	$36,970
3	Middle- Income	$50,479
4	Upper Middle-Income	$67,129
5	High-Income	$101,259
6	Highest Income	$182,624

Geocoding the rental survey data and incorporating cluster membership

The rental survey data were geocoded using ArcGISMap. The reference file was the street center line file in the CAGIS database. We checked the locations using (1) Zip code information, (2) municipality and City of Cincinnati's Statistical Neighborhoods data, (3) Hamilton County Auditor's property data (http://www.hamiltoncountyauditor.org/realestate/), and (4) the Mapquest software. Information on each rental property's cluster group membership (one through six) was obtained from the census clusters through the GIS overlay function and was combined with the geocoded rental survey. That is, we identified the census block group for each rental unit and in doing so linked particular housing units to particular clusters. Figure 1 shows the spatial distribution of the rental units among the six clusters.

Figure 1: Census clusters and 2004 rent survey units.



Regression analysis

The regression feature of SPSS, along with the above merged data base, was used to account for variations in the rents for the 805 units based on (1) the number of bedrooms; (2) the location of the property in one of the six census clusters; and (3) the type of building. Since the latter two factors are nominal variables (i.e., the numbers do not have any inherent meaning), we created a set of "dummy" variables where the variable equals 1 if the property has that characteristic and it equals 0 if it does not. There were no rental properties in cluster six; hence we only created and

utilized five dummy variables for cluster membership. Following discussions with CMHA and HCDCD staff, we recoded building type into three categories: (1) multi-family apartments (high-rise, mid-rise, garden, four-plex); (2) houses (single-family, row); and (3) duplexes. Fifteen cases lacked information on apartment type; hence the sample size for the regression analysis was 790 rather than 805.

The building type dummy variables included HOUSE, houses; MULTIFAM, multi-family apartments; and DUPLEX, duplexes. The census block group cluster dummies included POOR, poor neighborhoods; LOWMID, lower-middle-income neighborhoods; MIDINC, middle-income neighborhoods; UPMID: upper-middle income neighborhoods; HIGHINC, high-income neighborhoods; and HIGHEST, highest-income neighborhoods.

In addition, the variable, BEDRMS, represents the number of bedrooms in each unit. Using HOUSE as the reference variable for building type and UPMID as the reference variable for cluster type we created the following regression model for predicting gross rent of a given unit: Gross rent = a * (constant) + b * MULTIFAM + c * DUPLEX + d * POOR + e * LOWMID + f * MIDINC + g * HIGHINC + h *BEDRMS

The initial regression run based on 790 cases accounted for more than three-fifths (62 percent) of the variation in gross rent, a fairly good explanatory result. Not surprisingly, the most important predictor of rents was the number of bedrooms, beta=.752. (We present the beta statistic, the standardized regression coefficient, here because it allows us to compare the strength of the different variables as predictors). Three other factors were significant predictors: living in a multi-family apartment (beta=-.075), living in a poor neighborhood (beta=-.208), and living in a low- to middle-income neighborhood (beta=-.160).

An analysis of residuals identified the "outliers," the ten cases with the largest gap (3.34 standard deviations or more) between the predicted and actual rents. Dropping these ten outliers and rerunning the regression analysis led to an increase in the amount of variation explained (from 63 percent to 67 percent). The same four variables that were significant predictors in the initial regression run remained significant in the second one as well. The unstandardized regression coefficients were used to create the following equation to predict gross rents:

