A New Approach to Estimating the Production Function for Housing: Appendices

Dennis Epple

Carnegie Mellon University and NBER

Brett Gordon

Columbia University

Holger Sieg

Carnegie Mellon University and NBER

March 24, 2009

I. A Cobb-Douglas Example

To illustrate the result in Propositions 1 and it is useful to consider an example. Consider a Cobb-Douglas production function $Q = M^\alpha L^{1-\alpha}$ which implies that $q = m^\alpha$. Solving the firm's optimization problem yields:

$$(1) \qquad\qquad m(p_q) \;=\; (\alpha p_q)^{\frac{1}{1-\alpha}}$$

$$s(p_q) \;=\; (\alpha p_q)^{\frac{\alpha}{1-\alpha}}$$

as a consequence we have:

$$(2) \qquad\qquad v(p_q) \;=\; p_q \, s(p_q)$$

$$\;=\; \alpha^{\frac{\alpha}{1-\alpha}} \, p_q^{\frac{1}{1-\alpha}}$$

Inverting this function yields

$$(3) \qquad\qquad p_q(v) \;=\; \alpha^{-\alpha} \, v^{1-\alpha}$$

Moreover, it is straightforward to verify that the zero profit condition implies that:

$$(4) \qquad\qquad r(v) \;=\; (1-\alpha)\, v$$

Hence the the relationship between p_l and $r(v)$ is linear.

We can recover the Cobb-Douglas production function from this linear relationship. The differential equation for the supply function is given by:

$$(5) \qquad\qquad (1-\alpha)\,[s + p_q s'] = s$$

This can be rewritten as $\frac{s'}{s} = \frac{\alpha}{(1-\alpha)p_q}$. Integrating and rearranging, we obtain the following supply function:

$$(6) \qquad\qquad\qquad\qquad s \;\; = \;\; c\, p_q^{\frac{\alpha}{1-\alpha}}$$

where c is the constant of integration. This example illustrates how we can recover the supply function up to a constant of integration. As with any commodity, units for measuring quantity may be chosen arbitrarily as long as price per unit is chosen accordingly. Hence the constant of integration can be set by choice of any convenient normalization.

II.   Multiple Production Factors

Thus far we have shown that we can recover the housing production function if there are only two input factors. In this section we show how to extend the main results to accommodate multiple input factors. It is sufficient to consider a model with three input factors since the three-factor case easily generalizes to a model with more than three inputs. Let the three input factors be denoted by M, N, and L. For example, we can think of n as labor.[1] We assume that p_n and p_l, vary by location while p_m is fixed.[2]

We assume that we observe $V = p_q Q$, N, L, p_n, and p_l. We do not observe p_q and Q.

As before let

$$(7) \qquad \pi(p_q, p_n, p_l) \quad = \quad p_q \, s(p_q, p_n) \; - \; m(p_q, p_n) \; - \; p_n \, n(p_q, p_n) \; - \; p_l$$

denote the indirect profit function, where $s(p_q, p_n)$ is the supply function per unit of land, and $n(p_q, p_n)$ is the indirect factor demand per unit of land. The Envelope Theorem implies that

$$(8) \qquad \frac{\partial \pi}{\partial p_q} \quad = \quad s(p_q, p_n)$$
$$\frac{\partial \pi}{\partial p_n} \quad = \quad -n(p_q, p_n)$$

Using a similar logic as in the previous section, we can show that

$$(9) \qquad v \quad = \quad p_q \, s(p_q, p_n)$$

---

[1] John Kennan and James Walker (2005) document large differences in wage rates across metropolitan areas in the U.S. Thus the three-input production function could potentially be identified by inter-metropolitan wage variation coupled with inter- and intra-metropolitan land price variation.

