The Shape of Mobility: Measuring the Distance Decay function of household mobility

John R. Hipp*

Adam Boessen

September 23, 2011

DRAFT: Do not quote or cite without permission from the author

Word count: 3,396

Word count (including references): 3,630

Running Head: "Distance of moves"

* Department of Criminology, Law and Society and Department of Sociology, University of California, Irvine. Address correspondence to John R. Hipp, Department of Criminology, Law and Society, University of California, Irvine, 3311 Social Ecology II, Irvine, CA 92697; email: john.hipp@UCI.edu. This research is supported in part by NSF grant BCS-0827027

[1] The research in this paper was conducted while the first author was a Special Sworn Status researcher of the U.S. Census Bureau at the Triangle Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau. This paper has been screened to ensure that no confidential data are revealed.

The Shape of Mobility: Measuring the Distance Decay function of household mobility

Abstract

This study uses a unique dataset that provides information on the distance of the most recent move for a large sample of households in 22 metropolitan areas over three waves. With this information, we flexibly estimate the distance decay function for these moves for the entire sample. We also estimate the distance decay function for a series of subpopulations based on key demographic information. These bivariate estimates allow us to estimate the entire distance decay functional form for these subpopulations. Finally, we estimate multivariate models to assess the extent to which various subpopulations differ from the general population in their distance decay function, controlling for other characteristics.

Keywords: residential mobility, distance decay, gravity model,

Bio

John R. Hipp is an Associate Professor in the department of Criminology, Law and Society, and Sociology, at the University of California Irvine. His research interests focus on how neighborhoods change over time, how that change both affects and is affected by neighborhood crime, and the role networks and institutions play in that change. He approaches these questions using quantitative methods as well as social network analysis. He has published substantive work in such journals as *American Sociological Review, Criminology, Social Forces*, *Social Problems, Mobilization, City & Community, Urban Studies* and *Journal of Urban Affairs*. He has published methodological work in such journals as *Sociological Methodology*, *Psychological Methods*, and *Structural Equation Modeling*.

Adam Boessen is a doctoral student in the Department of Criminology, Law and Society at the University of California - Irvine. His primary research interests include the community of context of crime, spatial analysis, social network analysis, and juvenile delinquency. His work uses quantitative methodologies to examine the relation between residential mobility and crime, the measurement and conceptualization of neighborhoods, and the impact of incarceration on juvenile offenders.

The Shape of Mobility: Measuring the Distance Decay function of household mobility

Demographers have long been interested in the question of how far households tend to travel when engaging in residential mobility. Studies have explored this question using different data structures, and different data analytic techniques, providing key insights. Nonetheless, a challenge for answering this question is that one needs information on all moves by households. Studies are frequently limited to information on only intra-metropolitan moves, or inter-metropolitan moves, which precludes estimating the entire distance decay function to capture the distance of moves over all possible moves.

This long line of demography literature focused on the question of the physical distance of moves by households when they engage in residential mobility has established quite clearly that households are more likely to move shorter distances rather than longer distances. Studies have shown this by focusing on the relative likelihood of moves within a metropolitan area versus inter-regional moves, and showing that within-metropolitan area moves are relatively more frequent. Indeed, within-neighborhood moves are relatively more frequent than expected by random chance, also emphasizing the importance of distance as a constraining force in such mobility decisions. Another body of literature has estimated the distance decay function of moves by households within a particular metropolitan region. A well-known challenge to such latter studies is that their estimate of the distance decay function is truncated because 1) they are only capturing moves within the metropolitan area (and therefore are unable to estimate the tail of the general residential mobility function given that they, by definition, have no information on such longer moves), and 2) there is a natural truncation in the estimated functional form given the particular spatial distribution of households in the city or metropolitan area.

In this paper, we extend this literature by using a sample of households in 22 different metropolitan areas over three time points to estimate the distance function captured by the most recent move by the household.

Mobility literature

Scholars have studied the physical distance of moves by households in several fashions. We next briefly describe this literature.

Inter-metropolitan mobility

Numerous studies have focused on the aggregated flows of households in an effort to estimate the distance of residential moves. This literature has largely built on gravity flow models: the notion that distance has important effects inhibiting mobility. These studies have focused on both inter-metropolitan area mobility flows, as well as intra-urban mobility flows (Haynes and Fotheringham 1984). These studies have consistently shown the importance of the gravity flow model: that is, households are more likely to move to closer destinations rather than further ones. For example, one study looked at aggregate flows between metropolitan areas in the 1930's and detected such an effect (Anderson 1956). Studies have also viewed mobility flows between states (Plane and Mulligan 1997) and inter-metro area mobility flows (Galle and Taeuber 1966).

