The Production Function for Housing: Evidence from France

Pierre-Philippe Combes
Aix–Marseille School of Economics and Science Po

Gilles Duranton
University of Pennsylvania

Laurent Gobillon
Paris School of Economics

Our paper

We propose a non-parametric approach to (partially) identify the production function for housing $H(K, T)$

We implement it using data on land prices and construction costs for single-family homes in France

Three results:

- The production function is nearly constant returns and Cobb-Douglas in land and structure
- The cost share of structure is around 0.8
- We formally reject constant returns and Cobb-Douglas

Motivations

Understanding of housing supply important because...

1/ Housing $\approx$ 25% of household expenditures in US and France

2/ Construction industry important for business cycle:
$\approx$ 8% of workforce in 2011 in France,
pro-cyclical and perhaps at the root of some fluctuations

3/ Housing supply is central to our understanding of cities

Housing production transforms demand for location
into land use and housing consumption

- The supply of housing is at the heart of urban models
- Housing durability is a key driver of urban dynamics
- How housing is produced has implications on welfare effects of land use regulations

Classical parametric approach in the literature

Main issue: $P\,H$ (housing value) is observed, but not H

Parametric approach since Muth (1969)

Supposing that H is CES, elasticity of substitution recovered from:

$$\log\left(\frac{K}{T}\right) = c + \sigma \log\left(\frac{R}{T}\right) + \epsilon$$

with R the land rent

Issues:

- K not observed (replaced using zero-profit condition)
- Measurement error on land prices (IV)
- Parcel heterogeneity (controls + IV)

Our approach

Express H as a function of observables using theory restrictions:

- Profit-maximizing builders choose structure K for a parcel T
- First-order condition with respect to K: $P\,\partial H/\partial K = r$
- Free entry: $P\,H - rK = R(T)$
- Eliminate P: $\partial \log H/\partial K = r/(r\,K + R(T))$
- Integrate this differential equation over K for given T

For given T, only variations of H with respect to K are identified

Source of identification: variations in land prices due to variations in demand for location (leading to variation in optimal structure)

Extension to deal with parcel heterogeneity affecting supply

For the estimations, we use information on T, K and R

Most closely related paper

Epple, Gordon and Sieg (AER, 2010)

Like us:

- Non-parametric estimation of the housing production function
- Use restrictions imposed by theory on $H = H(K, T)$

But:

- Use of different observables: PH/T and R/T
- Less direct approach
- Impose constant returns to scale (which we formally reject)
- Less suitable data
- No attempt to deal with parcel heterogeneity

Roadmap

1/ Model of housing production to impose theoretical restrictions

$\Longrightarrow$ (partial) identification and formula for housing production

2/ Estimation strategy based on this formula

3/ Presentation of the Survey of Developable Land Prices

4/ Results on the shape of production function

5/ Robustness checks when factors affecting local housing supply are taken into account

6/ Extensions: parametrized version of production function and full identification when imposing CRS

Housing production

Housing services H produced with land T and structure K

Production function $H(K, T)$ strictly increasing and concave

At location x, unit price of housing $P(x)$ and parcel area T taken as given by builders

Builders maximize their profit with respect to K; profit:

$$\pi(x) = P(x)H(K, T) - rK - R(P(x), T)$$

with r the user cost of structure and R the land rent

First-order condition

First-order condition for profit maximization with respect to K:

$$P(x)\frac{\partial H(K,T)}{\partial K} = r$$

Optimal structure: $K^* = K^*(P(x), T)$

One-to-one correspondance between unit price of housing and optimal structure

$\Longrightarrow P(x) = P(K^*, T)$ and $R(P(x), T) \equiv R(K^*, T)$

Zero-profit condition

Free-entry $\Longrightarrow$ zero-profit condition:

$$R(K^*, T) = P(K^*, T)H(K^*, T) - rK^*$$

Since P is not observed in the data, we substitute for it using FOC

We obtain a partial differential equation:

$$\frac{1}{H(K^*, T)} \frac{\partial H(K^*, T)}{\partial K} = \frac{r}{rK^* + R(K^*, T)}$$

Identification of housing production function

Suppose that for given T, location desirability varies such that:

$P(x)$ distributed over $[\underline{P}, \overline{P}]$

Then the optimal structure K^* covers the interval $[\underline{K}, \overline{K}]$

where $\underline{K} = K^*(\underline{P}, T)$ and $\overline{K} = K^*(\overline{P}, T)$