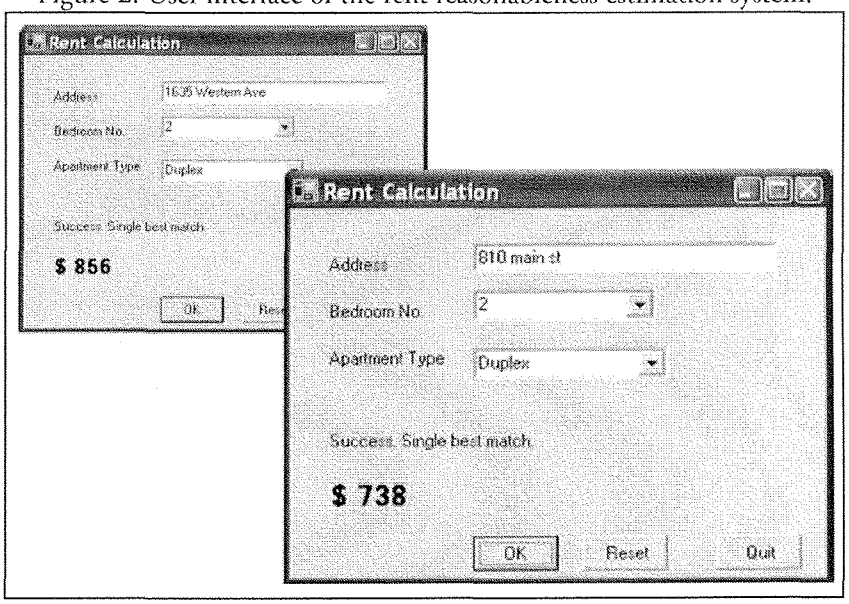
Gross rent = 529.830 – 74.268MULTIFAM – 34.951DUPLEX – 118.643POOR –92.102LOWMID – 48.657MIDINC + 24.411HIGHINC + 180.639BEDRMS

EVALUATING OF THE CMHA/HCDCD 2005 RENT REASONABLENESS ESTIMATION METHODOLOGY

The CMHA/HCDCD 2005 Rent Reasonableness Estimation

In early 2005 we notified CMHA and HCDCD that we were interested in helping them to load the regression model on to mid-capacity portable handheld computers that inspectors could use on-site. A graduate student developed a regression model system for predicting gross rent. After a staff person keys in an address, the apartment type and the number of bedrooms, the system would respond by indicating an estimated reasonable rent for that property. Figure 2 shows two snapshots of the rent estimation system. As shown, two units of the same size and apartment type may have different gross rents when they are in different clusters.

Figure 2: User interface of the rent reasonableness estimation system.



The system provides the base for reasonable rent estimation. That is, if the property were in poor condition, or if it had unusual amenities, an inspector may make adjustments in the estimates.

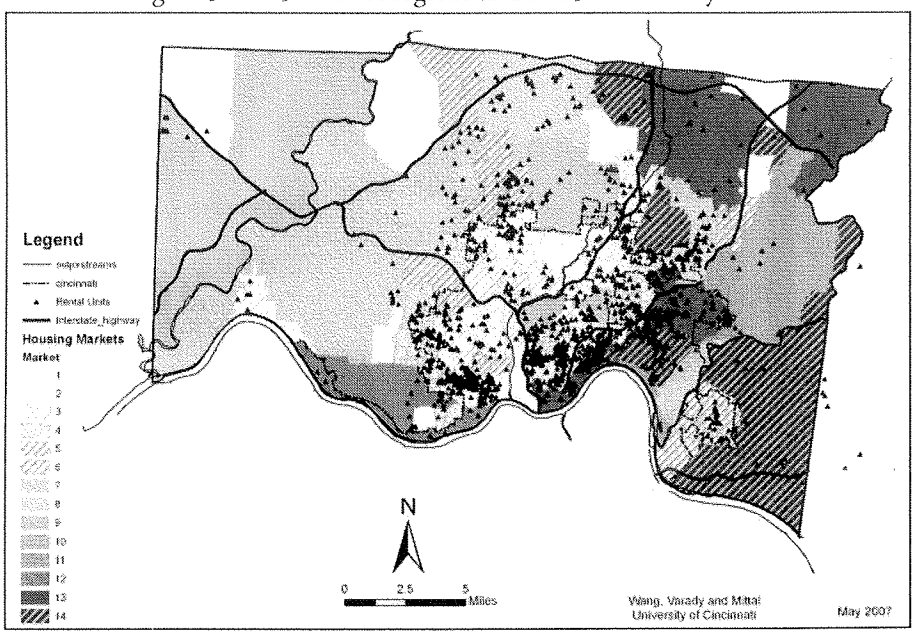
Shortly after our communication with CMHA and HCDCD, CMHA bought a new software system for handling its housing voucher client data. According to a CMHA staff member, the new system precluded the use of the regression model that we had developed. Instead the new software led to the re-introduction of housing market regions, the flawed approach that had led to our original involvement in

the rent reasonableness project.