Although early studies focused almost exclusively on the gravity flow model and the importance of physical distance, later work increasingly incorporated other possibly important characteristics of metropolitan areas that might influence migration destinations (Ferguson and Kanaroglou 1995). Such characteristics as amenities, and economic opportunities, were tested for their effect on mobility flows. For example, one study extended the gravity flow model by including various sociocultural characteristics of the destination as predictors of direction of mobility for inter-state flows (Herting, Grusky, and Van Rompaey 1997).

Intra-metropolitan mobility

Another body of literature has focused on the mobility decisions of individual households. This literature has generally focused on mobility flows within a particular city or metropolitan area. For example, one study looked at the distance of intra-metropolitan area moves in Cedar Rapids, IA (Brown, Horton, and Wittick 1970), whereas another study studied such moves within Milwaukee (Clark 1976). Although a strength of such studies is that they can estimate more precisely the distance decay function of moves, they nonetheless are estimating a constrained function. In part, this distance decay function is constrained only to intra-meteropolitan moves, and cannot estimate the distance decay function of longer moves. Instead, such studies are forced to interpolate beyond the data to estimate the form of the distance decay function. Second, these studies are constrained to the actual size of the metro area and the specific peculiarities of how the population is distributed spatially (Clark 1986). That is, the presence of bodies of water, or mountains, that create open spaces without any population cause problems for estimating a parametric distance decay function.

Studies of intra-metropolitan moves have found that there are differences in distance moved based on the characteristics of the household. For example, there is evidence that blacks engage in shorter moves (Clark 1986).

An important consequence is that the specific shapes of metropolitan areas can impact distance decay functions, sometimes in manners that can be predicted analytically (Taylor 1971). That is, the spatial footprint of the population in an area constrains the possible mobility choices of households, which then will strongly affect the possible shape of the distance decay function that is estimated for households.

Distance decay functions

There are a number of distance decay functions that can be estimated. Morrill and Pitts (Morrill and Pitts 1967) discussed four for various types of flows: Pareto, exponential, lognormal, and Pareto-exponential. They argued that existing evidence suggested that exponential functions seemed more appropriate for migrations and marriage distances. They then displayed examples of fitted curves for a handful of cities (Morrill and Pitts 1967). In this study we flexibly estimate the distance decay functional form using multiple polynomials, rather than attempting to adjudicate between these particular parametric forms.

Although the existing evidence of inter-regional mobility flows and intra-metropolitan studies is that distance has a strong diminishing effect on mobility probabilities (Clark 1986), there are important limitations. Notably, existing literature has not been able to estimate this entire distance decay function. Studies focusing on intra-metropolitan flows are only able to estimate this functional form at the shorter end of the mobility distance scale. These studies are forced to extrapolate beyond their data to assess the effect of longer distances on mobility flows. And studies looking at inter-metropolitan area flows are able to get estimates of the long tail of such mobility flows, but again must extrapolate to shorter moves. It is an open question whether the functional form estimated for short moves would in fact be appropriate for long moves. Likewise, it is not clear that the functional form estimated for long moves would in fact fit for short moves.

We are able to explore this question here given that we have a sample of households from a large number of metropolitan areas, and we have information on the distance of their most recent move no matter where they were moving from in the entire United States. We describe the data next.

Data and Methods

Data

The American Housing Survey (AHS) conducts surveys of about 4,000 housing units from each of a large number of metropolitan areas across the U.S in various years. Every two years the AHS surveys a subset of the metropolitan areas: as a result, a particular metropolitan area is surveyed approximately every four years. Because of this variability in the actual year of the survey, we are sometimes combining metropolitan areas from slightly different years. That is, whereas the "waves" are labeled 1987, 1991, and 1995, these "waves" actually contain the data for the nearest year in which a particular metropolitan area was surveyed. For instance, whereas in the wave 1987 some of the metropolitan areas were actually surveyed that year, some of the metropolitan areas were actually surveyed in 1985. We have a very large sample of households in 23 metropolitan areas in 1987, 22 in 1991, and 23 in 1995. Pooling across these 68 metropolitan years, we have 66,383 tract years.[1]

By using special access to a Census Research Data Center, we were able to place AHS residents into 1980 census tracts. Thus, we knew the current tract of residence. The AHS asks respondents to report on the zip code of their previous residence. Thus, we have geographic