For a given optimal structure K^* (and T), housing production is:

$$\log H(K^*, T) = \int_{\underline{K}}^{K^*} \frac{r}{rK + R(K, T)} \mathrm{d}K + \log Z(T)$$

with $Z(T)$ a positive function and log the natural logarithm

Empirical sources of identification

(Partial) identification because housing prices vary

across locations

When holding area T constant, spatial variations in housing prices

$\Longrightarrow$ spatial variations in land prices

$\Longrightarrow$ spatial variations in optimal structure

For given area T, estimation of housing services as K varies

up to a multiplicative constant $Z(T)$

Estimated land rent

For given T and K, one single land rent according to theory

This land rent is estimated by kernel smoothing:

$$\widehat{R}(K, T) = \sum_i \omega_i R(K_i, T_i)$$

with:

$$\omega_i = \frac{L_{h_K}(K - K_i) L_{h_T}(T - T_i)}{\sum_i L_{h_K}(K - K_i) L_{h_T}(T - T_i)}$$

where $L_h(x) = \frac{1}{h} f\left(\frac{x}{h}\right)$ with $f(\cdot)$ the normal density

and $h_X = N^{-1/6} \sigma(X)$ with $\sigma(X)$ the empirical standard deviation

Estimation of housing production function

Housing production estimation using empirical counterpart of theoretical formula

Trapezoid approximation of integral gives:

$$\widehat{\log H}(K_i, T) = \sum_{j=2}^{i} \left(\frac{c_{j-1} + c_j}{2} \right) (K_j - K_{j-1})$$

where:

$$c_j = \frac{r}{rK_j + \widehat{R}(K_j, T)}$$

Implementation

In practice:

- 9 values for T: deciles (1^{st} to 9^{th})
- 900 values for K: equi-distributed between 1^{st} and 9^{th} deciles

We measure K with structure price

User cost of structure: $r = 6\%$
(long-term interest rate 5% + annual depreciation 1%)

Land rent: $R \approx r_T P_T$ where P_T: land price

User cost of land: $r_T = 3\%$
(long-term interest rate 5% - annual appreciation 2%)

Data on land prices and Structure

French Survey of Developable Land Prices (2006-2012)
(*Enquête des Prix des Terrains à Bâtir - EPTB*)

Data from building permits for an individual house

Nb. observations ranges from 49,000 in 2009 to 127,000 in 2012
Increasing coverage over time

Prices of the parcel and construction (decorated or not)
Location at municipality level
Characteristics: type of acquisition, area, intermediary, services

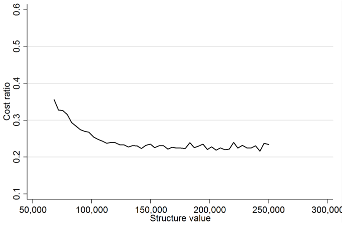
We keep only parcels purchased in mainland France at survey date

Sample: 388,805 land sales in rural areas and 352 urban areas
(city = urban area)

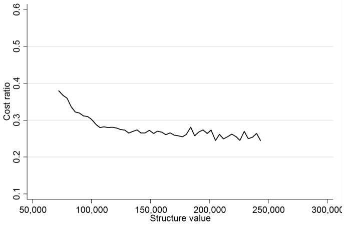
Descriptive Statistics (2012 Euros)

Variable	Mean	St. deviation	1^{st} decile	Median	9^{th} decile
Entire country:					
Parcel area	1,156	947	477	883	2,079
Construction cost	127,551	55,003	78,440	115,000	190,667
Parcel value	63,387	58,164	19,673	50,000	120,000
Parcel value per m^2	80	86	14	58	166
Urban areas:					
Parcel area	1,048	821	449	820	1,883
Construction cost	131,616	57,599	80,140	118,000	199,750
Parcel value	73,115	62,518	27,017	58,271	135,000
Parcel value per m^2	96	94	22	72	192
Greater Paris:					
Parcel area	839	673	329	665	1,493
Construction cost	151,298	73,727	89,173	132,850	236,605
Parcel value	142,010	108,598	69,155	124,419	220,000
Parcel value per m^2	237	193	67	182	466

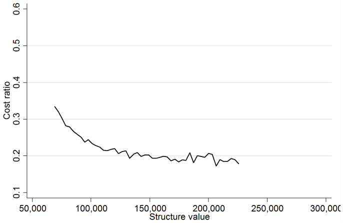
Cost ratio $R/(rK)$ as a function of structure K: graphs



