# A New Approach to Estimating the Production Function for Housing

By Dennis Epple, Brett Gordon, and Holger Sieg*

*Dating to the classic works of Alonso, Mills, and Muth, the production function for housing has played a central role in urban economics and local public finance. This paper provides a new flexible approach for estimating the housing production function which treats housing quantities and prices as latent variables. The empirical analysis is based on a comprehensive database of recently built properties in Allegheny County, Pennsylvania. We find that the new method proposed in this paper works well in the application and provides reasonable estimates for the underlying production function. (JEL : C51, L11, R12)*

Understanding the relationship between housing markets and the urban economy has been at the center of research in modern urban economics.[1] Dating at least to the classic works of William Alonso (1964), Edwin Mills (1967), and Richard Muth (1969), the production function for housing has played an important role in explaining how households and firms sort in an urban economy. Variation in amenities and publicly provided goods (e.g., education) across locations impacts the demand for locations. This in turn gives rise to variation across locations in land prices, investment in structures per unit of land, and housing prices. The housing production function is the funda-

---

[1]New residential construction is an important industry in the U.S. economy. According to Lawrence Smith, Lawrence Kenneth Rosen and George Fallis (1988), it accounts for five percent of GNP and employs four percent of the labor force. Housing services are one of the largest single items in a typical household's consumption bundle. The cost of housing services and household operation comprises 39 percent of the consumer price index.

mental technological channel through which variation in attractiveness of locations and variation in household willingness to pay for housing and amenities are translated into land use patterns, housing consumption, and the spatial distribution of households in metropolitan areas. The importance of such spatial patterns may be magnified through endogenous variation in neighborhoods, variation which may reflect local social interactions, heterogeneity in household demographic characteristics and preferences, or spatial agglomeration effects on firm productivities.[2]

More recently, a number of researchers have forcefully argued that housing supply is also key to understanding the dynamic development of urban areas. Edward Glaeser and Joseph (2005) focus on the role that durable housing stocks play in mediating urban decline.[3] Edward Glaeser, Joseph Gyourko, and Ravin Saks (2005) consider the relationship between urban growth and new residential construction. They argue that the housing supply elasticity is the key parameter that governs growth dynamics. If, for example, housing supply is price inelastic, a positive regional productivity shock or an increase in the attractiveness of a city will primarily lead to higher paid workers and more expensive houses. If the housing supply is elastic, one might expect smaller price changes and larger adjustments in city sizes. Understanding the production and supply of new residential housing is, therefore, important from a dynamic perspective.

The production function for housing entails a powerful abstraction. Houses are viewed as differing only in the quantity of services they provide, with housing services being homogeneous and divisible.[4] Thus, a grand house and a modest house differ only in the number of homogeneous service units they contain. In addition, housing is presumed to be produced from land and non-land

---

[2] Firm location decisions in the presence of agglomeration effects are studied in Robert Lucas and Esteban Rossi-Hansberg (2002).

[3] Similarly, Jan Brueckner and Stuart Rosenthal (2005) document the existence housing cycles and study the impact of housing on the gentrification of neighborhoods.

[4] Richard Muth (1960) and Ed Olson (1969) introduced the assumption that there exists an unobservable homogeneous commodity called housing services.

factors via a constant returns production function. Estimating the housing production function is then an important undertaking to facilitate the calibration of computational models and to provide evidence about the validity of the strong assumptions entailed in the concept of a housing production function.

Estimating housing production functions is, however, challenging.[5] The key problem encountered in estimation is that that the quantity of housing services per dwelling and the price per unit of housing services are not observed by the econometrician. Instead, we observe the value of a housing unit which is, by definition, the product of the price and quantity. Separating these two sources is a daunting task, especially if we are not willing to rely on *ad hoc* decomposition procedures. The main objective of this paper is to develop and apply a new technique for estimating the housing production function which properly treats the quantity and the price of housing services as latent variables unobserved by the econometrician.