[1] The metropolitan areas are: Anaheim, Baltimore, Buffalo, Chicago, Cleveland, Dallas, Denver, Fort Worth, Houston, Indianapolis, Los Angeles, Milwaukee, Minneapolis, Oakland, Philadelphia, Phoenix, Portland OR, Riverside-San Bernardino, Sacramento, Saint Louis, San Diego, San Francisco, Seattle, Tampa, Washington DC

"containers" of the origin and destination of the most recent move for each respondent in the sample. These are not ideal measures, and therefore pose specific challenges. First, zip codes are not ideal geographic containers given that they were created by the Postal Service for the express goal of delivering mail, and therefore do not necessarily map onto the concept of "neighborhood". Furthermore, zip codes can change boundaries quite readily, with minimal documentation of the changes over time. As a best approximation of the zip code boundaries over the time period of our study, we used zip code boundaries from 1991.[2] From these boundary files, we computed the latitude/longitude of the center point of each of these zip codes. From the same source, we obtained 1980 tract boundaries, and computed the latitude/longitude of the center point of each of these tracts.

Dependent Variables

We computed the distance in miles between the centroid of the tract of residence and the zip code that the household reported moving from. We log transformed this value.

Independent variables

In the initial analyses in which the outcome was logged distance of the previous move, we created a series of standard demographic measures capturing information on the household head or the household in general to assess their relationship with move distance. We created indicators of whether the household head is African American, Latino, Asian, or other race (with white as the reference category). We computed a measure of years of education of the household head, and a measure of household income (logged). We capture the distinction between owners and renters with an indicator of owners. We created measures of age of household head, and age squared, to capture possible nonlinearities in distance moved over the life course. We capture

[2] These were obtained from the MABLE/Geocore website located at the Missouri Census Data Center website (http://mcdc2.missouri.edu/websas/geocorr90.shtml).

household composition with indicators of currently married, widowed, or divorced (with never married as the reference category), and indicators of whether the household has children less than school aged (less than 6 years of age, but no school-aged children), whether the household has school aged children (aged 6 to 18 years), with the reference households without children under the age of 18.

In the models capturing the distance distribution, we first ordered the sample from shortest to longest distance, and then created an integer indicator ranging from 1 to the sample size based on this ordering (thus, the household with the shortest distance move is coded 1, the household with the second shortest distance move is coded 2, on up to the household with the longest distance move which is coded to the sample size value). We used the same procedure for each of the sub-samples on which we computed the distance distribution. We also created various polynomials for these integer indicators to capture nonlinearities in the distance function.

The summary statistics for the variables used in the analyses are presented in Table 1.

<<<Table 1 about here>>>

Methods

In all models, the outcome measure is logged distance of the most recent move. The models are estimated as ordinary least squares regressions. In the models estimating the distance function of the most recent move, the predictor variables are the integer indicator showing the relative distance of the move for a particular household, and as many polynomials of this as are necessary to reasonably capture this distance function.

Results

We first present the summary statistics for the distance of moves for the sample and various sub-populations of the sample. As seen in Table 1, the median distance of moves in this sample is 4.4 miles, whereas the mean distance is 114 miles; this of course implies the unsurprising considerable skew given that some moves can be for extremely long distances, whereas the bulk of moves are for very short distances. For those in poverty, moves are of shorter distance (median 3.3 miles). Also, whites tend to move longer distances (median is 4.9 miles) compared to African Americans and Latinos (a median of about 3.3 miles). Also, owners move much longer distances compared to renters: 6.4 miles compared to 3.7 miles for the median values.

We next turn to the nonlinear models capturing the distance function of the most recent move for the entire sample, and various subsamples. In these models, we only include the ranking indicator, and its various polynomial. Nonetheless, these models do an extremely good job of explaining this distribution. As evidence of this, the R^2's for the models are as follows: .995 for the full sample, .992 for African Americans, .996 for whites, .995 for Latinos, .997 for households in poverty, .994 for owners, .995 for renters, .994 for households within children less than school age, .995 for households with school aged children, .996 for households with no children. Thus, the flexible distance function we are estimating is essentially explaining these patterns.

