Accounting for Spatial Variation of Land Prices in Hedonic Imputation House Price Indexes

Jan de Haan* and Yunlong Gong**

* Statistics Netherlands / Delft University of Technology

** Delft University of Technology

EMG Workshop 2014

Outline

- Background
- A simplification of the ‘builder’s model’
  - Basic ideas
  - Adding characteristics and linearizing
- Location and spatial nonstationarity of land prices
  - Two models
  - The MWGR method
- Hedonic imputation price indexes
- Empirical results
- Conclusions
  - Fancy econometric methods are unnecessary ….

Background

Uniqueness of properties mainly due to location

A simplification of the ‘builder’s model’

Builder’s model (Diewert, de Haan and Hendriks, 2015): value of property p_i^t is sum of value of land and value of structure:

$$p_i^t \qquad {}^t z_{iL}^t \qquad {}^t z_{iS}^t \qquad u_i^t$$

z_{iL}^t : plot size in square meters

z_{iS}^t : living space in square meters

^{t} : price of land per square meter

^{t} : price of structure (living space) per square meter

A simplification of the ‘builder’s model’

Writing in linear form, using (multiplicative) dummies D_{ia}^{t} for age category, and reparameterizing:

$$p_i^t = {}^{t}z_{iL}^{t} + \sum_{a=1}^{A} {}_{a}^{t}D_{ia}^{t} z_{iS}^{t} + u_i^t$$

No restrictions on parameters

Functional form is neither continuous nor smooth

Adding structure characteristics (number or rooms, type of house)

Only categorical variables; dummies D_{ir}^{t}

Ignoring interaction terms and reparameterizing

A simplification of the ‘builder’s model’

Fully linear model:

$$p_i^t = {}_t z_{iL}^t + \sum_{a=1}^{A} {}_t^a D_{ia}^t + \sum_{r=1}^{R} {}_t^r D_{ir}^t \; z_{iS}^t \; u_i^t$$

Normalizing (dividing by structure size):

$$\sum_{r} {}_t^a \; {}_t^r \; {}_t^{ir} \; {}_t^i$$

p_i^{t*} p_i^t / z_{iS}^t : “normalized” property price

r_i^t z_{iL}^t / z_{iS}^t : ratio of plot size to structure size

Straightforward estimating equation (including intercept)

Location and spatial nonstationarity of land prices

Location is capitalized into price of land not price of structures

1) Price of land (only) varies across postcode areas *k*:

$$p_{ik}^{t} = \sum_{k=1}^{K} p_{k}^{t} D_{ik}$$

p_k^t : price per square meter of land for area *k*

D_{ik} : multiplicative dummy for *k*

2) Price of land differs at property level:

$$p_i^{t*} = r_i^t \sum_{a=1}^{A} p_a^t D_{ia}^t \sum_{r=1}^{R} p_r^t D_{ir}^t + \varepsilon_i^t$$

Hedonic imputation price indexes

Hedonic double imputation house price indexes: Laspeyres, Paasche and Fisher

$$P^{0t}_{\text{Laspeyres}} \quad \frac{\sum_{i \in S^0} p_i^{t(0)}}{\sum_{i \in S^0} p_i^0} \quad \text{(defined on base period sample)}$$

Predicted prices:

$$p_i^0 \qquad p_i^{0*} z_{iS}^0 \qquad {}_i^0 z_{iL}^0 \left[{}^0 \qquad \sum_{a=1}^{A} {}_a^0 D_{ia}^0 \qquad \sum_{r=1}^{R} {}_r^0 D_{ir}^0 \right] z_{iS}^0$$

Estimated quality-adjusted prices:

$$p_i^{t(0)} \qquad {}_i^t z_{iL}^0 \left[{}^t \qquad \sum_{a=1}^{A} {}_a^t D_{ia}^0 \qquad \sum_{r=1}^{R} {}_r^t D_{ir}^0 \right] z_{iS}^0$$

Hedonic imputation price indexes

$$P^{0t}_{Laspeyres} \quad s^0_L P^{0t}_{L,Laspeyres} \quad s^0_S P^{0t}_{S,Laspeyres}$$

Estimated value shares for land and structures, s^0_L and s^0_S , sum to 1 due to double imputation

