Constructing HPIs

RE420: URBAN AND REGIONAL ECONOMICS

Review of HPI

House Price Index

- **A house price index (HPI)** measures the price changes of residential housing as a percentage change from some specific start date (which has an HPI of 100).
- Popular HPI Construction Methodologies
  1. Representative Property HPI Method
  2. Average/Median HPI Method
  3. The Hedonic HPI Model
  4. The Repeat-Sales HPI Model

The Representative Property Method

- The Representative Property Price Index defines a representative property and collect the price of the property in each period.

- Advantages:
  - Easy and straightforward to understand
  - Requires a small data size.
- Limitations:
  - The representative property may not be comparable across different markets
  - It does not utilize information from other properties
  - The chosen representative property may no longer accurately reflect the housing market.

Average/Median Price Index

- Average or Median Price Index collect the sale price data for all properties transacted during the period.

- Advantage:
  - Easy to calculate and understand
  - Considers all property transactions rather than focusing one
- Limitations:
  - Requires a much larger data collection than Representative Property Method.
  - Average/Median Price Index fails to control for changes in the *quality* or *mix* of properties

The Hedonic House Price Index Model

- The Hedonic House Price Index measures property values by accounting for various characteristics that affect price, such as size, location, and amenities.
- A Standard Hedonic Model (review Lecture Note 7 for more details!)

$$\ln P_{it} = \alpha + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \sum_{t=2}^{T} \gamma_t D_t + \varepsilon_{it}$$

  - P_i is transaction price of property i, and is expressed in logarithmic form because the housing transactions prices are log-normally distributed;
  - X_1 represents any continuously measured property, locational and neighborhood hedonic characteristics (e.g., lot size);
  - X_2 represents any discretely measured property, locational and neighborhood hedonic characteristics (e.g., number of bedrooms, presence of garage);
  - D_{it} is indicator variables which take value of 1 if property i transacted during period t and 0 otherwise (year dummy)

The Hedonic House Price Index Model

- The Hedonic House Price Index measures property values by accounting for various characteristics that affect price, such as size, location, and amenities.
- A Standard Hedonic Model (review Lecture Note 7 for more details!)

$$\ln P_{it} = \alpha + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \sum_{t=2}^{T} \gamma_t D_t + \varepsilon_{it}$$

  - P_i is transaction price of property i, and is expressed in logarithmic form because the housing transactions prices are log-normally distributed;
  - X_1 represents any continuously measured property, locational and neighborhood hedonic characteristics (e.g., lot size);
  - X_2 represents any discretely measured property, locational and neighborhood hedonic characteristics (e.g., number of bedrooms, presence of garage);
  - D_{it} is indicator variables which take value of 1 if property i transacted during period t and 0 otherwise (year dummy)

The Hedonic House Price Index Model

- Specifically, the Hedonic House Price Index utilize the values of γ_t , the coefficient estimates of year dummies
- Changes in the γ_t values reflect annual changes in house price levels after accounting for the heterogeneous characteristics of individual housing units

The Hedonic House Price Index Model

- Advantages:
  - Measures the house price changes while keeping the constant quality
    - Factors out the quality differences of housing units, separating **the pure price change** from the change in quality

- Limitations:
  - Data Intensive: Requires extensive data collection and sophisticated statistical techniques.
  - Omitted Variable: Impossible to know all the variables that affect housing prices, and omitting the key variables will affect the estimated values for γ_t.

The Repeat-Sales HPI Model

- The Repeat-Sales House Price Index model extends the hedonic house price index model by utilizing the two paired transaction observations on the same property

- **The Hedonic Model**

$$\ln P_{it} = \alpha + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \gamma_t \qquad (1)$$

$$\ln P_{is} = \alpha + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \gamma_s \ \ (\mathrm{t} \neq s) \qquad (1')$$

- **The Repeat-Sales Model**: subtract eq. (1') from eq. (1),

$$\ln \frac{P_{it}}{P_{is}} = (\alpha - \alpha) + \beta_1 (\ln X_{1i} - \ln X_{1i}) + \beta_2 (X_{2i} - X_{2i}) + (\gamma_t - \gamma_s) \qquad (2)$$

  - Since the hedonic characteristics remain unchanged, the first three terms in RHS equals to zero.