[2] The case in which the price of only one input factor varies by location is formally equivalent to the two input case in Section 2.1. All inputs that do not vary in price can be treated as a single composite good.

is a monotonic function of p_q holding p_n fixed. The inverse function $p_q = p_q(v, p_n)$ therefore exists. As a consequence, the generalization of Proposition 1 holds and there exists an equilibrium locus that relates the p_l to v given p_n:

$$(10) \qquad\qquad p_l \quad = \quad r(v, p_n)$$

Using a similar argument, we can also show that there exist a function $n^*(v, p_n) = n(p_q(v, p_n), p_n)$. The existence of the equilibrium locus $p_l = r(v, p_n)$ in equation (10) then implies the following alternative representation of the indirect profit function:

$$(11) \qquad\qquad \pi^*(v(p_q, p_n), p_n) \quad = \quad r(p_q \; s(p_q, p_n), p_n) - p_l$$

Taking partial derivatives, we have:

$$(12) \qquad\qquad \frac{\partial \pi^*}{\partial p_q} \quad = \quad \frac{\partial r}{\partial v} \frac{\partial v}{\partial p_q}$$
$$= \quad \frac{\partial r}{\partial v} \left[s(p_q, p_n) + p_q \frac{\partial s(p_q, p_n)}{\partial p_q} \right]$$

and

$$(13) \qquad\qquad \frac{\partial \pi^*}{\partial p_n} \quad = \quad \frac{\partial r}{\partial v} \frac{\partial v}{\partial p_n} + \frac{\partial r}{\partial p_n}$$
$$= \quad \frac{\partial r}{\partial v} p_q \frac{\partial s(p_q, p_n)}{\partial p_n} + \frac{\partial r}{\partial p_n}$$

Note that in equilibrium

$$(14) \qquad\qquad \pi(p_q, p_n) \quad \equiv \quad \pi^*(p_q, p_n)$$

$$n(p_q, p_n) \quad \equiv \quad n^*(p_q s(p_q, p_n), p_n)$$

4

We thus obtain the generalization of Proposition 2:

The supply function per unit of land is the solution to the following fundamental system of partial differential equations:

$$
(15) \qquad s(p_q, p_n) = \frac{\partial r}{\partial v}\left[s(p_q, p_n) + p_q \frac{\partial s(p_q, p_n)}{\partial p_q}\right]
$$

$$
-n^*(p_q s(p_q, p_n), p_n) = \frac{\partial r}{\partial v} p_q \frac{\partial s(p_q, p_n)}{\partial p_n} + \frac{\partial r}{\partial p_n}
$$

Note that the functions r and n^* only depend on observables. In general, a closed-form solution to the system of partial differential equations will not exist and we must rely on numerical methods to compute the supply function. Once we have obtained the supply function, we can recover the production using the following simple algorithm. Given that we know the functions $n^*(v, p_n)$ and $m^*(v, p_n)$, we can compute

$$
(16) \qquad n(p_q, p_n) = n^*(p_q s(p_q, p_n), p_n)
$$

$$
m(p_q, p_n) = m^*(p_q s(p_q, p_n), p_n)
$$

By varying p_q and p_n We can therefore trace out the relationship between n, m and $q = s(p_q, p_n)$.[3]

---

[3]Generalizing the results to production functions with more than three input factors is straightforward.

III. Interpreting the Errors as Productivity Shocks

The equation

$$(17) \qquad\qquad p_l = r(v)$$

is an equilibrium relationship between two endogenous variables, the price per unit of land, p_l, and the value of housing per unit of land, v. To estimate equation (17), we must provide a characterization of the sources of disturbances in this relationship. It is natural to presume that there is error in measuring the price per unit of land and the value of of housing per unit of land:

$$
\begin{aligned}
\tilde{p}_l &= p_l + \epsilon_{p_l} \\
\tilde{v} &= v + \epsilon_v
\end{aligned}
$$

The ϵ_v may reflect either measurement error or productivity shocks. As an example of the latter, architectural plans that look good on paper may produce a house that lacks aesthetic appeal, a negative shock to the value of the dwelling, or a house that is viewed as having unusual "curb appeal," a positive shock. To develop this in more detail, suppose the realized price per unit of housing services for a particular dwelling is

$$(18) \qquad\qquad \tilde{p}_q = p_q + \epsilon_p$$

where ϵ_p is independent of p_q and ϵ_p is not observed until the dwelling is completed. The firm's profit per unit land is then:

$$(19) \qquad\qquad \tilde{\pi} = \tilde{p}_q q(m) - p_m m - p_l$$

Assuming home builders are risk neutral, the firm then maximizes the expected profit:

$$(20) \qquad \pi = p_q q(m) - p_m m - p_l$$

Our theoretical development then proceeds exactly as before. In particular, the supply function is in terms of the expected price, p_q, and the zero-expected-profit condition is likewise defined in terms of p_q.