E.g. Laspeyres price index for land:

$$P^{0t}_{L,Laspeyres} \quad \frac{{}_{i\ S^0} \ {}^{t}_{i} z^0_{iL}}{{}_{i\ S^0} \ {}^{0}_{i} z^0_{iL}}$$

Big influence of properties with relatively large value shares (large plot sizes and high land prices)

Empirical results

Three models estimated, separately for each year:

1) No variation in land prices allowed ("OLS")

2) Variation across postcodes ("OLSD")

3) Variation across individual properties ("MWGR")

[60 neighboring properties used in MWGR estimations]

According to (corrected) AICc as well as RMSE:

OLSD performs better than OLS

MWGR performs better than OLSD

for each year

Empirical results

Parameter estimates for structure characteristics, 2007

	OLS	OLSD	MGWR
Intercept	1561.00** (46.93)	1472.04** (55.59)	1633.70** (75.35)
Building period:1960-1970	-367.23** (26.85)	-310.09** (36.97)	-411.55** (45.21)
Building period:1971-1980	-308.01** (24.19)	-255.16** (35.68)	-378.17** (44.86)
Building period:1981-1990	-230.45** (24.21)	-178.98** (34.25)	-259.74** (45.46)
Building period:1991-2000	-54.42* (22.41)	-58.87* (27.87)	-124.04** (38.28)
Terrace	-326.68** (35.80)	-286.66** (36.78)	-309.05** (42.04)
Corner	-303.89** (32.67)	-280.98** (32.67)	-278.44** (35.04)
Semidetached	-156.63** (49.37)	-165.54** (49.85)	-195.84** (52.39)
Duplex	-171.43** (31.49)	-149.10** (31.63)	-170.19** (33.94)

Empirical results

Intercept measures price of living space per square meter for detached houses built after 2000

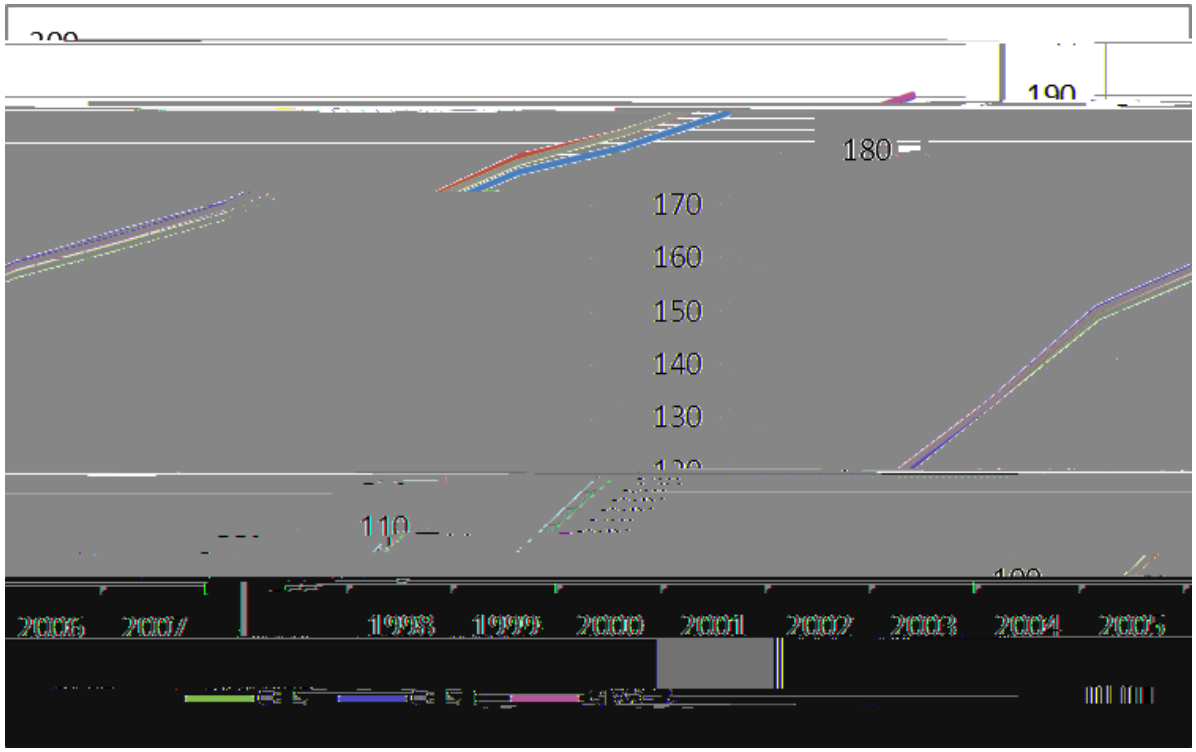
Large difference between intercepts for MWGR and OLSD

