Program files for "Metropolitan Land Values" by David Albouy, Gabriel Ehrlich, and Minchul Shin

Albouy, D., G. Ehrlich, and M. Shin (2017): "Metropolitan Land Values," *Review of Economics and Statistics*, Forthcoming.

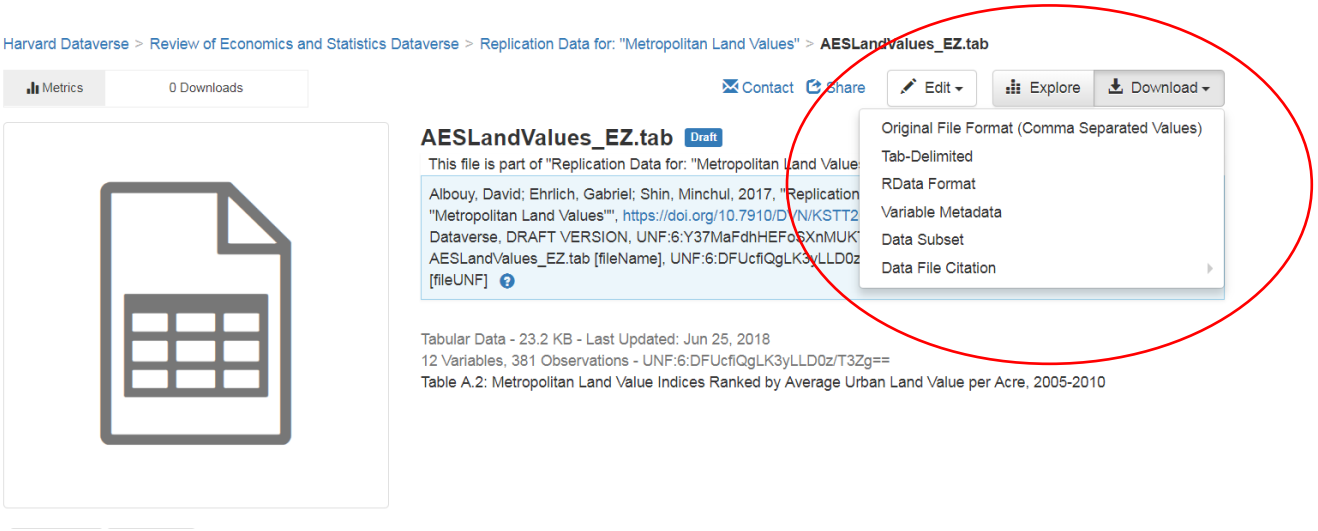
Last revised: 2018-6-25
First version: 2017-7-08

<Land value estimates>

If you just want to use our land value estimates (Table A2), please download the following file:

AESLandValues_EZ.tab

This file can be downloaded in various formats (e.g., csv, RData, etc.)



<Program>

All required files are in the zipped file.

We describe our program in the following three steps:

1. Step 1: Loading the multiple data sets
2. Step 2: Estimation with and without shrinkage; covariate estimation.
3. Step 3: Land value computation, figures, and tables

We describe how to access the underlying land transactions data at the end of the document.

Step 1: Loading the multiple data sets

The following m-file merges and rearranges multiple data sets for estimation

1. File: **main_new01_data.m** for our main estimator
    a. This file loads multiple data sets (explained below) and combines them to generate 'data_all.mat', which is used for estimation (step 2 below).

This process loads multiple data sets

1. all_20170217.csv: This dataset is based on Costar, which has restricted access. Thus, it is not included in the actual files. Another document explains how to get them. This csv file contains the following (by column)
    a. origorder: unique id to each land transaction observation (assigned by us not Costar).
    b. Pmsafips: cmsa/pmsa combination code to assign the unique MSA identifier.
    c. Year: Transaction year
    d. Y: log price per acre
    e. X1 – X13: Purpose of use (dummy variables)
        i. X1: no purpose
        ii. X2: commercial
        iii. X3: industrial
        iv. X4: retail
        v. X5: single family
        vi. X6: multi family
        vii. X7: office
        viii. X8: Apartment
        ix. X9: Hold for development
        x. X10: Hold for investment
        xi. X11: mixed use
        xii. X12: medical
        xiii. X13: parking
    f. X14: log lot size
    g. Lati and loni: latitude and longitude of the land transaction
    h. Year_dummy1 – Year_dummy6 (dummy variables)
    i. Latj and lonj: latitude and longitude of the corresponding MSA. (not used in the actual estimation, we use ch_lat and ch_lon instead)
    j. Pop2k_c, hu2k_c, landb_c, waterb_c, areab_c: MSA level population, housing units, land area, water area, total area from the census 2k. (not used in the actual estimation, we use Area_us instead).
    k. Ch_lat and ch_lon: latitude and longitude of MSA center.
    l. Area_ua and afact_ua: urbanized area of the corresponding MSA from the census 2k.
    m. Min_dist, min_dist2: distance from the lot to the corresponding city center.
    n. Dist2coast: distance from the lot to the closest coastal line.
    o. Final note: Y, X1-X14, lati, loni, min_dist, min_dist2, and dist2coast are based on CoStar dataset. Thus, we replace them with missing values.
2. name_20170217.mat: matlab file that contains MSA level information (names and the total number of observations by MSAs).
3. censustracts2_ch.csv: tract level information. Included variables (by column) are