We plot each of these distance decay functions. For the complete sample, we see in Figure 1 that the log relationship captures the middle range of the distance of moves, but there are bends at the two ends of the distribution. That is, given the slope in the middle of this figure, there are fewer very short distance movers and fewer very long distance movers compared to the proportion in the middle portion of this distribution. Thus, based on these estimates, the 10th percentile of move distances is 0.95 miles, the 20th percentile is 1.68 miles, the 30th percentile is 2.33 miles, the 40th percentile is 3.2 miles, and the median is 4.45 miles. At the other end of the distribution, the 80th percentile is 11.75 miles and the 90th percentile is 37.7 miles. Thus, only about 8.5% of the sample moved 50 or more miles, 5.7% moved 100 or more miles, and 3.3% moved 200 or more miles.

<<<Figure 1 about here>>>

We next focus on the distance distribution of moves for various subpopulations of our sample. We illustrate the distance distribution for those in poverty by overlaying it on the distribution for the entire sample in Figure 1. As expected, we see that those in poverty move shorter distances than the complete sample. At virtually all points of the distribution, households in poverty move shorter distances than those of the entire sample. For example, at the 30th percentile, the typical household in the sample moved 37% farther than a household in poverty (2.33 to 1.7 miles), and this relative gap remains constant up through the 70th percentile. However, this gap widens for the longest moves: at the 80th percentile this difference is 56% farther, and at the 90th percentile it is 127% farther. It appears households in poverty are particularly unlikely to make the longest moves---those that are inter-regional.

Turning to the effect of race/ethnicity, we see that Latinos and African Americans move shorter distances than the entire sample. Figure 2 compares the distance distributions of Latinos and African Americans to the complete sample, and shows that African Americans move shorter distances than the complete sample at all points in the distribution (the line depicting their distance distribution is always below that of the complete sample). At the 30th percentile, and average household in the sample moves 59% farther than does an African American household. Although this gap falls to 38% and 34% at the 50th and 70th percentiles, it rises to 73% and 123% at the 80th and 90th percentiles. The pattern is relatively similar for Latinos, as they always move shorter distances than the complete sample except for those moving the most extremely long distances, who actually match and slightly exceed the distance moved among long distance movers in the entire sample. At the 30th percentile, the average household in the sample moved 48% farther than a Latino household. This gap fell to 34% at the 50th percentile, 27% at the 70th percentile, and 22% at the 90th percentile.

<<<Figure 2 about here>>>

When comparing households based on the presence of children, there are few differences for short distance moves (see Figure 3). It appears that an equal proportion of households with very young children, school aged children, or no children, move similar distances among the 2/3 shortest distance moves. However, there are differences for longer moves: those with very young children move considerably shorter distances than those with school aged children or no children. By the 80th percentile, households with children 6 to 18 years old move about 58% farther than households with children 5 and younger, and this gap grows to 134% at the 90th percentile. Households without children move about 58% farther than households with children 5 and younger at the 80th percentile, and over twice as far at the 90th percentile.

<<<Figure 3 about here>>>

There are also consistent differences in the distance of moves for owners and renters at all distances in the distance distribution, as seen in Figure 4. Although owners are less likely than renters to move, when they do so, they move longer distances. Owners tend to move 60 to 80% farther than renters at all points in the distribution, with the only narrowing of the gap occurring at the 5% longest moves.

<<<Figure 4 about here>>>

Multivariate results

We next assessed the partial effects of these various socio-demographic characteristics on the distance of moves, holding constant these demographic characteristics. The results of this model are shown in Table 2. Given that the outcome is logged distance, we can interpret these coefficients in terms of percentage changes in the distance of moves. Holding these demographic characteristics constant, an African American household moves about 50% less far than does a white household (the reference category). Whites in general move the longest distances, as Latinos move 32% less far than whites, Asians move 25% less far, and other race households move 28% less far.

<<<Table 2 about here>>>

Holding constant these characteristics, higher SES households move greater distances, consistent with findings from prior research. Each additional year of education increases the distance moved about 8%, whereas each additional $1000 in income (CHECK THIS) increases the distance moved almost 10%. Although prior studies nearly always show that renters move more frequently than owners, we see evidence that renters move shorter distances than owners (on average, about 7% shorter distances, holding constant these other characteristics).

The effect of age on the distance of moves is u-shaped. At the youngest and oldest ages, households are most likely to move the longest distances. The inflection point is at 37 years of age, suggesting that households with a 37 year old household head move the shortest distances, on average. For example, a 20 year old moves, on average, about 6% farther than a 37 year old, holding constant these other household characteristics. A 60 year old moves, on average, about 12% farther than a 37 year old, holding constant these other household characteristics.

