README FOR: AN EMBARRASSMENT OF RICHES: CONFRONTING OMITTED VARIABLE BIAS AND MULTISCALE CAPITALIZATION IN HEDONIC PRICE MODELS

Joshua K. Abbott and H. Allen Klaiber

The Review of Economics and Statistics, November 2011, 93(4): 1331–1342

README

(A) Stata code needed to replicate the study is included as "Spatial_HT_ReStat.do." Instructions for running the code are contained as comments in the attached Stata .do file.

(B) The following variables and data sources are used in the current paper and attached .do file:

a. Housing characteristics: price, square footage, lot acres, stories, bathrooms, year built, garage presence, pool presence. These are confidential data and may be obtained as described in section [E].
b. Land use characteristics: subdivision open space, local parks, city parks, regional parks, schools, rail, canals. These are confidential data and may be obtained in section [E].
c. Census demographics: race, age. These are public data available from the U.S. Census Bureau at http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.
d. Spatial attributes: distance to central Phoenix. This data was created using ArcGIS.

(C) Stata 11/MP and ArcGIS 9.3 were used in the analysis.

(D) For data descriptions see section (B) and (E) along with the attached Stata .do file.

(E) Data associated with housing characteristics and land use characteristics is not in the public domain. However, the process to acquire these data is straightforward and is described below. There are two primary sources of data required to replicate this analysis: assessor transactions data and GIS data on open space.

a. The Maricopa county assessor provides real estate transactions data, subdivision identifiers, as well as property use codes for a small fee. This data may be obtained following the procedures outlined at the following web address:

http://mcassessor.maricopa.gov/assessor/. Using the assessor provided data alone; it is possible to replicate all data elements required for estimation with the exception of non-subdivision open space.

b. To supplement the assessor data, we also purchased transactions data from a private data vendor, Dataquick. This data primarily served as a "check" on the assessor data and would not be needed, in principle, to replicate the analysis undertaken in this paper. However, this data did allow us to fill in a small number of missing housing characteristics for a small subset of houses thereby expanding the data sample available for estimation.

c. To obtain the data used to characterize non-subdivision open space, including city and regional parks, we relied on confidential data obtained from Arizona Parks and Recreation. These data are available on request as the "Statewide Recreation Inventory." We were able to obtain these data at no cost after signing a confidentiality agreement. These data were provided in GIS shapefile format.

(F) Code to estimate the model is attached, but due to confidentiality agreements, the data is not included in the same directory. The process to acquire the data is outlined in section (E).

(G) This paper was exempted from providing the data due to confidentiality agreements required to obtain the data. The code has been published.