CMHA and HCDCD contracted with a consulting firm to develop a rental market database to be used in conjunction with the new software to determine rent reasonableness for units proposed for or assisted under the tenant-based Section 8 program. The consultant developed a list of Hamilton County rental properties from a variety of primary and secondary data sources. Nearly 6,000 telephone calls were made to achieve a data base of 2,698 properties. An advantage of this data base was that it included a larger number of single-family houses than we were able to identify for our data base. Since this was a "blind study" callers did not disclose the ultimate purpose of their phone calls. Callers asked about rental costs, utilities and amenities included in the rental costs. Once the data was gathered and validated, it was geo-coded for census tract, census block group, and longitude and latitude coordinates (Merusi Partners Inc., 2005).

Based on the rent survey data and the consultant's analysis of trends in values across the county, CMHA and HCDCD staff aggregated 206 of the 230 census tracts in Hamilton County into 14 housing markets based on rental rate, bedroom size, location of the property surveyed, and similar demographic characteristics of the census tracts. Figure 3 shows the distribution of the 2,698 cases across the 14 housing market regions. Twenty-four census tracts were not included in the 14 housing markets because of the absence of information on rental units.

Figure 3: 2005 Market regions and 2005 rent survey units.



After the housing market regions were identified, the housing authorities used a table to estimate the reasonable rent for a given rental unit. For illustrative purposes, Table 3 shows a part of the table. For a rental unit in question, the housing inspector first must identify the census tract the unit is located in based on its street address (which requires additional data processing not reflected in the table). Then the housing inspector identifies the market region based on the table. Together, the housing market region and bedroom number determines the reasonable rent for the unit. CMHA's methodology also included a plus or minus 12 percent adjustment system—to be implemented by a housing inspector on-site—to modify the initial reasonable rent estimate to take into account characteristics of the apartment, the building and the surrounding area (see Merusi Partners Inc., 2005).

Table 3: An excerpt of the rent reasonable estimation table adopted by CMHA

No. of Bedrooms	Market 1	Market 2	Market 3	Market 4
0 Bedroom	$351	$360	$370	$379
1 Bedroom	$419	$429	$440	$452
2 Bedrooms	$565	$580	$595	$610
3 Bedrooms	$734	$753	$772	$792
4 Bedrooms	$770	$790	$810	$831
5 Bedrooms	$883	$906	$929	$953
6 Bedrooms	$1,001	$1,027	$1,053	$1,080
7 Bedrooms	$1,117	$1,145	$1,175	$1,205
Census Tracts	Census Tract MA 1	Census Tract MA 2	Census Tract MA 3	Census Tract MA 4
	67	80	22	29
	68	81	23	30
	69	82.01	25	32
	73	82.02	27	33
	77	83	74	59
	85.02	84	75	99.01
	86.01	91	78	99.02
	87	100.02	79	102.01
	88	102.02	85.01	107

Source : CMHA, 2005

Here is an example to illustrate CMHA's chosen approach, for estimating reasonable rent for a three bedroom unit located in the census tract 100.02. From the bottom of the table, we can see that the census tract 100.02 is part of the housing market region 2. The reasonable rent for a three bedroom in housing market region 2 is $753. It should be pointed out that building type is not considered directly in calculating the reasonable rent.