When the dwelling is completed, the builder obtains the following revenue from sale of the house:

$$(21) \qquad \tilde{v} = (p_q + \epsilon_p)q = v + \epsilon_v$$

where $\epsilon_v = q \cdot \epsilon_p$

Thus, whether ϵ_v is measurement error or the result of uncertainty about the appeal of the completed structure, we have:

$$(22) \qquad \tilde{p}_l = r(\tilde{v} - \epsilon_v) + \epsilon_p$$

If we assume that ϵ_v and ϵ_p are mutually independent and independent of p_l and v, we can estimate this relationship using methods developed by Hausmann, Newey, and Powell (1995) and Newey (2001). Implementing these estimation approaches requires instruments.

IV.   Housing Production in Wake County, North Carolina

This section provides an additional application of our methodology to Wake County in North Carolina. Wake County provides free access to a detailed database containing information on property characteristics and real estate assessment data.[4] The appraisals are conducted by the Wake County Revenue Department. An appraiser in the Wake County Revenue Department indicated that, to assure their land is being treated on an equal footing with undeveloped properties, owners of developed properties have the right to know land value separate from building value.

While extensive documentation is available regarding the methods for land valuation in Allegheny County, less detail is available regarding the approach used in Wake County. Our sense is that Wake County follows a similar methodology, using the comparable sales method to value land. In the absence of land sales, they extrapolate the land value based on the residual obtained from subtracting the depreciated building value from the sale price.[5]

The complete database lists 343,886 properties, of which 258,866 are habitable residential properties. We limit our subsample to the set of residential properties built after 2000 and remove any properties that do not have positive lot area sizes and market values listed. After geocoding the properties, this produces a final sample size of 6,263. Table 1 provides descriptive statistics. While the mean and median value per unit land and price of land are both higher than in Pittsburgh, lot sizes in the two areas are roughly the same. This may reflect the fact that the Raleigh-Durham area

---

[4]The data are freely available for download at http://www.wakegov.com/tax/downloads/default.htm.

[5]In some cases, land values may be inferred by rule of thumb. For example, visual inspection of a plot of land prices versus property values per unit land reveals a subset of properties for which land and property values are noticeably at variance with the pattern evident in the remainder of the data. We removed those observations for the results reported here.These properties are relatively new and close to the city center. We speculate that land valuation may have been more difficult for these properties, perhaps due to a paucity of recent vacant land sales available for comparison.

has experienced rapid population growth—and demand for new housing—over the past decade.[6]

Table 1. Descriptive Statistics

Sample of Residential Real Estate					
Variable	Mean	Median	Stdev	Min	Max
Value per unit of land, v	27.20	22.21	24.83	0.3863	229.66
Price of land, p_l	5.82	4.85	4.80	0.1739	34.44
Lot area (sq ft)	27323	13068	42221	871	422968
Distance (miles)	10.68	10.80	4.17	0.1934	21.1459

The size of the residential sample is 6,263.

The results of estimating the function $r(v)$ are in Table 2. We used distance (in miles) to the city center and zip-code dummies as instruments for the HNIP estimates. The parameter estimates are slightly higher than the values we obtained in Pittsburgh. This discrepancy could be due to fundamental differences in the technology of housing production between the areas. The warmer climate in Wake County may reduce the share of non-land inputs required to produce a given quantity of housing services because there is less need to build houses that can withstand the rigors of the seasonal weather variations in the north.

Given the estimates of $r(v)$, we estimate the supply function for a variety of specifications. The average price elasticity, based on elasticity for each observation in the sample, ranges from 2.91 in the quadratic case to 4.00 in the fully nonparametric case. These values are broadly consistent with those obtained in our application to Allegheny County.

---

[6] "Raleigh's growth is showing," *News and Observer* (Raleigh, NC), June 21, 2006.

