House sales, Louisiana, 2005

i : individual house

$n = 1080$

PRICE (\$)

SQFT (area, square feet)

$1 \text{ ft} \simeq 0.3048 \text{ m}$

$1 \text{ ft}^2 \simeq 0.093 \text{ m}^2$

BEDROOMS (#)

BATHS (#)

AGE (years)

OCCUPANCY $\begin{cases} 1 & \text{seller in} \\ 2 & \text{vacant} \\ 3 & \text{rented} \end{cases}$

POOL $\begin{cases} 1 & \text{yes} \\ 0 & \text{no} \end{cases}$

STYLE $\begin{cases} 1 & \text{traditional} \\ \vdots & \\ 11 & \text{Canadian colonial} \end{cases}$

FIREPLACE $\begin{cases} 1 & \text{yes} \\ 0 & \text{no} \end{cases}$

WATERFRONT $\begin{cases} 1 & \text{yes} \\ 0 & \text{no} \end{cases}$

DOM (#)

"continuous" PRICE, SQFT, AGE, DOM

"count" BEDROOMS, BATHS

"category" OCCUPANCY, STYLE (careful with m.e's)

"binary" POOL, FIREPLACE, WATERFRONT (exclusive, exhaustive)

— cannot include both categories e.g. pool, no pool.

— if omitted category not large, danger of near multicollinearity

(b) (i) PRICE $H_0: SK = 0 \quad H_1: SK \neq 0 \quad 95\%$

$$\frac{n}{6} SK^2 \sim \chi^2(1)$$

$$\underbrace{\frac{1080}{6}(6.291909)^2}_{=7125.8614} \gg \underbrace{\chi^2_{0.95}(1)}_{=3.841} \qquad \boxed{r\ H_0\ 95\%}$$

(ii) DOM $H_0: KT = 3 \quad H_1: KT \neq 3 \quad 99\%$

$$\frac{n}{24}(KT-3)^2 \sim \chi^2(1)$$

$$\underbrace{\frac{1080}{24}(12.35337-3)^2}_{=3936.84887} \gg \underbrace{\chi^2_{0.99}(1)}_{=6.635} \qquad \boxed{r\ H_0\ 99\%}$$

(iii) log(SQFT) H_0: "normal" $(SK=0, KT=3)$ $\quad H_1: \setminus H_0 \quad 90\%$

$$JB = \frac{n}{6}\left[SK^2 + \frac{(KT-3)^2}{4}\right] \sim \chi^2(2)$$

$$\underset{=1.045963}{JB} < \underbrace{\chi^2_{0.90}(2)}_{=4.605} \qquad \boxed{d\ n\ r\ H_0\ 90\%}$$

Prob 0.59

(c) EQ01. $k = 11$.

(i), (iii)

$$\log \widehat{PRICE} \simeq \underset{(0.23)}{6.14} + \underset{(0.036)}{0.70} \log SQFT - \underset{(0.017)}{0.036}\ BEDROOMS$$

seller occupies

$$+ \underset{(0.021)}{0.22}\ BATHS - \underset{(0.00053)}{0.0054}\ AGE + \underset{(0.019)}{0.058}\ \text{"}OCCUPANCY=1\text{"}$$

$$+ \underset{(0.032)}{0.054}\ POOL - \underset{(0.017)}{0.083}\ \text{"}STYLE=1\text{"}$$

$$+ \underset{(0.020)}{0.062}\ FIREPLACE + \underset{(0.034)}{0.16}\ WATERFRONT + \underset{(0.000091)}{0.000046}\ DOM$$

(ii) $\boxed{\beta_1}$ SQFT - elasticity - of - PRICE

i.e. SQFT ↑ 10% PRICE ↑ 7%.

(iv)

individual

<u>significant</u>

99% constant, log SQFT, BATHS, AGE, OCCUPANCY=1, STYLE = 1, FIREPLACE, WATERFRONT

95% BEDROOMS

90% POOL

<u>insignificant</u> DOM i.e. no (significant) effect on PRICE

joint

$F = 280.22$ $[0.00]$ H_0: all coeffs. except constant $= 0$

$H_1: \backslash H_0$.

r H_0 99%.

model "explains something"

(v) $R^2 \simeq 72\%$.

(vi) $\bar{R}^2 \simeq 72\%$.

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k}\right)(1-R^2) \Longrightarrow \bar{R}^2 \xrightarrow[k \text{ fixed}]{n \uparrow \infty} R^2$$

(vii) $$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k} = \frac{81.97970}{1080-11} \simeq 0.077$$

$$(\hat{\sigma})^2 = (0.276926)^2 \simeq 0.077$$

(viii) log PRICE — dependent

(elasticity) log SQFT + 0.70 SQFT $\uparrow$ 100% PRICE $\uparrow$ 70%

e.g. $100m^2$ \$310,000

$200m^2$ \$527,000

BEDROOMS − 0.036 BEDROOMS $\uparrow$ 1 PRICE $\downarrow$ 4%

— controlling for other variables...