We also see important differences for household structure. Although prior studies have frequently shown that single person households tend to be quite mobile, we see that when they do move, married households tend to move much farther than single person households (about 30% farther, on average, holding constant these other demographic characteristics). Widowed households also move farther than single person households (about 8% farther, on average). However, households with children move shorter differences than those without children. Households with school-aged children (aged 6 to 18 years) move about 7% shorter distances on average than households without children, and households with very young children (less than 6 years of age) move about 15% shorter distances than households without children.

Conclusion

We have extended the literature on the distance traveled by households when changing residence. Whereas prior literature often is constrained to assessing this distance decay function either on short moves only (within a metro area) or on long moves only (inter-regional moves), we were able to estimate the distance of household moves over the complete population of moves within the U.S.

We were able to assess the distance decay of mobility for various subpopulations. We saw that racial/ethnic minorities, and households living in poverty, move the shortest distances. This effect was found both in bivariate and multivariate models. We showed that these can be quite substantial effects, especially at the longest distances.

We acknowledge some limitations of this study. Because we had no information on the prior location of residence for those who immigrated from another country, we were unable to estimate the extreme ends of the distance decay distribution. We did not account for the recency of the move; instead, we only utilized information on the most recently reported move. Furthermore, as discussed, the imprecision of the geographic containers that the household lives in and came from (tracts and zip codes) introduce considerable measurement error into our estimates of the short distance end of this functional form.

In conclusion, understanding the complete distance decay function for households, as well as subsets of households, provides key information for understanding various demographic processes. Understanding the specific shape of this functional form for specific subpopulations provides unique information regarding the character of mobility for these groups beyond measures assuming proportional differences in mobility distance.

References

Anderson, Theodore R. 1956. "Intermetropolitan Migration: A Correlation Analysis." *American Journal of Sociology* 61:459-462.
Brown, Lawrence A., Frank E. Horton, and Robert I. Wittick. 1970. "On Place Utility and the Normative Allocation of Intra-Urban Migrants." *Demography* 7:175-183.
Clark, W. A. V. 1976. "Migration in Milwaukee." *Economic Geography* 52:48-60.
—. 1986. *Human Migration*. Beverly Hills: Sage.
Ferguson, Mark R. and Pavlos S. Kanaroglou. 1995. "Utility Variability within Aggregate Spatial Units and its Relevance to Discrete Models of Destination Choice." Pp. 243-269 in *New Directions in Spatial Econometrics*, edited by L. Anselin and R. J. G. M. Florax. New York: Springer.
Galle, Omer R. and Karl E. Taeuber. 1966. "Metropolitan Migration and Intervening Opportunities." *American Sociological Review* 31:5-13.
Haynes, Kingsley E. and A. Stewart Fotheringham. 1984. *Gravity and Spatial Interaction Models*, Edited by G. I. Thrall. Beverly Hills: Sage.
Herting, Jerald R., David B. Grusky, and Stephen E. Van Rompaey. 1997. "The Social Geography of Interstate Mobility and Persistence." *American Sociological Review* 62:267-287.
Morrill, Richard L. and Forrest R. Pitts. 1967. "Marriage, Migration, and the Mean Information Field: A Study in Uniqueness and Generality." *Annals of the Association of American Geographers* 57:401-422.
Plane, David A. and Gordon F. Mulligan. 1997. "Measuring Spatial Focusing in a Migration System." *Demography* 34:251-262.
Taylor, Peter J. 1971. "Distances within Shapes: An Introduction to a Family of Finite Frequency Distributions." *Geografiska Annaler. Series B, Human Geography* 53:40-53.

Tables and Figures

Table 1. Summary statistics for distance of moves for various subsamples

				Percent moved to same tract (3 different estimates)		
	Median distance	Mean distance	Std. Dev distance	Upper bound	Proportion overlap	Proportion overlap 2
Full sample	4.4	114.2	422.9	21.5%	16.1%	2.3%
Poverty	3.3	90.8	374.8	23.4%	18.2%	2.3%
White	4.9	128.3	445.6	18.0%	13.3%	2.1%
Black	3.2	65.9	330.9	24.3%	18.6%	1.8%
Latino	3.3	60.3	278.1	21.2%	17.3%	2.1%
Owner	6.4	116.8	421.5	16.0%	11.3%	2.1%
Renters	3.7	112.4	423.9	21.6%	16.8%	2.1%
Bachelors degree or more		166.8	513.7	15.6%	11.1%	1.8%
More than HS degree, but not bachelors		115.3	427.5	18.0%	13.5%	2.0%
High school degree		90.4	370.9	20.6%	15.6%	2.3%
Less than HS degree		61.8	294.1	25.4%	20.3%	2.3%
Perceive crime	3.6	93.5	386.6	23.4%	17.8%	2.3%
Do not perceive crime	5.1	128.3	445.1	16.5%	12.3%	1.9%