The Repeat-Sales HPI Model

- From eq. (2), the repeat-sales model can be simplified to:

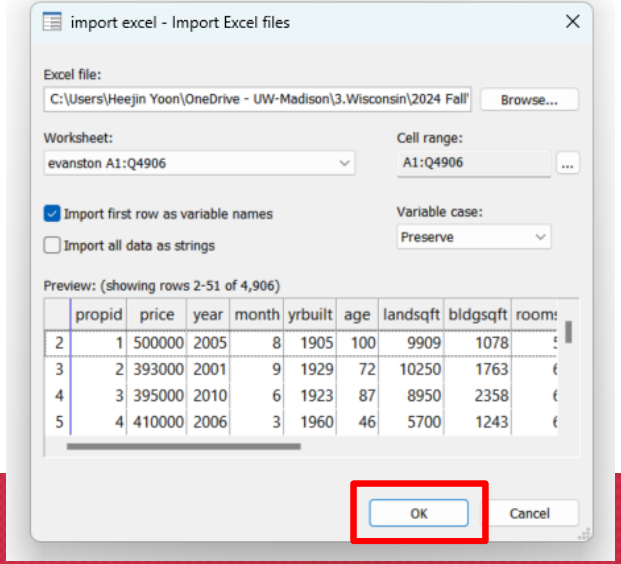
$$\ln \frac{P_{it}}{P_{is}} = \gamma_t - \gamma_s$$

- Advantages:
  - Based on the same theoretical consideration as the Hedonic House Price Index
  - Data burden is lower than the Hedonic HPI, because all X_i's are differenced out
- Limitations:
  - Selection Bias: Rely on properties that sell multiple times– these properties may not represent the overall market
  - Violation of constant quality assumption (Meese and Wallace, 1991): the same property may experience changes in housing quality

Computing HPIs Using Stata & Excel

Import the Housing Transaction Data

- File > Import > Data to Excel spreadsheet (*.xls;*.xlsx)
- Browse and select the file *housing transactions v2.xlsx.*
- Check "Import first row as variable names" and click "OK"



1. Average/Median HPI

Average/Median Price Index

Let's start from the Average/Median HPI Model.

1. Tabulate the average transaction prices by each year:

   – Command: `tabulate year, summarize(price)`

```
. tabulate year, summarize(price)
```

year	Summary of price		
	Mean	Std. dev.	Freq.
2000	352456.81	205702.31	340
2001	362530.84	253994.12	355
2002	424676.2	294506.9	409
2003	436728.58	256155.97	432
2004	479291.47	356295.22	395
2005	519468.94	334838.52	365
2006	546087.28	321749.17	316
2007	553604.73	288896.16	244

Average/Median Price Index

2. Tabulate the median transaction prices by each year (2 step):

i. Generate the variable *median_price:*

- Command: `bysort year: egen median_price = median(price)`

ii. Report the variable *median_price* by each year:

- Command: `tabulate year, sum(median_price)`

```
tabulate year, summarize(median_price)
```

year	Summary of median_price Mean	Std. dev.	Freq.
2000	310000	0	340
2001	305000	0	355
2002	350000	0	409
2003	375000	0	432
2004	390000	0	395
2005	424000	0	365
2006	465000	0	316
2007	497500	0	244

Why the standard deviation is zero for all years?

Average/Median Price Index

3. Copy & paste the "Mean" values from Steps 1. and 2. to Excel.

	A	B	C
1	Year	Average House Price	Median House Price
2	2000	352,456.81	310,000.00
3	2001	362,530.84	305,000.00
4	2002	424,676.20	350,000.00
5	2003	436,728.58	375,000.00
6	2004	479,291.47	390,000.00
7	2005	519,468.94	424,000.00
8	2006	546,087.28	465,000.00
9	2007	553,604.73	497,500.00

Average/Median Price Index

4. Convert the values to HPI by normalizing the 2000 value to 100.

Command: =B2/$B$2*100, =B3/$B$2*100, =B4/$B$2*100, …

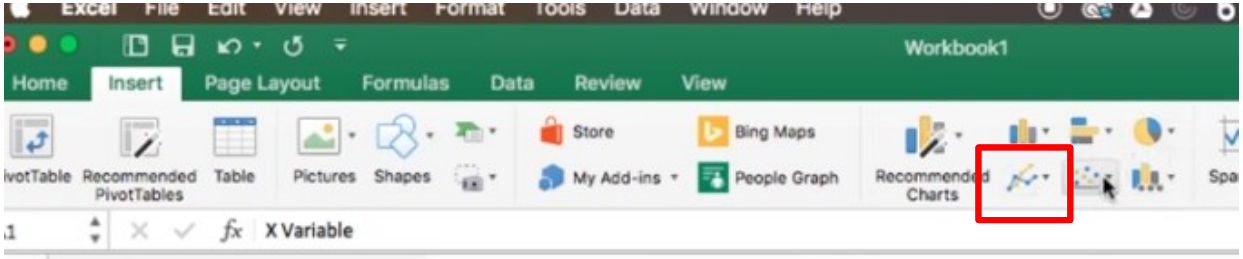
A	B	C	D	E
Year	Average House Price	Median House Price	Average HPI	Median HPI
2000	352,456.81	310,000.00	=B2/$B$2*100	100.00
2001	362,530.84	305,000.00	102.86	98.39
2002	424,676.20	350,000.00	120.49	112.90
2003	436,728.58	375,000.00	123.91	120.97