According to CMHA's consultant, the existence of micro-markets, small niche markets that skew values in certain census tracts, necessitated expanding the number of market regions from 10 to 14 (Merusi Partners, Inc., 2005). We believe that the addition of four more market regions did not solve the problem of housing market regions being too large and too heterogeneous. Our working hypothesis is that our regression modeling approach which is based on census block group clusters can produce more accurate estimates than the tabular method that CMHA adopted. The remainder of this paper tests this hypothesis.

Construction and comparison of reasonable rent regression models

In Fall 2005, we requested and then received the consultant's rental market survey dataset. In addition to the market region and number of bedrooms, the dataset includes gross rent, address, census tract, building age, and building type. The consultant had retrieved building type data from the Hamilton County Assessor's website using the following categories:

401: Commercial - Apartments, 4 To 19 Units
402: Commercial - Apartments, 20 To 39 Units
403: Commercial - Apartments, 40+ Units
510: Residential - Single Family
520: Residential - Two Family Dwellings
530: Residential - Three Family Dwelling

We used the 2005 data set to construct two regression models, one with the six clusters as we defined earlier and the other with the 14 market regions defined by CMHA/HCDCD staff. The common independent variables between the two models were the number of bedrooms, building type, and building age. As with the 2004 study we converted housing market regions and building types into two sets of dummy variables. We used "510: Residential—Single Family" as the reference building type. For the Cluster regression model, we used "high income" cluster as the reference cluster. There are no rental properties found in the "highest income" cluster. For the Market region regression model, we used "Market Region 14" as the reference housing market region. Tables 4 and 5 portray the results from the two regression models.

In general, the two regression models performed well. The model based on housing market regions accounted for more of the variation in the rents than that of the cluster model (57 percent versus 53 percent). The geographic variables added to the explanatory power of both models, adding an additional seven percent in the case of the 14 housing market regions and an additional three percent in the case of the 5 clusters. Compared to the most affluent cluster (Cluster 5), rents in other clusters are significantly lower. The results for the market regions model are broadly similar, that is, all of the markets had lower rents than market 14. Since the latter had the highest average rents this is hardly a surprising result. The results do show that rents in market regions 10 and 11 are not significantly different from the reference region

Table 4: Regression results: Housing and neighborhood characteristics affecting gross rents (clusters of census blocks).

	Without geographical variables					With geographical variables				
Variable	B	St. Error	t	Beta	Sig.	B	St. Error	t	Beta	Sig.
4 to 19 unit apartments[a]	-221.491	14.109	-15.698	-.361	.000	-207.426	13.929	-14.892	-0.338	.000
20 to 39 unit apartments	-.177.918	28.224	-6.304	-.096	.000	-153.538	27.783	-5.526	-0.083	.000
40 unit apartments or more	-90.595	15.001	6.039	-.135	.000	-86.904	14.754	-5.89	-0.13	.000
Two family dwellings	-115.023	13.181	-8.726	-.202	.000	-110.500	13.011	-8.493	-0.194	.000
Three family dwellings	-.144.559	16.842	-8.583	-.160	.000	-136.946	16.477	-8.311	-0.152	.000
Building age	-.381	.149	-2.551	-.045	.011	-.169	0.148	-1.146	-0.02	NS
Number of bedrooms	159.783	4.480	35.662	.560	.000	158.369	4.411	35.914	0.555	.000
Lowest income cluster[b]						-.196.267	17.882	-10.976	-0.347	.000
Lower income cluster						-.171.272	17.322	-9.888	-0.324	.000
Middle income cluster						-89.433	19.897	-4.495	-0.104	.000
Upper middle income cluster						162.506	31.882	-5.097	-0.081	.000
Constant	549.279	19.289	28.476		.000	692.313	25.367	27.291		.000
F	374.014					268.225				
df	2633					2629				
Significance	.000					.000				
Adjusted R^2	.497					0.527				

Notes: a. Single family building was the reference category

b. Cluster 5, the high income one, was the reference category

Table 5: Regression results: Housing and neighborhood characteristics affecting gross rents