Table 2. Outcome is logged distance of most recent move

	Coef	SE	
Black	-0.504	-(21.49)	**
Latino	-0.323	-(11.75)	**
Other race	-0.276	-(3.36)	**
Asian	-0.249	-(6.22)	**
Years of education	0.079	(26.53)	**
Owner	0.071	(3.99)	**
Household income	0.097	(6.31)	**
Married	0.301	(13.60)	**
Widowed	0.082	(1.98)	*
Divorced	-0.007	-(0.28)	
Age (X 100)	-1.666	-(5.31)	**
Age squared (X 100)	0.023	(6.76)	**
Only children less than 6 years of age	-0.149	-(6.04)	**
Presence of children 6 to 18 years of age	-0.074	-(3.83)	**
Wave 2 indicator	0.053	(2.92)	**
Wave 3 indicator	0.076	(4.18)	**
Intercept	0.961	(13.52)	**

*** p < .01 (two-tail test), * p < .05 (two-tail test), † p < .05 (one-tail test). T-values in parentheses.*

Figure 1. Logged distance of most recent move for full sample and those in poverty

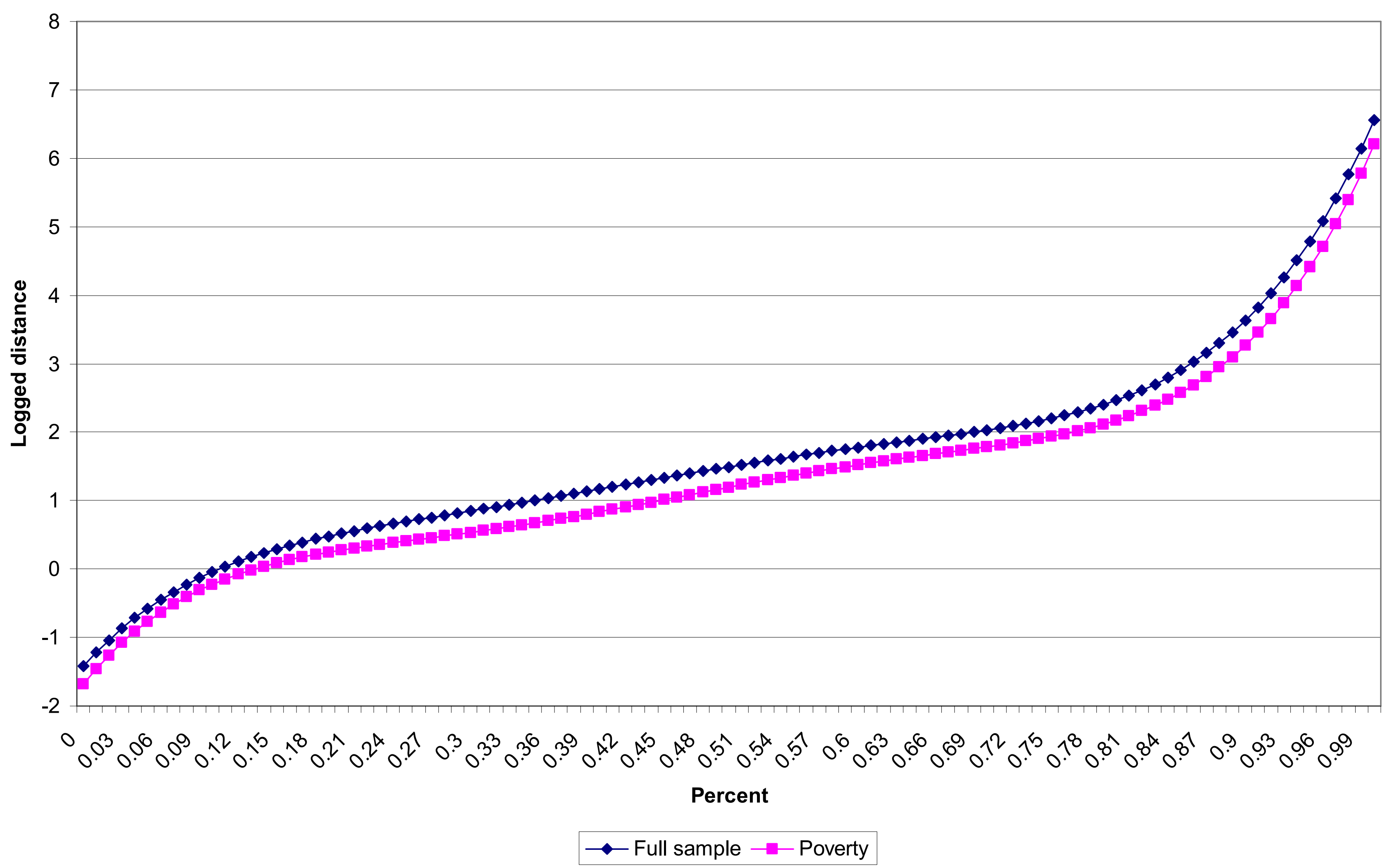

Figure 2. Logged distance of most recent move for blacks, Latinos and full sample

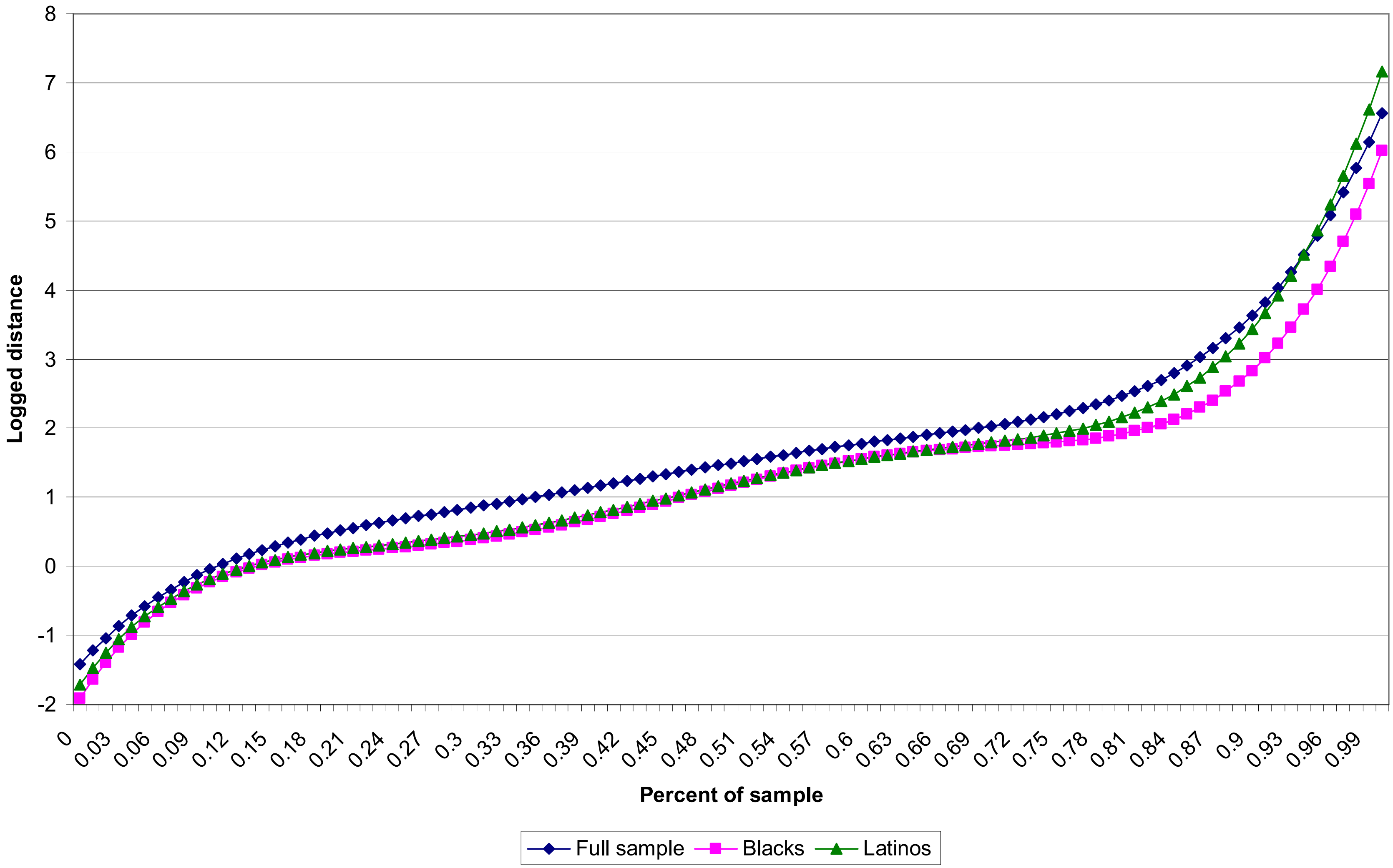

Figure 3. Logged distance of most recent move for households based on presence of children