   a. combofips: CMSA/PMSA identifier.
   b. Ch_lat and Ch_lon: latitude and longitude of corresponding MSA center.
   c. Lati and Loni: latitude and longitude of tract centroid.
   d. Min_dist and min_dist2: distance from tract centroid to the city center.
   e. Pop2k, pop09: population from 2000 and 2009 census.
   f. Hus2k: housing units from 2000 census.
   g. Landsqmi: total land area of the tract.
   h. Landsqmi_ua: total land area that is part of urbanized area
4. censustracts2_d2c.csv: additional tract level information
   a. First column: arbitrary id number for the tract centroid
   b. Second column: distance from the tract center to the closest coastal line
   c. Third column: latitude of the tract center
   d. Fourth column: longitude of the tract center

Step 2: Estimation with and without shrinkage; covariate prediction

1. Land value estimation with shrinkage
   a. File: **main_new04_full_logp1_A_s3.m**
   b. Takes 'data_all.mat' from the step 1.
   c. Estimates land value model (Eq 1 in the main text).
   d. Stores estimation output in mat format: 'results_full_logp1_A_s3_REML.mat'
2. Unshrunken estimation for Figure 2
   a. File: **main_new04_full_unshrunken_REML_s3.m**
   b. Same as above but with/without shrinkage.
   c. This file stores estimation output in mat format: 'results_shrunken_s3.mat'
3. Covariate estimation for determining predicted value
   a. File 1: **main_new201_Xestpred.m** (for the shrinkage estimator)
   b. File 2: **main_new201_Xestpred_unshrunken.m** (for the unshrunken estimator)
   c. This file estimates a model for covariates (Eq 14 in the appendix)
   d. Takes 'data_all.mat' from the step 1.
   e. Outputs: 'results_estX_REML.mat' and 'results_estX_shrunken_s3.mat'.

Step 3: Land value computation, figures, and tables

All results in Step 3 are stored in "Program path\ figure_draft_rev"

1. Tables for the "Metropolitan Land Value Indices"
   a. File: **main_new211_fig_for_paper_part1.m**
   b. This file generates numbers for our main Tables (Table 2 and Table A.2)
   c. This file uses outputs from Step 1 and Step 2.
   d. This file generates excel file "tab_lv_log2m6_estX_M3.xls"
2. Figure 1
   a. File: **main_new201_fig_for_paper_Houston.m**
   b. This file uses outputs from Step 1 and Step 2.
3. Figure 2

    a. File: **main_new212_fig_for_paper_part2.m**
    b. This file uses outputs from Step 1 and Step 2.
4. Figure 3 and Table A.3
    a. File: **DavisPalumboComps.do** (Stata file) – this file makes the comparison figures to the Davis and Palumbo index and assembles the data for Table A.3 with the same comparisons.
    b. File: **tab_lv_log2m6_estX_M3.xls** – this file contains the MSA –level land value estimates from the Matlab programs
    c. File: **Davis Palumbo Land Prices Panel.xlsx** – this file contains the Davis and Palumbo Land Price Data (sheet "landdata-msas-2016q1" contains the raw data downloaded from http://datatoolkits.lincolninst.edu/subcenters/land-values/metro-area-land-prices.asp; the sheets "for stata" and "for stata 2" have data to be imported into Stata, and the sheet "Table A3" contains the output of the Stata do file formatted nicely.)
    d. File: ahs_smsalot_2011.dta – this file contains data from the American Housing Survey regarding typical lot sizes and is used to compare our per-acre land value estimates to Davis and Palumbo's land value estimates estimates per typical lot.
5. Land values for missing MSAs
    a. File: main_new205_missing_estX.m
    b. This file uses outputs from Step 1 and Step 2.
    c. This file generates excel file "tab_lv_missing_log2m6_estX_M3.xls"
6. Table A.1 (parameter estimates)
    a. File: **main_new104_final_coeff.m**
    b. This file generates excel file "tab_parameters_log2m6.xls"

CoStar COMPs Data Access

The land transactions dataset used in this paper is proprietary data from the CoStar COMPs Database. It is ordinarily available to academic researchers free of charge through the CoStar University program: http://origin-www-vi.costar.com/SpecialPrograms/CoStarUniversity/Access.aspx. The data is available for download as PDF files. We used the program Nitro Pro (https://www.gonitro.com/) to convert the data into machine-readable format.