A	B	C	D	E
Year	Average House Price	Median House Price	Average HPI	Median HPI
2000	352,456.81	310,000.00	100.00	100.00
2001	362,530.84	305,000.00	=B3/$B$2*100	98.39
2002	424,676.20	350,000.00	120.49	112.90
2003	436,728.58	375,000.00	123.91	120.97

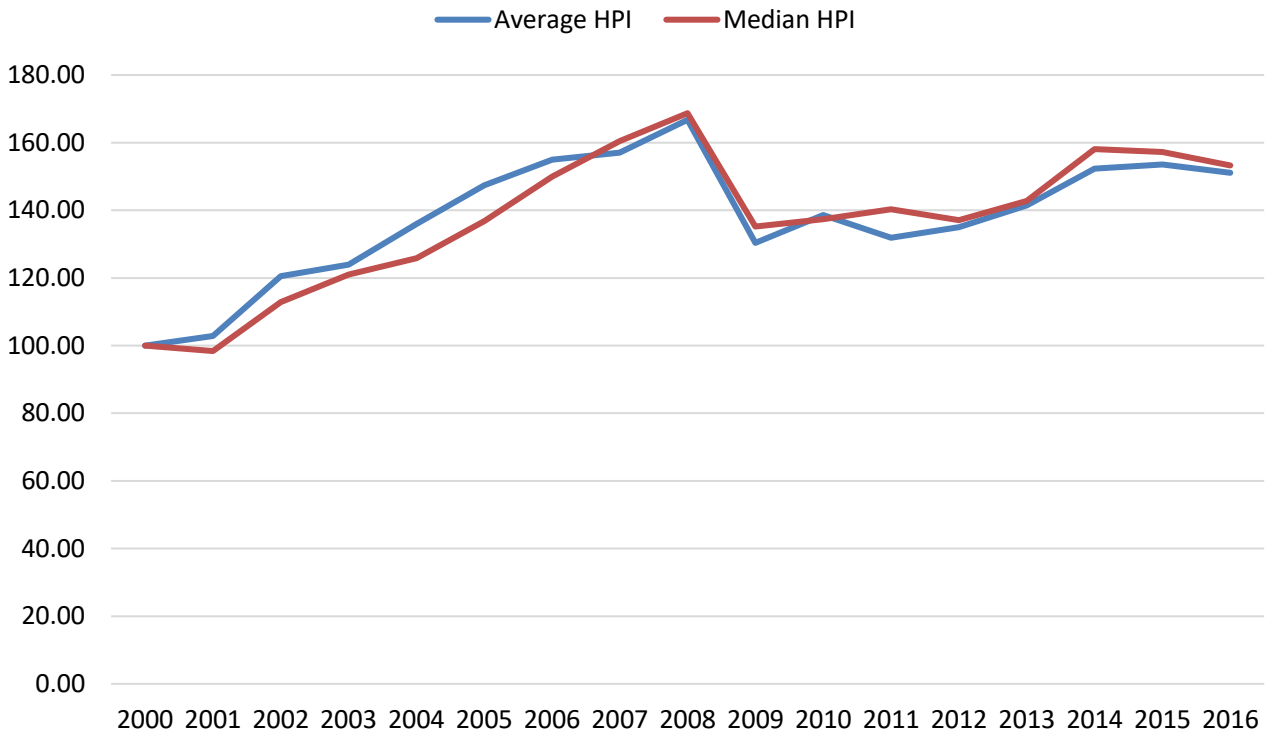
Average/Median Price Index

5. Plot the Average HPI and Median HPI.

Select the data > Insert > Line Graph



Average/Median Price Index



Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms

1. Tabulate the average transaction prices by each year, only for 1-4 bedrooms:
   - Command: `tabulate year if bedroom>=1 & bedroom<=4, summarize(price)`

```
. tabulate year if bedroom>=1 & bedroom<=4, summarize(price)
```

year	Summary of price Mean	Std. dev.	Freq.
2000	312495.25	160127.7	302
2001	333587.22	226358.82	325
2002	374629.98	198548.17	369
2003	382476.09	181866.73	387
2004	436298.4	324694.49	356
2005	469943.88	271896.33	325
2006	484146.5	253634.43	277
2007	525500.26	273114.57	322

Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms

2. Tabulate the average transaction prices by each year, only for 5-8 bedrooms:

   – Command: `tabulate year if bedroom>=5 & bedroom<=8, summarize(price)`

```
. tabulate year if bedroom>=5 & bedroom<=8, summarize(price)
```

year	Summary of price Mean	Std. dev.	Freq.
2000	670046.05	249934.49	38
2001	676086.67	322416.05	30
2002	886352.63	541378.14	40
2003	903300	322568.83	45
2004	871741.03	394893.2	39
2005	921860	495727.93	40
2006	986025.64	405338.02	39
2007	837113.64	297028.7	22

Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms

3. Copy & paste the "Mean" values from Steps 1. and 2. to Excel.

A	B	C
Year	Average House Price (1-4 Bedrooms)	Average House Price (5-8 Bedrooms)
2000	312,495.25	670,046.05
2001	333,587.22	676,086.67
2002	374,629.98	886,352.63
2003	382,476.09	903,300.00
2004	436,298.40	871,741.03
2005	469,943.88	921,860.00
2006	484,146.50	986,025.64
2007	525,509.26	837,113.64
2008	528,296.05	870,985.79

Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms

4. Convert the values to HPI by normalizing the 2000 value to 100.

Command: =B2/$B$2*100, =B3/$B$2*100, =B4/$B$2*100, …

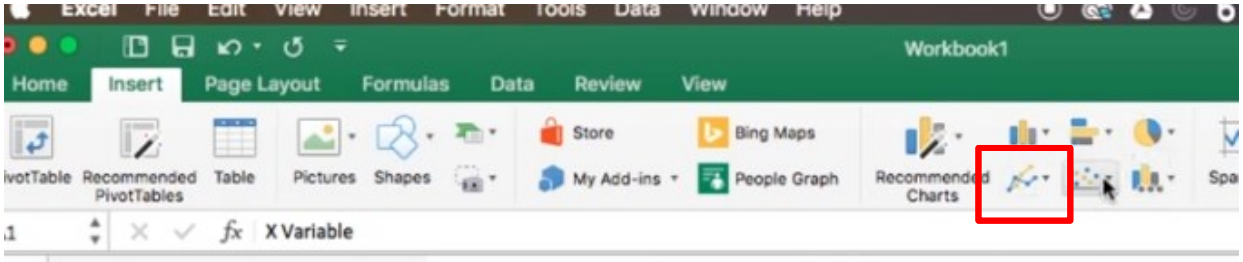
Year	Average House Price (1-4 Bedrooms)	Average House Price (5-8 Bedrooms)	Average HPI (1-4 Bedrooms)	Average HPI (5-8 Bedrooms)
2000	312,495.25	670,046.05	=B2/$B$2*100	100.00
2001	333,587.22	676,086.67	106.75	100.90
2002	374,629.98	886,352.63	119.88	132.28

Year	Average House Price (1-4 Bedrooms)	Average House Price (5-8 Bedrooms)	Average HPI (1-4 Bedrooms)	Average HPI (5-8 Bedrooms)
2000	312,495.25	670,046.05	100.00	100.00
2001	333,587.22	676,086.67	=B3/$B$2*100	100.90
2002	374,629.98	886,352.63	119.88	132.28

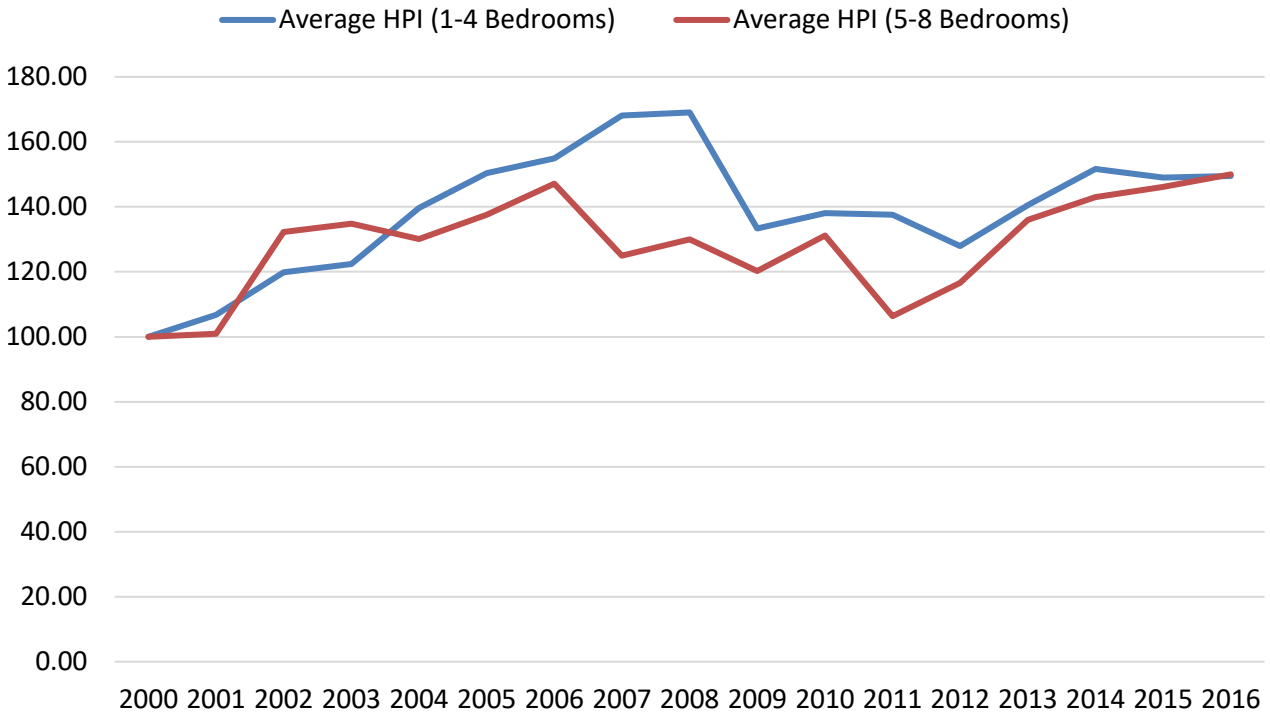
Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms

5. Plot the Average HPI for 1-4 Bedrooms and the Average HPI for 5-8 Bedrooms

Select the data > Insert > Line Graph



Extension: Average HPIs for 1-4 Bedrooms and 5-8 Bedrooms



2. Hedonic HPI

Hedonic HPI Model

Next is the Hedonic HPI. Let's come back to Stata and run Hedonic Regression Model.

1. Create logarithms of price and bldgsqft

   Command: `gen ln_price = ln(price)`

   Command: `gen ln_bldgsqft = ln(bldgsqft)`

2. Run the linear regression: $$\ln(Price_{it}) = \alpha + \beta_1 bedrooms + \beta_2 bathrooms + \beta_3 \ln(bldgsqft) + \beta_4 centair + \beta_5 fireplace + \gamma_{2001} D_{year=2001} + \cdots + \gamma_{2016} D_{year\_2016} + \varepsilon_{it}.$$

   Command: `regress ln_price bedrooms bathrooms ln_bldgsqft centair fireplace i.year`

Hedonic HPI Model

```
. regress ln_price bedrooms bathrooms ln_bldgsqft centair fireplace i.year
```

Source	SS	df	MS		
Model	1087.91584	21	51.8055161	Number of obs	= 4,905
Residual	792.058383	4,883	.162207328	F(21, 4883)	= 319.38
				Prob > F	= 0.0000
				R-squared	= 0.5787
				Adj R-squared	= 0.5769
Total	1879.97422	4,904	.383355265	Root MSE	= .40275

ln_price	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
bedrooms	.0108292	.0101169	1.07	0.284	-.0090045	.030663
bathrooms	.0639268	.012628	5.06	0.000	.0391702	.0886833
ln_bldgsqft	.7758501	.0290961	26.67	0.000	.7188087	.8328915
centair	-.0171451	.0129034	-1.33	0.184	-.0424417	.0081514
fireplace	.294652	.0135324	21.77	0.000	.2681225	.3211816
year						
2001	.0492443	.0305742	1.61	0.107	-.010695	.1091835
2002	.1893342	.029561	6.40	0.000	.1313814	.2472871
2003	.2488985	.0292265	8.52	0.000	.1916016	.3061955
2004	.3082835	.0298027	10.34	0.000	.2498568	.3667101
2005	.4174222	.030375	13.74	0.000	.3578736	.4769708
2006	.4961729	.0314812	15.76	0.000	.4344555	.5578903

Hedonic HPI Model

3. Copy & paste the coefficient values for the year dummies (2001 to 2016).

   And, for the 2000 coefficient value, enter 0.

Year	Coefficient
2000	0.0000
2001	0.0492
2002	0.1893
2003	0.2489
2004	0.3083
2005	0.4174
2006	0.4962
2007	0.4926

Hedonic HPI Model

4. Take the exponential function of the coefficient, since the dependent variable of the linear regression was the logarithm of price ($\because \exp(\ln(price)) = price$)

   Command: =exp(B2)

Year	Coefficient	Exp(Coefficient)
2000	0.0000	1.00
2001	0.0492	=EXP(B3)
2002	0.1893	1.21
2003	0.2489	1.28

Hedonic HPI Model

5. Multiply 100 to the Exp(Coefficient) values to make the 2000 HPI value to 100.

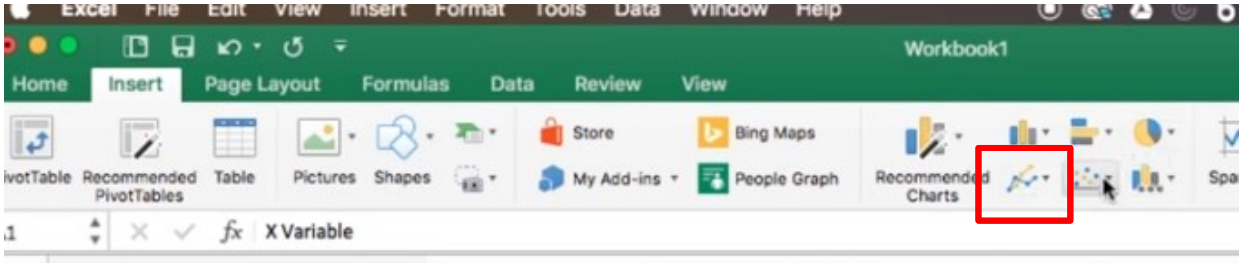
Command: =C2*100

Year	Coefficient	Exp(Coefficient)	Hedonic HPI
2000	0.0000	1.00	=C2*100
2001	0.0492	1.05	105.05
2002	0.1893	1.21	120.84
2003	0.2489	1.28	128.26