$$\log y = \alpha + \beta x + \text{controls} + u$$

$$\frac{\partial \log y}{\partial x} = \beta = \frac{\partial y / y}{\partial x} = \frac{100 \ \partial y / y}{100 \ \partial x}$$

$$\Rightarrow 100 \ \partial y / y = (100 \beta) \ \partial x \quad \text{i.e. } x \uparrow 1 \quad y \uparrow 100\beta \%$$

BATHS + 0.22 BATHS $\uparrow$ 1 PRICE $\uparrow$ 22%

AGE − 0.0054 AGE $\uparrow$ 10 PRICE $\downarrow$ 5%

"OCCUPANCY = 1" + 0.058 seller in PRICE $\uparrow$ 6%

$$\log y = \alpha + \beta D + \text{controls} + u \ ; \ D \text{ binary}$$

$$y = e^{\alpha + \beta D} \ldots$$

$$y_1 = e^{\alpha + \beta} \quad (D = 1)$$

$$y_0 = e^{\alpha} \quad (D = 0)$$

$$\frac{y_1 - y_0}{y_0} = \frac{e^{\alpha}(e^{\beta} - 1)}{e^{\alpha}} = e^{\beta} - 1$$

if β small,

$e^{\beta} \simeq 1 + \beta$

so $100(e^{\beta} - 1) \simeq 100\beta$

i.e. $D: 0 \rightarrow 1$ $y \uparrow 100(e^{\beta} - 1)\%$

POOL + 0.054 POOL PRICE $\uparrow$ 5%

"STYLE = 1" − 0.083 traditional style PRICE $\downarrow$ 8%

FIREPLACE + 0.062 FIREPLACE PRICE $\uparrow$ 6%

WATERFRONT + 0.16 WATERFRONT PRICE $\uparrow$ 16%

DOM + 0.0000046 insignificant — no effect!

(ix) $$JB_{\hat{u}} = \frac{n-k}{6}\left[SK^2 + \frac{(KT-3)^2}{4}\right] \sim \chi^2(2)$$ JB r H_0 99%

raw data $\frac{n}{6}$

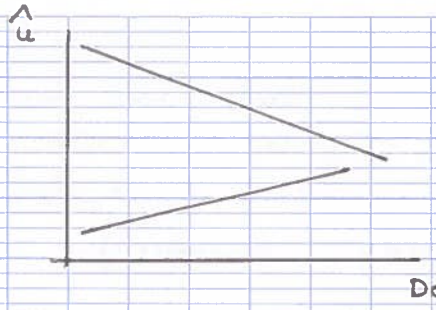
fitted errors $\frac{n-k}{6}$

$$= \frac{n-k}{n} JB \simeq 126.60 >> \chi^2_{0.99}(2)$$ $JB_{\hat{u}}$ r H_0 99%

$$\frac{1080-11}{1080}(127.9029) \qquad \chi^2_{0.99}(2) = 9.21$$

note: $JB_{\hat{u}} \rightarrow JB$ as $n \uparrow \infty$ k fixed

(x)



$\hat{u} = Mu$

$E(u) = 0 \qquad Var(u) = \sigma^2 I$

$E(\hat{u}) = 0 \qquad Var(\hat{u}) = E(Muu'M')$

$= M\,Var(u)\,M$

$= \sigma^2 M$

$(M = M' = M^2)$

if $Var(u) \neq \sigma^2 I$,

$Var(\hat{u}) = M\,Var(u)\,M$

e.g. as DOM ↑ (house on market for long time) — fewer shocks to price (depends more on characteristics), DOM ↓ — more price variation cf. characteristics

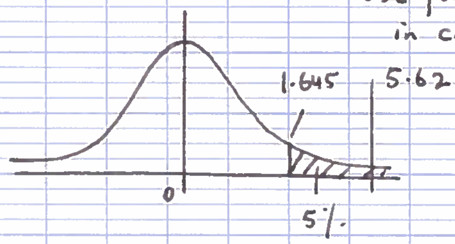
— plot suggests possible heteroscedasticity (need to test this). cf. i (here, DOM_i)

(d). $H_0: \beta_1 = 0.5 \qquad H_1: \beta_1 > 0.5 \qquad 95\%$.

$$t = \frac{0.70 - 0.5}{0.036} \simeq 5.62 \sim t(n-k) \simeq N(0,1)$$

use full accuracy in computation

$n - k = 1080 - 11 = 1069$



r H_0 95%.

⇒ evidence that SQFT ↑ 10% PRICE ↑ more than 5%.