	Without geographical variables					With geographical variables				
Variable	B	St. Error	Beta	t	Sig.	B	St. Error	t	Beta	Sig.
4 to 19 unit apartments[a]	-.221.491	14.109	-.361	-.15.698	.000	-206.474	13.276	-15.552	-0.337	.000
20 to 39 unit apartments	-.177.918	28.224	-.096	-.6.34	.000	-157.883	26.415	-5.977	-0.085	.000
40 unit apartments or more	-90.595	15.001	-.135	-.6.039	.000	-123.045	14.234	-8.645	-0.184	.000
Two family dwellings	-.115.023	13.181	-.202	-8.726	.000	-111.299	12.442	-8.946	-0.196	.000
Three family dwellings	-.144.559	16.842	-.160	-8.563	.000	-143.945	15.737	-9.147	-0.16	.000
Building age	-.381	.149	-.045	-2.551	.011	-.622	0.148	-4.2	-0.073	.000
Number of bedrooms	159.783	4.480	.560	35.662	.000	158.056	4.183	37.782	0.554	.000
Market 1[b]						-241.410	13.304	-18.145	-0.406	.000
Market 2						-249.055	20.588	-12.097	-0.191	.000
Market 3						-164.501	19.065	-8.628	-0.136	.000
Market 4						-199.026	14.878	-13.377	-0.258	.000
Market 5						-230.688	27.014	-8.54	-0.122	.000
Market 6						-184.166	16.104	-11.436	-0.2	.000
Market 7						-189.876	24.975	-7.603	-0.11	.000
Market 8						-168.713	17.87	-9.441	-0.158	.000
Market 9						-104.921	18.722	-5.604	-0.09	.000
Market 10						-47.881	44.112	-1.085	-0.014	NS
Market 11						-48.880	30.986	-1.577	-0.022	NS
Market 12						-224.864	34.565	-6.506	-0.089	.000
Market 13						-131.989	14.945	-8.831	-0.166	.000
Constant	549.279	19.289		28.476	.000	744.677	22.18	33.574		.000
F	374.014					173.002				
df	2633					2620				
Significance	.000					.000				
Adjusted R^2	.497					.566				

Notes: a. Single family buildings was the reference category

b. Market 14 was the reference category

—market region 14. These results imply that the CMHA/HCDCD scheme may include more housing markets than are needed.

For both models the number of bedrooms proved to be the strongest predictor of rents and in both, multi-family units rented for less than single-family units when all other relevant characteristics were taken into account. It is difficult to understand why building age had a significant negative impact on rents for the housing market region model but not for the cluster one. However, it is possible in the case of the large housing market regions, building age served as a proxy for other neighborhood characteristics.

Comparison of model estimations and surveyed data

To validate the performance of our two models we applied the two regression models developed in the previous section to the 2004 rental survey data (i.e., 274 rental units with utility data). In addition, we calculated rents for the same 274 units following the CMHA tabular method that is, using the market region and bedroom size alone to predict rent. We next calculated and compared the differences between each of the predicted rents and the actual rent for the same rental unit.

The comparison of the average differences for the three prediction methods clearly demonstrates the advantage of the regression modeling approach (The detailed results are not included here but are available from the authors). On average, the market region regression model predicted rent was $16 lower than the actual rent while the cluster regression model was lower by $17. On the other hand, the tabular approach produced errors that were six times as large (i.e., on average $94 lower than the actual rent).

The preceding highlights the value of the regression modeling approach in estimating reasonable rents. Both of the regression estimates are superior to the estimates produced by the market tabular approach. The fact that the market region model performs similarly to the cluster approach might at first glance seem to provide justification to the CMHA for continuing to use the market region strategy. The next section shows why such a choice would be a mistake.

Comparison of clusters and market regions

CMHA's current methodology for computing reasonable rents, the "market tabular approach," is based on the assumption that the 14 housing market regions are relatively homogeneous. After all, CMHA's stated reason for expanding the number of markets from 10 to 14 was to create separate markets for what had previously been micro-markets or niches in the 10 larger markets.