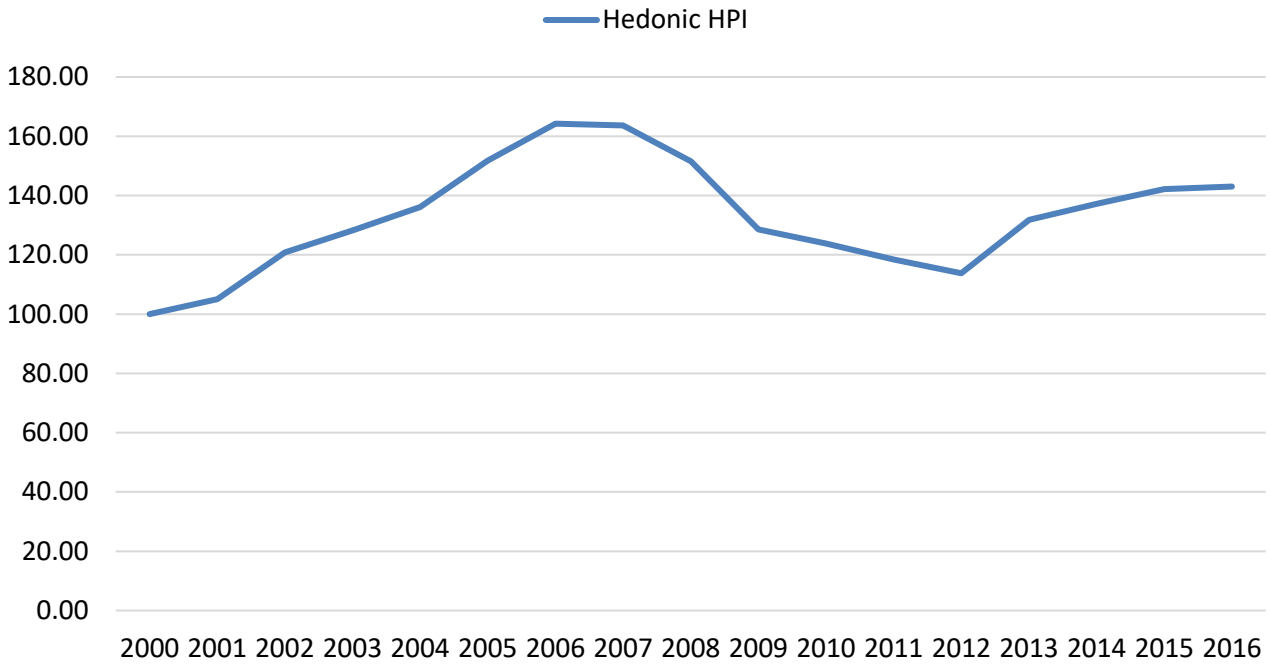
Hedonic HPI Model

6. Plot the Hedonic HPI

   Select the data > Insert > Line Graph



Hedonic HPI Model



3. (Optional) Repeat-Sales HPI

Repeat-Sales HPI Model

The last practice is the Repeat-Sales HPI Model

1. Run the same linear regression as the Hedonic HPI Model.
   But this time, we add one additional term to ensure the estimation is done within the same properties that are transacted more than twice.

   Command: `regress ln_price i.propid bedrooms bathrooms ln_bldgsqft centair fireplace i.year`

   This will take some time. While waiting for the Stata results, I will answer the questions.

Repeat-Sales HPI Model

bedrooms	0	(omitted)				
bathrooms	0	(omitted)				
ln_bldgsqft	0	(omitted)				
centair	0	(omitted)				
fireplace	0	(omitted)				
year						
2001	.1927407	.1250085	1.54	0.130	-.058744	.4442253
2002	.1776825	.0913309	1.95	0.058	-.0060515	.3614165
2003	.3835455	.0833742	4.60	0.000	.2158183	.5512728
2004	.4567866	.0808398	5.65	0.000	.294158	.6194152
2005	.5301381	.0871304	6.08	0.000	.3548543	.7054219
2006	.585651	.0926486	6.32	0.000	.399266	.7720359
2007	.383019	.1171721	3.27	0.002	.1472992	.6187388
2008	.4748535	.1495316	3.18	0.003	.1740347	.7756724
2009	.3090678	.133791	2.31	0.025	.0399149	.5782206
2010	.4402269	.1176293	3.74	0.000	.2035873	.6768664
2011	.2354484	.2257394	1.04	0.302	-.2186807	.6895776
2012	.1850152	.1076532	1.72	0.092	-.0315551	.4015855
2013	.2984959	.1096969	2.72	0.009	.0778141	.5191777
2014	.2350642	.1122369	2.09	0.042	.0092727	.4608558
2015	.3603864	.125347	2.88	0.006	.1082207	.6125521
2016	.3035756	.1001968	3.03	0.004	.1020056	.5051456
_cons	11.84357	.1725428	68.64	0.000	11.49645	12.19068

Why are these variables omitted?