Structures become less expensive as they get older

Detached houses are more expensive than other types of houses

All coefficients differ significantly from zero

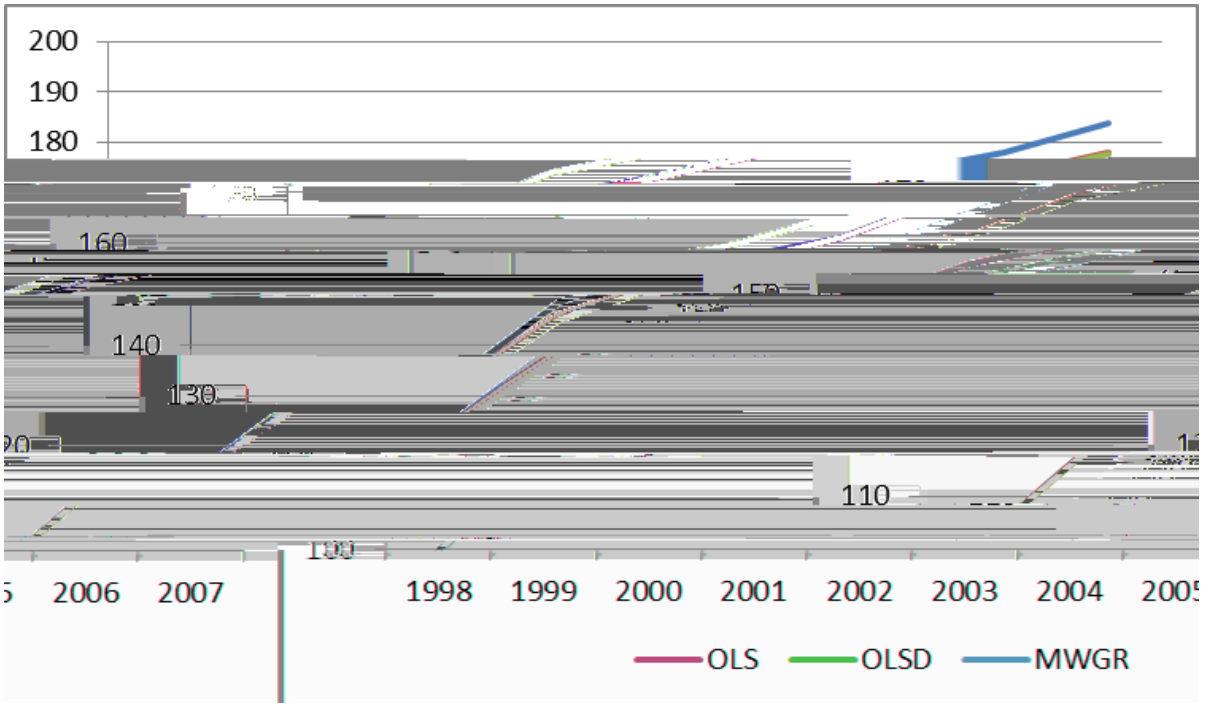
Chained hedonic imputation Laspeyres house price index