As discussed earlier, the 731 census block groups formed six clusters based on similarity with respect to five census variables: percent African American, percent attending college, median household income, proportion of households in poverty, and proportion of homeowners. Therefore, if the 14 market regions are indeed ho-

mogeneous, they should present a good fit with the six clusters. That is, each market region should exclusively, or almost exclusively, fall in one cluster. However, four of these one cluster markets comprised between half and three fifths of the total area of the cluster.

Table 6: Summary of areas by cluster and market region (% of clusters).

Area (% of Row)	Cluster							
Market	Outside	1	2	3	4	5	6	Grand Total
Outside	5.0%	0.5%	14.2%	21.2%	5.9%	53.1%	0.0%	100.0%
1	2.0%	48.4%	41.2%	8.4%	0.0%	0.0%	0.0%	100.0%
2	0.0%	56.8%	30.3%	9.1%	0.0%	3.8%	0.0%	100.0%
3	0.0%	38.4%	61.6%	0.0%	0.0%	0.0%	0.0%	100.0%
4	0.0%	19.2%	56.3%	12.3%	0.0%	12.2%	0.0%	100.0%
5	0.0%	0.5%	23.4%	22.3%	0.0%	53.8%	0.0%	100.0%
6	0.0%	11.7%	46.1%	14.3%	20.3%	7.6%	0.0%	100.0%
7	0.0%	5.9%	25.6%	29.8%	3.9%	34.8%	0.0%	100.0%
8	0.0%	2.2%	29.5%	36.4%	5.9%	26.0%	0.0%	100.0%
9	0.0%	1.7%	47.4%	42.4%	3.9%	4.7%	0.0%	100.0%
10	0.0%	0.0%	53.7%	46.3%	0.0%	0.0%	0.0%	100.0%
11	0.0%	0.0%	10.5%	4.7%	16.6%	16.8%	51.5%	100.0%
12	0.0%	0.0%	47.0%	8.1%	17.7%	27.2%	0.0%	100.0%
13	0.1%	5.2%	39.9%	8.5%	23.9%	22.4%	0.0%	100.0%
14	0.0%	1.3%	22.9%	10.7%	35.3%	28.0%	1.8%	100.0%
Grand Total	0.6%	8.3%	32.7%	21.0%	11.2%	22.8%	3.4%	100.0%

Table 6 shows that this is not the case. Reading across each row one sees the proportion of the area of each market region devoted to each cluster type ("Outside" refers to parts of Hamilton County that were not assigned to any market region or cluster). In general, the results highlight the heterogeneity of the 14 housing market regions. In no case is a market region comprised exclusively (or even nearly exclusively) of one cluster. Most of the market regions are distributed among two or more clusters. Market 14 is the most heterogeneous of all. Thirty five percent of the area is devoted to cluster four, 28 percent to cluster five, 23 percent to cluster two while the remaining area is distributed among clusters 1 and 6.

This heterogeneity also can be seen by comparing the map of clusters (Figure 1) and the map of the housing market regions (Figure 3). The cluster map shows that the clusters do not have much geographical connectivity. That is, two adjacent block groups often have completely different socio-economic characteristics and are in different clusters. On the other hand, many block groups that are

located far away from one other share similar socio-economic characteristics and are in the same cluster. Such spatial distribution patterns clearly demonstrate the problem with CMHA's housing market regions.

CONCLUSIONS

When we started the project our aim was to develop a cost-effective approach that the Cincinnati Metropolitan Housing Authority and the Hamilton County Department of Community Development could use to develop reasonable rent estimates at a geographical level below the census tract. We developed such a methodology for doing this based on a combination of cluster and regression analyses.

The project changed unexpectedly when CMHA decided not to use our methodology because of the perceived requirements of its new housing voucher software. At that point our overriding concern became comparing our approach with CMHA/HCDCD's current approach—the tabular approach based on market regions, each of which is a combination of census tracts.