Repeat-Sales HPI Model

2. Copy & paste the coefficient values of year dummies (2001 to 2016).

   And, for the 2000 coefficient value, enter 0.

Year	Coefficient
2000	0.0000
2001	0.1927
2002	0.1777
2003	0.3835
2004	0.4568
2005	0.5301
2006	0.5857
2007	0.3830

Repeat-Sales HPI Model

3. Take the exponential function of the coefficient, since the dependent variable of the linear regression was the logarithm of price ($\because \exp(\ln(price)) = price$)

Command: =exp(B2)

Year	Coefficient	Exp(Coefficient)
2000	0.0000	1.00
2001	0.1927	=EXP(B3)
2002	0.1777	1.19
2003	0.3835	1.47

Repeat-Sales HPI Model

4. Multiply 100 to the Exp(Coefficient) values to make the 2000 HPI value to 100.

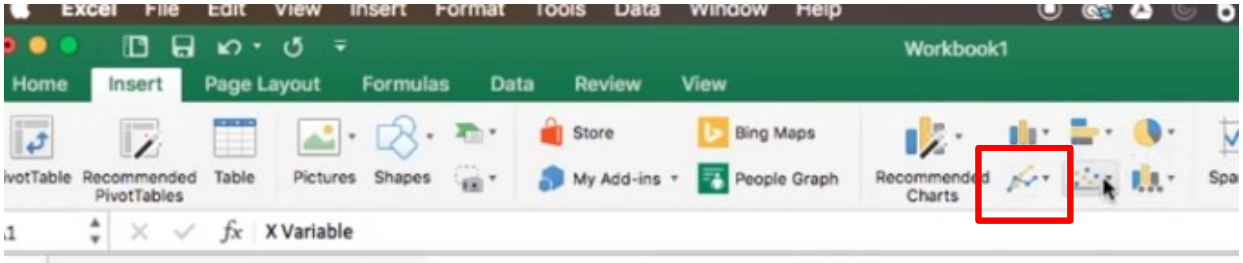
Command: =C2*100

	A	B	C	D
	Year	Coefficient	Exp(Coefficient)	Repeat-Sales HPI
	2000	0.0000	1.00	=C2*100
	2001	0.1927	1.21	121.26
	2002	0.1777	1.19	119.44
	2003	0.3835	1.47	146.75