Hardly any difference between OLSD and MGWR

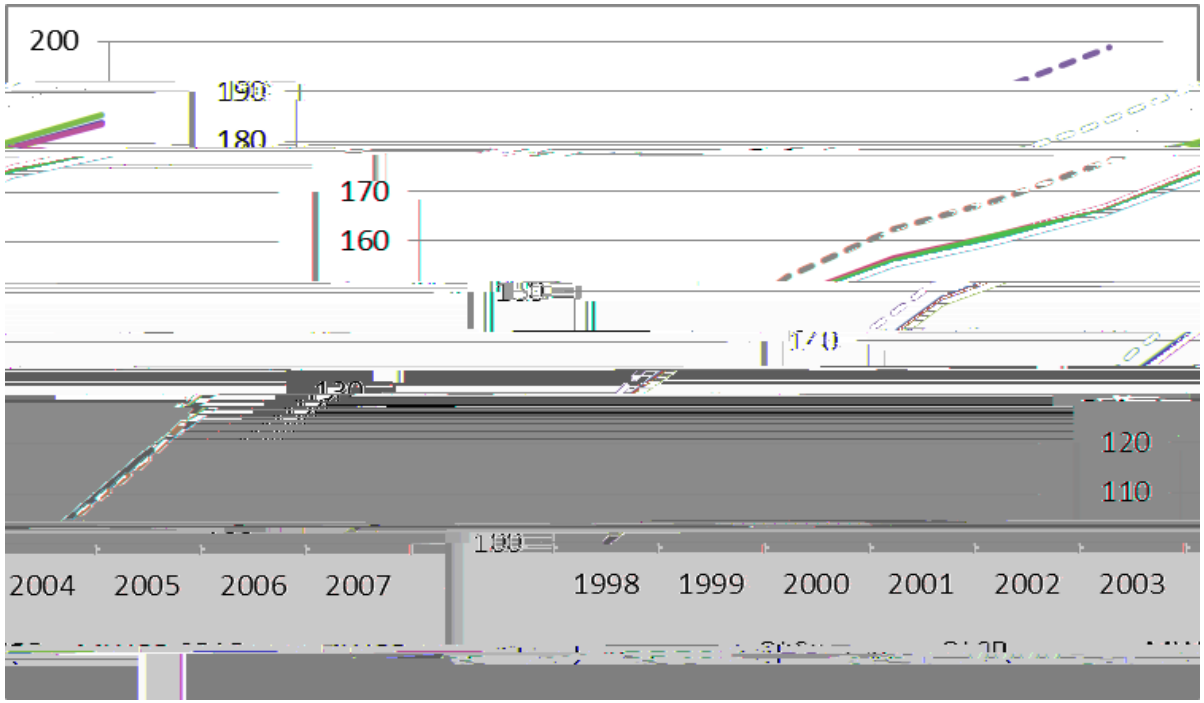
OLS has downward bias

Chained hedonic imputation Paasche house price index



OLS index upward biased

Chained hedonic imputation Fisher house price index



Fisher index insensitive to choice of hedonic model

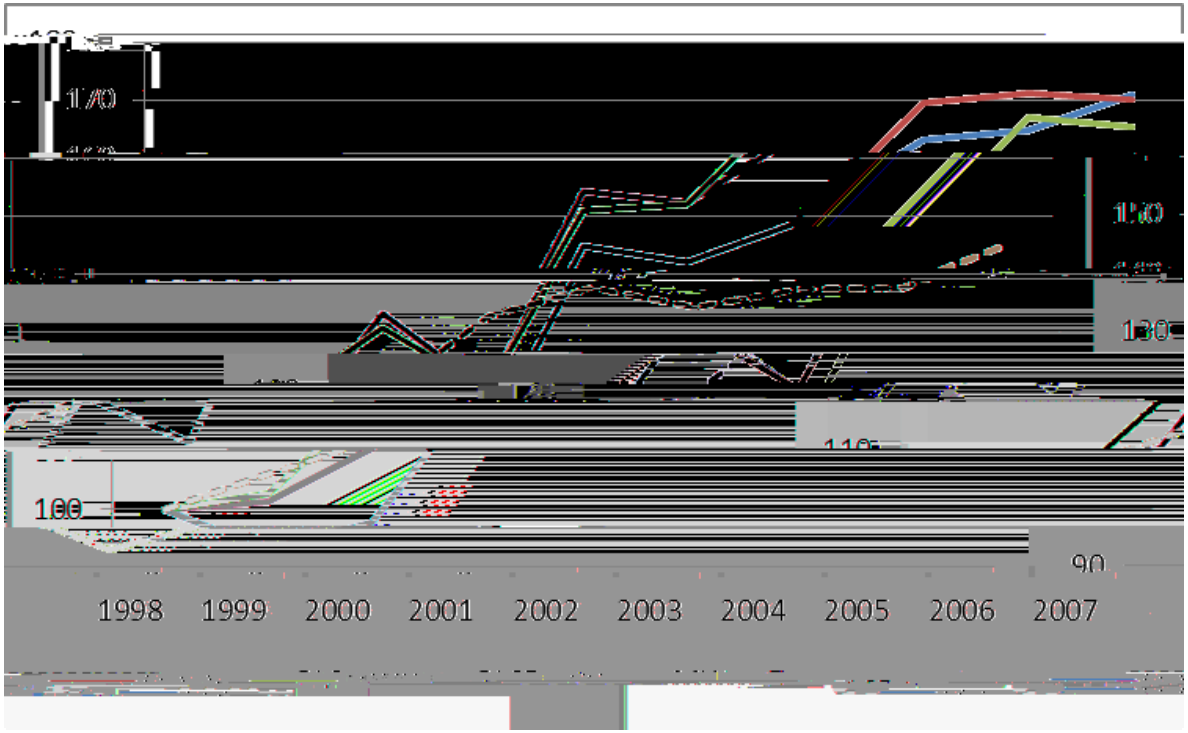
Official (nationwide) SPAR index rises much faster