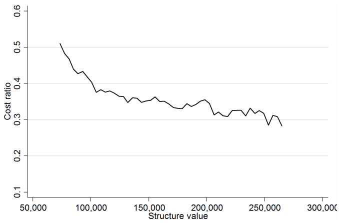
Entire country



All urban areas



UA, 50,000-100,000

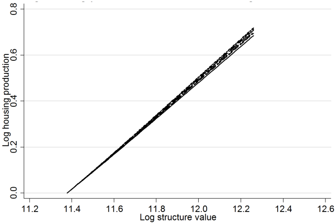


UA, $>$ 500,000 (excl. Paris)

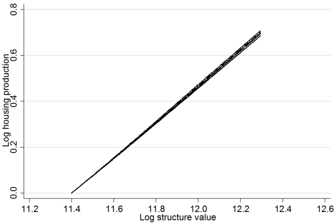
Cost ratio $R/(rK)$ as a function of structure K: comments

- Under Cobb-Douglas, $R/(rK)$ should be invariant in K
- This relationship is empirically mostly flat after 100,000 euros except for the largest urban areas
- Initial downward sloping may be due to measurement error (or by easy-to-build land fetching a higher price)
- Cost ratio larger in larger urban areas

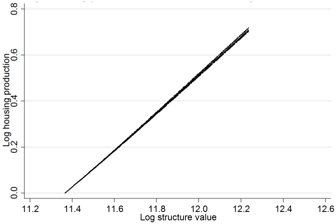
Log housing H vs. log structure K by area decile: graphs



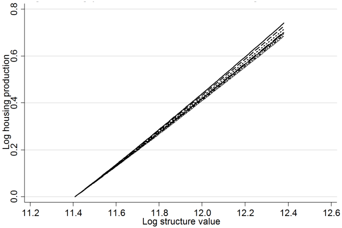
Entire country



All urban areas



UA, 50,000-100,000



UA, $>$ 500,000 (excl. Paris)

Log housing H vs. log structure K by area decile: comments (1/2)

- We plot $\log H$ as a function of $\log K$ for each decile of T as suggested by our framework
- All deciles behave in roughly the same way
- Only large urban areas appear to behave modestly differently
- The slope of these curves looks linear with slope 0.80

Log housing H vs. log structure K, OLS by area decile

Decile	1	3	5	7	9
$\log(K)$	0.779^a	0.796^a	0.808^a	0.819^a	0.818^a
	(0.00015)	(0.00020)	(0.00020)	(0.00018)	(0.00018)
R^2	1.00	1.00	1.00	1.00	1.00
Observations	900	900	900	900	900
$\log(K)$	0.360^a	0.228^a	0.266^a	0.362^a	0.329^a
	(0.00566)	(0.00857)	(0.01014)	(0.01083)	(0.01004)
$[\log(K)]^2$	0.018^a	0.024^a	0.023^a	0.019^a	0.021^a
	(0.00024)	(0.00036)	(0.00043)	(0.00046)	(0.00042)
R^2	1.00	1.00	1.00	1.00	1.00
Observations	900	900	900	900	900

a: significant at 1% level.

Log housing H vs. log structure K by area decile: comments (2/2)

- The coefficient on $\log K$ is highly significant, leads to a high R^2, and slightly varies across deciles
- Introducing higher-order terms in $\log K$ rejects Cobb-Douglas
- The production function of housing is slightly convex in K, and the share of non-land inputs is higher for larger parcels

Log housing H vs. log structure K, OLS by city size

City size class	UA	0-50	200-500	500+	Paris
$\log(K)$	0.784^a	0.832^a	0.785^a	0.730^a	0.700^a
	(0.00011)	(0.00012)	(0.00010)	(0.00023)	(0.00015)
R^2	1.00	1.00	1.00	1.00	1.00
Observations	8,100	8,100	8,100	8,100	8,100
$\log(K)$	0.365^a	-0.075^a	0.068^a	-0.091^a	-0.002
	(0.01577)	(0.00978)	(0.00647)	(0.01929)	(0.01213)
$[\log(K)]^2$	0.018^a	0.038^a	0.030^a	0.034^a	0.029^a
	(0.00041)	(0.00032)	(0.00027)	(0.00081)	(0.00051)
R^2	1.00	1.00	1.00	1.00	1.00
Observations	8,100	8,100	8,100	8,100	8,100

a: significant at 1% level; land decile fixed effects included.