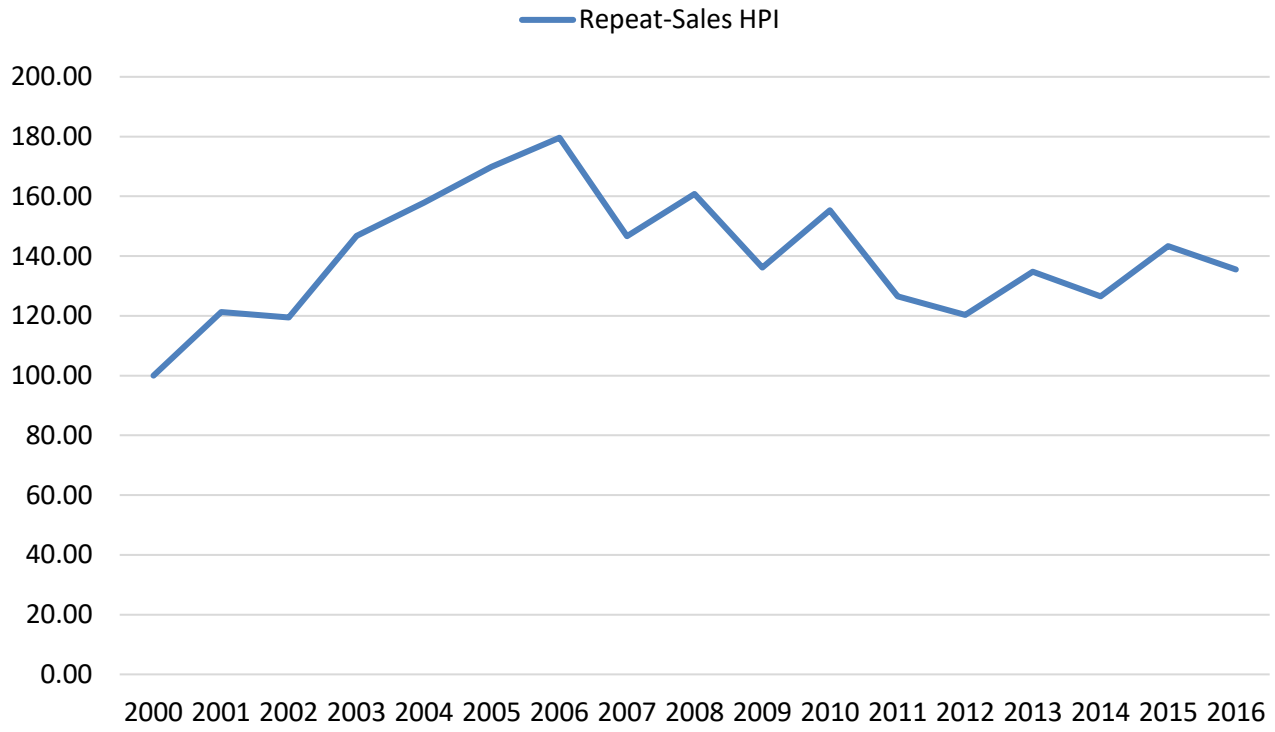
Repeat-Sales HPI Model

5. Plot the Repeat-Sales HPI

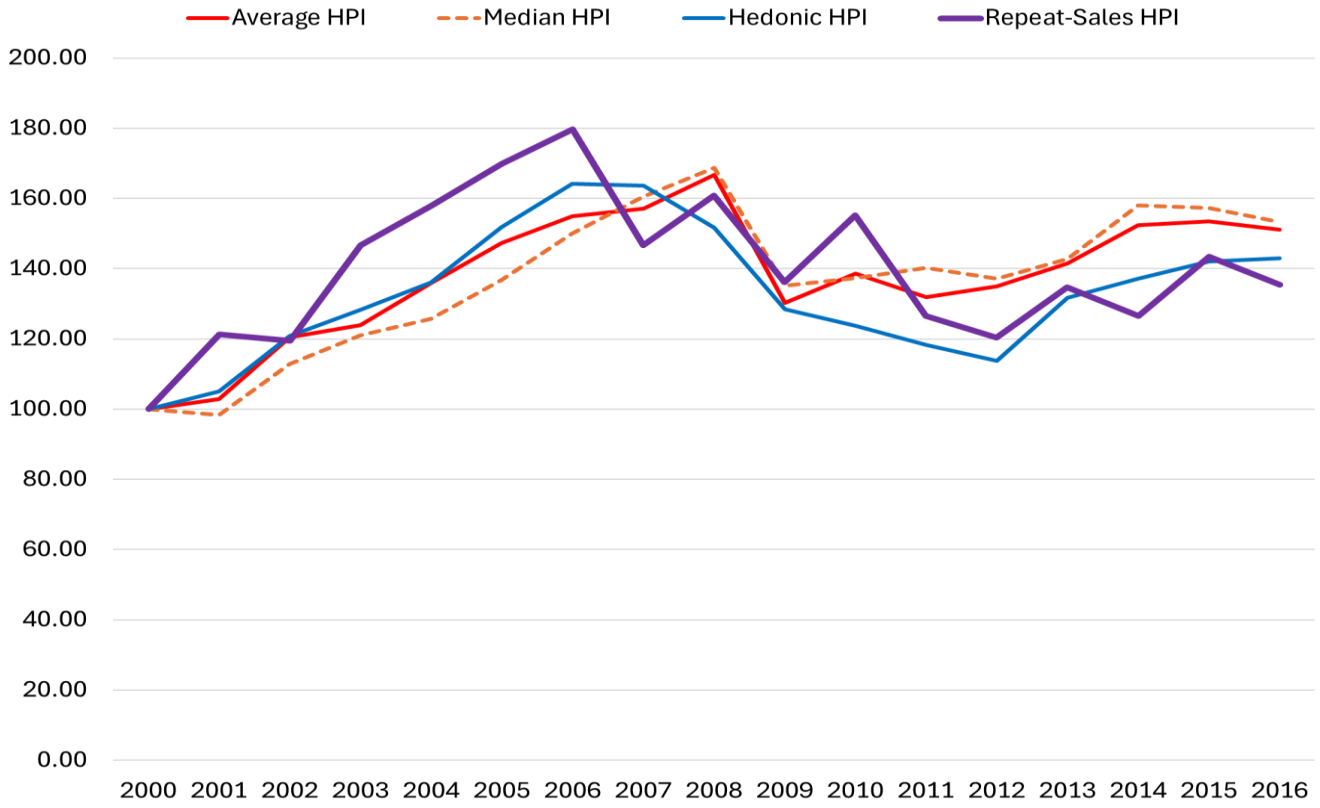
Select the data > Insert > Line Graph



Repeat-Sales HPI Model



Different HPI Models Together



Key Takeaways

- Practice the actual construction of various Housing Price Index (HPI) methodologies
- Recognize that evaluations of housing market conditions can differ based on the specific HPI being analyzed