Chained hedonic imputation Fisher price indexes for land

OLS and OLSD similar but MWGR very different

MGWR extremely volatile

Chained hedonic imputation Fisher price indexes for structures and official construction cost index



Differences much smaller than for land

Empirical results

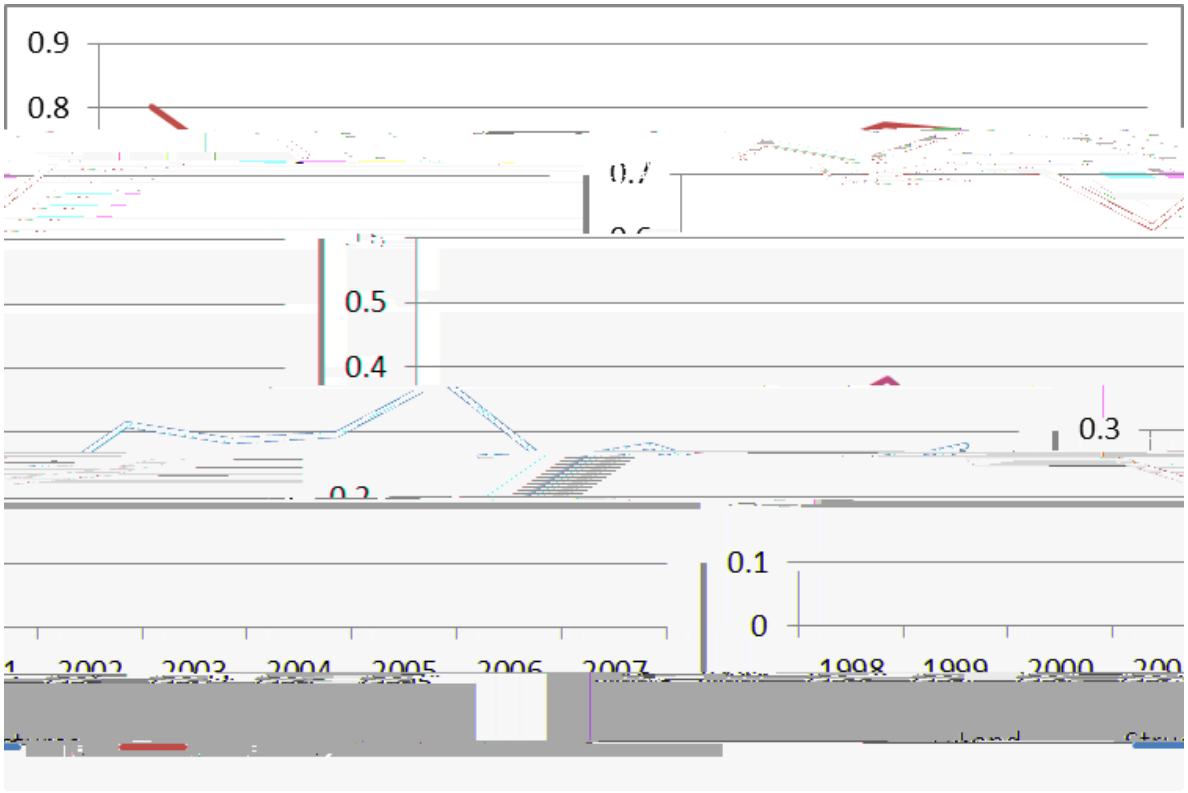
Are the trends on indexes for land and structures plausible?