Our regression modeling approach provides more accurate estimates of market rents than CMHA's tabular approach. In addition, our use of small census block groups to build relatively homogeneous clusters represents a significant improvement over CMHA's reliance on larger and more heterogeneous housing market regions. Our clusters correspond more closely to housing submarkets people consider when choosing where to live and when to move (Bates, 2006). Although CMHA has chosen not to use this innovative approach, other American public housing authorities may find it useful.

Epilogue

In Spring 2006, when CMHA issued its rent reasonable estimates based on its new software system landlords protested that allowable rents were too low and that the rent cuts were based on flawed data. In response, CMHA set up a five-member advisory panel consisting of CMHA staff, apartment industry representatives and neighborhood activists "to analyze rents in smaller geographic areas than prior studies" (Monk, 2006, 3). The last quote from a news story in a Cincinnati business newspaper was not exactly correct. We had prepared such an analysis—it was the focus of this paper—but it was ignored. The important thing, however, is that CMHA has finally recognized the need for a finer grain level of analysis than that of housing market regions, to produce accurate and fair reasonable rent estimates.

NOTE

1. Since CMHA's new adjustment system can be applied across the board to any of the three predicted rents we feel it is valid to compare the rent estimates before the adjustment.

APPENDIX 1: Methodology for Calculating Utility Costs

1. We prepared separate calculations of utility costs for different sized apartments (i.e., with apartments varying by number of bedrooms).
2. We added a $6 surcharge to the gas cooking costs.
3. For space heating, cooking, and water heating, we computed the average cost for units using gas and units using electricity and used that average in subsequent calculations.
4. We separated the homes into two categories: "single family" and "others." For the single-family category, we used the CMHA and HCDCD space heating allowances for single-family detached houses. For the "other category" we used the average of the "high rise" and "townhouse" allowances categories.
5. Where there was no data on who paid for water, we assumed that single-family tenants paid for water themselves while the landlord paid for water for homes in "the other" category.
6. Where there was no data on who paid for electricity, we assumed that the tenant paid for it.
7. Where there was no data on whether the space heating was gas or electric, we assumed gas heating was used. Where there was no data on who paid for the space heating, we assumed single-family tenants paid for space heating. There were 83 units in this category. For all other units, we assumed that if the tenant paid for the hot water, she paid for the space heating as well. Otherwise, we assumed that the landlord paid for space heating.
8. Where there was no data on whether the hot water was gas or electric heated we assumed the energy source was the same as was the case for space heating. That is, if heating was by gas we assumed that hot water was by gas and if heating was electric we assumed that hot water was by electricity too. Where there was no data on who paid for hot water, we assumed that whoever paid for space heating also paid for hot water as well.
9. Where there was no data on whether the cooking was by gas or electricity we used the information on how space heating was provided. Where there was no data on who paid for cooking, we assumed that the tenant paid for these costs.

REFERENCES

Bates, L. (2006) Does neighborhood really matter? Comparing historically defined neighborhood boundaries with housing submarkets. *Journal of Planning and Education Research*, 26:5-17.

Cincinnati Metropolitan Housing Authority (2005) Housing Choice Voucher Program Housing Allowances for Tenant Furnished Utilities and Other Services. Cincinnati.

HUD (U.S. Department of Housing and Urban Development) (2001) *2001* Voucher Program Guidebook. Washington, DC.

Knox, P. and McCarthy, L. (2005) *Urbanization: An Introduction to Urban Geography, 2nd Edition.* New York: Random House.

Merusi Partners Inc. (2005) HCV Program 2005 Rent Reasonableness Study. PowerPoint presentation to the Cincinnati Metropolitan Housing Authority and the Hamilton County Department of Community Development, November 18.

Monk, D. (2006) Housing authority panel, consultant to review rent. *Cincinnati Business Record*, May 19, p. 3.