Log housing H vs. log structure K by city size: comments

- The influence of K on housing production (cost share for CB) decreases with the size of urban area

Dealing with parcel heterogeneity

Identification strategy relies on (demand) variations for locations translating into variations in land prices and structure values

But land prices and structure values may also reflect (supply) variations in the ease of building

We can extend our approach to net out these (supply) variations

We only use variations of variables related to location in urban area and distance to center in that urban area

Kind of IV approach

Local supply factors

Local factors y directly affecting production $\Longrightarrow H = H(K, T, y)$

First-order condition:

$$P(x)\frac{\partial H(K,T,y)}{\partial K} = r \Longrightarrow K^* = K^*(P(x), T, y)$$

Zero-profit condition yields:

$$R(K^*, T, y) = P(x)H(K^*, T, y) - rK^*$$

Differential equation:

$$\frac{1}{H(K^*,T,y)}\frac{\partial H(K^*,T,y)}{\partial K} = \frac{r}{rK^* + R(K^*,T,y)}$$

It can be solved only for a given supply factor y

Adaptation of estimation approach

Specification before smoothing:

$$\log R_i = X_i a + Y_i b + \eta_i$$

X (resp. Y): location housing demand (resp. supply) factors

In practice, we estimate:

$$\log R_i = \beta_{c(i)} + \delta_{c(i)} d_i + B_i \gamma + \epsilon_i$$

β_c: city fixed effect, d_i: distance to city center,
B_i: other characteristics of parcel (intermediary, services, etc.)

Use of predicted rent (as well as predicted structure):

$$\widehat{\log} R_i = \hat{\beta}_{c(i)} + \hat{\delta}_{c(i)} d_i + \bar{B}\hat{\gamma}$$

Estimation results for all urban areas

Category	center	dist	dist+center	dist+center +res
$\log(K)$	0.779[a]	0.803[a]	0.783[a]	0.788[a]
	(0.00018)	(0.00007)	(0.00018)	(0.00018)
R^2	1.00	1.00	1.00	1.00
N. Obs.	8,100	8,100	8,100	8,100
$\log(K)$	2.789[a]	2.183[a]	3.663[a]	0.528[a]
	(0.01054)	(0.00592)	(0.01315)	(0.01801)
$[\log(K)]^2$	-0.084[a]	-0.058[a]	-0.121[a]	0.011[a]
	(0.00044)	(0.00025)	(0.00055)	(0.00076)
R^2	1.00	1.00	1.00	1.00
N. Obs.	8,100	8,100	8,100	8,100

a: significant at 1% level; land decile fixed effects included.

Parametric approximation of the production function

Compute the smoothed version of rent R on a 900 x 900 grid

Write the theoretical formula for cost share of structure. CES:

$$\frac{rK^*}{rK^* + R(K^*, T)} = \frac{\alpha K^{*(1-1/\sigma)}}{\alpha K^{*(1-1/\sigma)} + (1-\alpha)\, T^{1-1/\sigma}}$$

Minimize the sum of squared errors $\Longrightarrow \hat{\alpha},\, \hat{\sigma}$

Test whether the approximation is good:

- Construct counterfactual values of H on the 9 x 900 grid using its parametrized version
- Assess whether results are similar to those of non-parametric version when regressing $logH$ on $logK$ (and its square) for each decile of T

Test results of parametric approximation

Benchmark:

- Cobb-Douglas performs rather well but rejected
- CES: elasticity of substitution close to one, rejected
- 2nd order translog performs well (3rd order even better)

When "instrumenting":

- CB and CES still rejected, but CES performs better
- 2nd order translog still performs well (3rd order still better)

Full identification of the production function

Additional assumptions:

- Constant returns to scale
- Optimization with respect to T
- Land cost linear in T (but unit rent depending on x)

FOC with respect to T $\Longrightarrow$ full identification:

$$\log H(K,T) = C + \int_{\bar{T}}^{T} \frac{R(K,T)/T}{rK + R(K,T)} dT + \int_{\bar{K}}^{K} \frac{r}{rK + R(K,T)} dK + \int_{\bar{T}}^{T}\int_{\bar{K}}^{K} \frac{r\frac{\partial R}{\partial T}(K,T)}{[rK + R(K,T)]^2} dK dT$$

Fit of fully identified production function

Computation of H on the 9 x 900 grid

Regression of $logH$ on $logK$ (and its square) for each decile of T

Results differ from those when partial identification only

Production function would not be CRS

Conclusion and extensions

Non-parametric estimation of the production function for housing

Production function of housing services nearly Cobb-Douglas

Cost share of structure around 0.8

We estimated the production function taking building constraints as given

Going further:

Effect of land use restrictions (*Plans Locaux d'Urbanisme*)