No benchmark available for land

For structures: official (nationwide) construction cost index

- flattens during second half of sample period; price indexes for structures keep rising
- bias in construction cost index?
- house prices were still rapidly rising while construction cost index increased by only 4.9% during 2003-2007 (CPI: 5.8%)

Estimated value shares of land and structures, OLSD



Also very volatile

Structures share approximately 75%

Empirical results

Potential causes of volatility of the land and structure indexes

1) Small number of observations

2) Multicollinearity

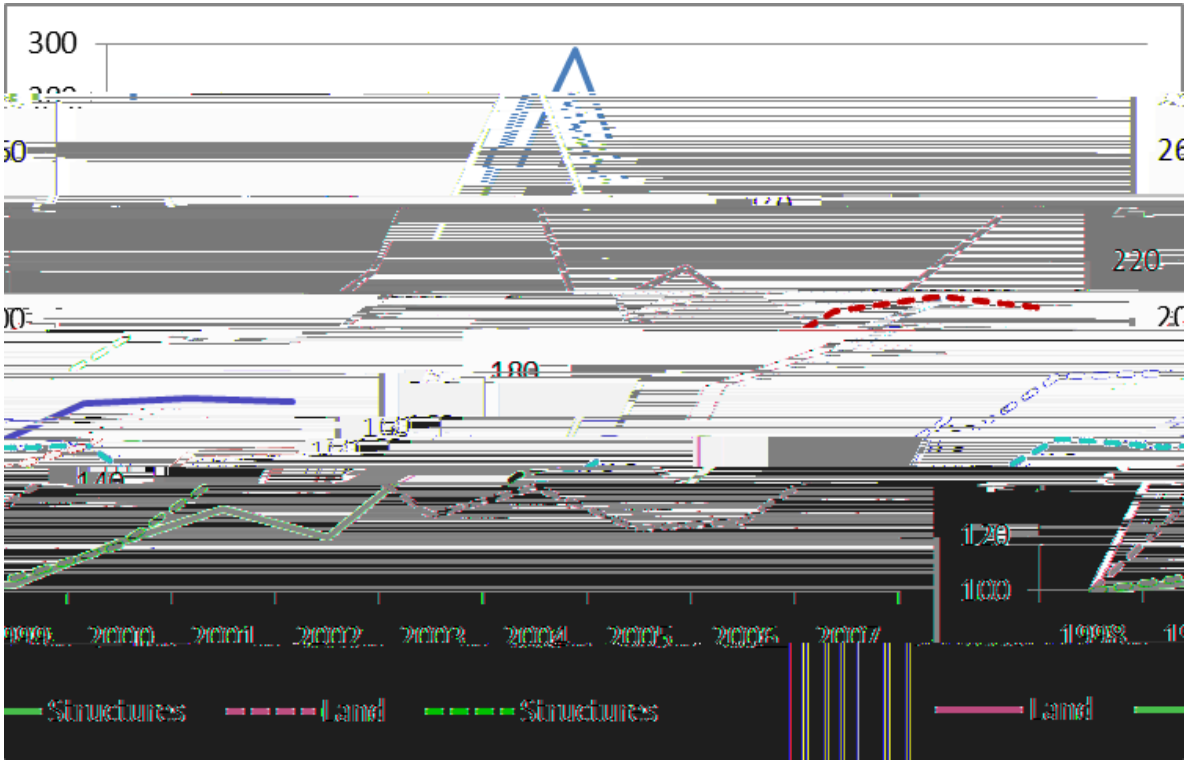
Land and structure price changes do not consistently show opposite signs; VIF for ratio of plot size to structures size is low

3) Heteroskedasticity

Yes (Breusch-Pagan test for OLS and OLSD)

4) Non-linear relation between normalized property price and ‘ratio’, and outliers

Chained hedonic imputation Fisher price indexes for land and structures, OLSD



Deleting all observations with ‘ratio’ larger than 5:

reduces volatility but changes trends dramatically

Conclusions

- The linearization and ‘normalization’ of the builder’s model is useful for estimating (overall) house price indexes
- Double imputation Fisher house price index is insensitive to choice of hedonic model, so …
- … no need to use spatial econometrics for estimating overall price index; postcode dummies will suffice

  (see also Hill and Scholz, 2014)
- Land and structure price indexes are very volatile due to outliers and nonlinear relation between normalized property price and land size to structure size ratio

  (and heteroskedasticity; multicollinearity not a big problem)