Table 2. Estimates of Equilibrium Locus

	OLS Estimates			
	Log-linear	Linear	Quadratic	Cubic
v		0.2000***	0.2395***	0.2251***
v^2			-0.0005***	-0.00008
v^3				0.000002***
Constant	-1.3152***			
$\log(v)$	0.9301***			
N	6,263	6,263	6,263	6,263
	HNIP Estimates			
	Log-linear	Linear	Quadratic	Cubic
v		0.2138 ***	0.2221***	0.1811***
v^2			-0.0002*	-0.0023**
v^3				0.00005***
Constant	-1.7677***			
$\log(v)$	1.0836***			
N	6,263	6,263	6,263	6,263

* indicates significance at the 90% level, ** at the 95% level,

and *** at the 99% level.

V. An Alternative Application: The Car Repair Service Industry

This alternative application focuses on the car repair service industry. We have obtained a unique data set that is based on surveys conducted for Underhood Service Magazine. Underhood Service targets repair shops that derive 50 percent or more of their revenue from the service and repair of under-the-hood systems. The contributing writers for Underhood Service are primarily the owners and managers of independent automotive repair businesses. Underhood Magazine is owned by Babcox which is located in Akron, Ohio, and has been in the automotive aftermarket publishing industry since 1920. Underhood magazine has conducted surveys of the industry for a number of years. Thanks to the generous help of Bob Roberts, the Marketing Research Manager of Babcox, we have obtained access to this data base.

Our analysis is based on the survey that was conducted in 2004. This survey was conducted in two parts. The 2004 survey was based on a random sample of 4000 subscribers. 102 people returned Part A and 139 returned Part B, for a total 6 % response rate.[7] 89 % of the respondents are the shop owners and 11 % are managers. Despite the low response rate, the sample seems to be representative and covers all regions of the U.S. Nearly all of the respondents of Part A of the survey operate a single repair shop. Each shop has an average of 4.4 (2.84) repair bays. 14% are in areas with a population greater than 500,000 individuals, 22% with populations ranging from 100,000 to 500,000, 32 % with populations between 15,000 and 100,000, and 32 % with populations below 15,000. The repair shops have an average of 3.7 full-time employees. About half the shops use part-time employees with an average of 1.8 part-timers. The total number of employees is broken down into job categories (owner, manager, technician, sales, clean-up, office, combo, and

---

[7]A respondent received a free 1 year extension on their subscription to Underhood Service magazine valued at $64.

other). The number of technicians ranges between one and six. Part B of the survey focuses on different aspects of the business. The majority are family owned and have been in business for almost 20 years on average. The mean hourly wage rate in the sample is $58.54 (14.01).

Table 3. Price Dispersion in Car Repair Services

state	min	max	state	min	max
Diagnostic Check only					
California	20	300	Ohio	35	98
Brake Repair					
California	110	600	Pennsylvania	75	400
Spark Plug Replacement					
Florida	60	400	Indiana	50	300
Oil and Lube Job					
Florida	25	100	Wisconsin	20	36

We have emphasized in this paper that output prices for similar types of good vary quite substantially in various industries in the U.S. One of the nice features of this data set is that it allows us to document this price dispersion. Table 3 illustrates the degree of price dispersion for various standard repairs in selected states. We find that there is a significant amount of price dispersion both within and across states.

We proceed and estimate a production function for basic car repair services. We combined the data from the two parts of the survey. After removing observations with incomplete information, we are left with 97 observations. The average number of technicians in our sample is 2.39 (1.17), the average annual salary of a technician is $38,016 (15,610), and the average annual revenue per technician is $ 144,364 (62,305). In our first stage regressions we use revenue per technician as the dependent variable and a technician's annual salary as the regressor. Again we estimated

different functional specifications of the first stage model. We find that noise in the data causes the higher-order polynomial forms to over-fit, which leads to the conditions in Proposition 1 not being satisfied. We conclude that the linear specification appears most reasonable. Note that this implies the production function is Cobb-Douglas. The estimates of associated the supply function imply that the price elasticity is 4.23. The elasticity of substitution is, of course, equal to one.

References

Kennan, John and Walker, James. 2005 . The Effect of Expected Income on Individual Migration

Decisions. Working Paper.