(e) EQ03 : EQ01 + $(\log SQFT)^2$ i.e. non-constant m.e. in log SQFT (non-constant SQFT-elasticity-of-PRICE)

$$\log PRICE = \beta_0 + \beta_1 \log SQFT + \beta_2 (\log SQFT)^2 + controls + u$$

$$\frac{\partial \widehat{\log PRICE}}{\partial \log SQFT} = \hat{\beta}_1 + 2\hat{\beta}_2 \log SQFT \quad ; \quad \hat{\beta}_1, \hat{\beta}_2 \text{ individually significant}$$

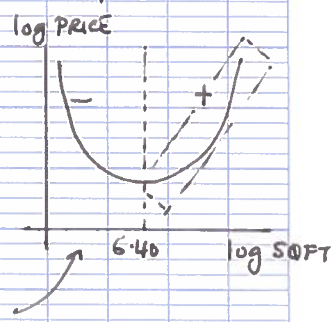
$\simeq -3.64 + 2(0.28) \log SQFT$

$\simeq -3.64 + 0.57 \log SQFT > 0$

as $\log SQFT > 6.40$

as $SQFT > 599.78$ feet^2

$(\simeq 56\ m^2)$



no observed log SQFT < 6.40 : data identifies + curvature

(f) $k = 11$ EQ01 leave 99% significant

$k = 8$ EQ02 : EQ01 − BEDROOMS, POOL, DOM

$k = 12$ EQ03 : EQ01 + $(\log \text{SQFT})^2$

$k = 10$ EQ04 : EQ03 − DOM, POOL

all individually significant 99% except BEDROOMS (almost 99% significant)

argument that EQ04 most suitable model.

— highest $\bar{R}^2$ across models $\left\{ \begin{array}{llll} \text{EQ01} & 0.7213 & \text{EQ03} & 0.7352 \\ \text{EQ02} & 0.7202 & \text{EQ04} & 0.7357 \end{array} \right.$

— individually significant parameters

— jointly significant parameters

— consider diagnostics

— consider out-of-sample performance (estimate model on subsample)

— plot $\hat{y}$ against y (scatter). : $\hat{u} = y - \hat{y}$ $\Rightarrow \hat{y} = y - \hat{u}$

(g) — not required

(a) PRICE min \$22,000 max \$1,580,000 median \$130,000

SK >> 0 some houses very high price (driven by same characteristics)

SQFT 662 feet^2 — 7897 feet^2 SK > 0

(62 m^2 — 734 m^2) very large !!

BEDROOMS few 1, 5, 6, 7, 8

2, 3, 4 well identified

mode 3

BATHS few 4, 5

1, 2, 3 well identified

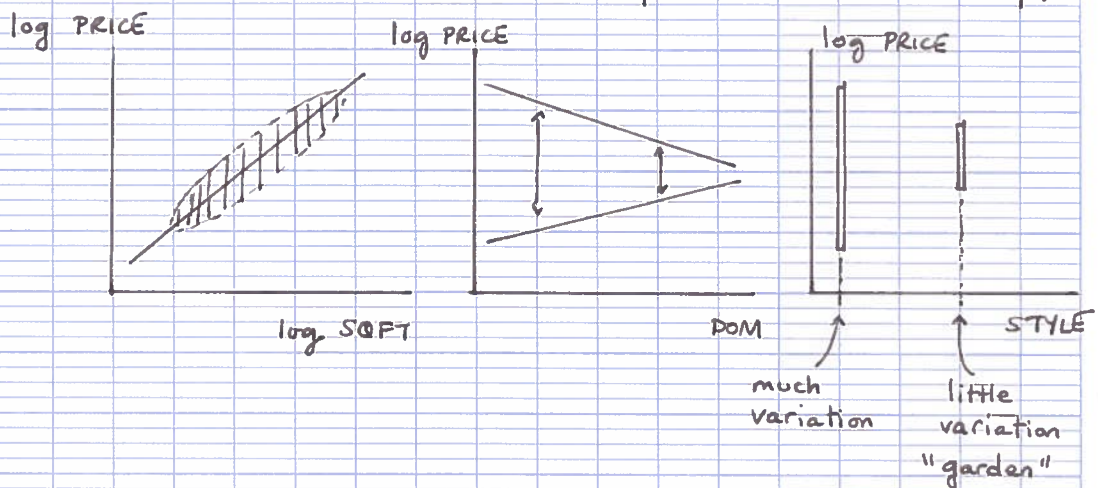
mode 2 (approx. 70% of all houses)

AGE modes 0 − 2.5 years

25 − 27.5 years

few very old houses; missing categories

OCCUPANCY	if = 3 omitted (rented), poorly identified	
POOL	mean 8%	
STYLE	mode 1 (traditional)	
	no 5 (mobile home)	
FIREPLACE	mean 56%	
WATERFRONT	mean 7% (few)	
DOM	0 – 728 days (0 – 2 years)	
	mean 74 days (1 std. dev. 95 days)	



(e) — advanced ; which m.e. is "largest" ? (in absolute value)

— compare like – with – like (binary mes ; 1 std. dev mes for non-binary)

binary :	"OCCUPANCY = 1"	POOL	"STYLE = 1"	FIREPLACE	WATERFRONT
	0.058	0.054	−0.083	0.062	$\boxed{0.16}$
non-binary : (1 std. dev. change)	log (SQFT)	BEDROOMS	BATHS	AGE	
	$\boxed{0.2862}$	−0.02533	0.1341	−0.09293	

$\hat{\beta}_j$ × std. dev. of variable (not std. error of $\hat{\beta}_j$)

WATERFRONT > "STYLE = 1" > FIREPLACE > "OCCUPANCY = 1" > POOL

log (SQFT) > BATHS > AGE > BEDROOMS