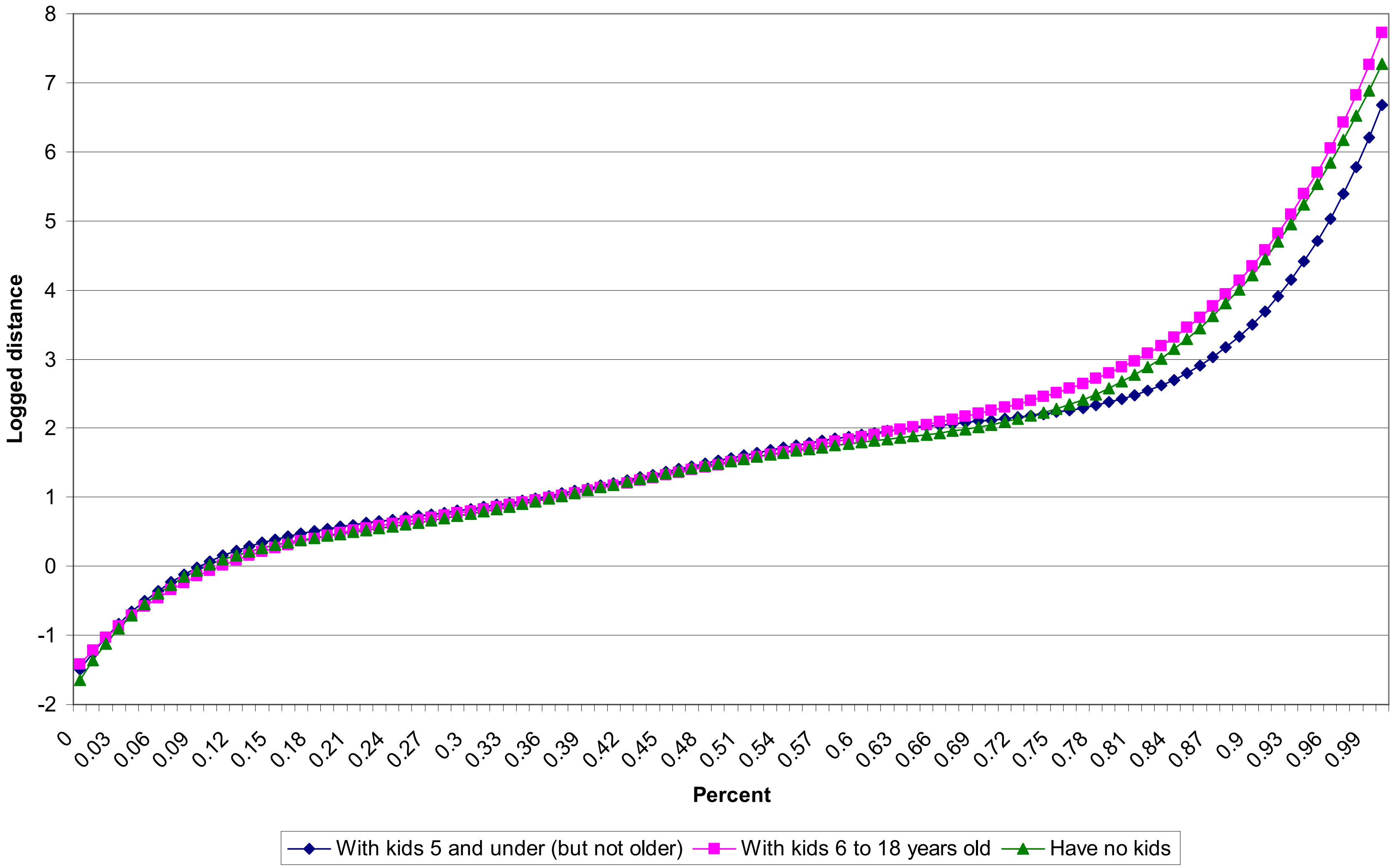

Figure 4. Logged distance of most recent move for owners and renters