Our approach to identification and estimation is thus based on duality theory.[6] We assume that the housing production function satisfies constant returns to scale.[7] We can, therefore, normalize output in terms of land use. While we do not observe the price or quantity of housing, we observe the value of housing per unit of land. We show in this paper that the price of housing is a monotonically increasing function of the value of housing per unit of land. Since the price of housing is unobserved, the attention thus focuses on value of housing per unit of land instead. Constant returns to scale also implies that profits of land developers must be zero in equilibrium. We exploit the zero profit

---

[5]Some early empirical papers on housing supply include Frank DeLeeuw and Nkanta Ekanem (1971) and Barton Smith (1976). John McDonald (1981) reviews the early empirical literature on housing production.

[6]Duality between the production function and the price possibility frontier was introduced by Paul Samuelson (1953), and is discussed, among others, in Edwin Burmeister and Kiyoshi Kuga (1970), Ronald Shepherd (1970), William Diewert (1973), and Dale Jorgenson and Lawrence Lau (1974).

[7]While this assumption is fairly standard in the literature on housing construction, it might be more controversial for other industries. However, Susanto Basu and John Fernald (1997) reports estimates for 34 industries in the U.S. that suggests that a typical industry has approximately constant returns to scale.

condition and derive an alternative representation of the indirect profit function as a function of the price of land and value of housing per unit of land.

Differentiating the alternative representation of the indirect profit function with the respect to the (unobserved) price of housing gives rise to a differential equation that implicitly characterizes the supply function per unit of land. Most importantly, this differential equation only depends on functions that can be consistently estimated by the econometrician. Moreover, we show that this differential equation has an analytical solution for some well-known parametric production functions such as Cobb-Douglas. In general, analytical solutions do not exist and we provide an algorithm that can be used to numerically compute the supply function per unit of land for arbitrary functional forms. With the supply function in hand, it is then straightforward to derive the housing production function. Finally, we show that the approach extends to the case with more than two input factors. In the general case the supply function per unit of an input is implicitly characterized by the solution to a system of partial differential equations. This system only depends on functions that can be estimated under suitable regularity conditions.

The theoretical results derived in this paper directly map into a flexible estimation procedure. We use semi-nonparametric and nonparametric techniques to estimate the alternative representation of the indirect profit function.[8] The derivative of the alternative representation of the indirect profit function is the key ingredient in the differential equation that characterizes the supply function per unit of land. Economic theory implies that the function that relates the price of land to the value of housing per unit of land must be monotonically increasing in the value of housing per unit of land and that the derivative of this function is bounded by one. One advantage of the

---

[8]William Diewert (1971) and Lawrits Christensen, Dale Jorgenson, and Lawrence Lau (1973) were the first to suggest the use of flexible (parametric) forms in estimation. Ronald Gallant (1981) introduced flexible semi-nonparametric techniques based on Fourier functions. Hrishikesh Vinod and Annan Ullah (1988) suggested the use of nonparametric kernel estimators.

semi-nonparametric approach adopted in this paper is that we can easily impose these types of shape restrictions in estimation. Thus, the approach proposed in this paper allows us to identify and estimate production functions with minimal functional form assumptions.

Our empirical application focuses on the Pittsburgh Metropolitan Area. We have obtained access to a data set which includes all housing units in Allegheny county. In contrast to some other publicly available data sets, this data set contains separate data for the value of land and the value of the dwelling.[9] We implement our estimation procedure using the sub-sample of housing units that were build after 1995. Despite the fact that there has been little population growth in the Pittsburgh metropolitan area during the past decades, we observe a large amount of new housing construction in that time period. We find that the approach suggested in this paper yields plausible and robust estimates of the underlying production function. Moreover, there exist some simple parametric forms which are broadly consistent with the data.

The rest of the paper is organized as follows. Section I discusses the main theoretical results regarding supply and production functions of housing. Section II discusses the estimation procedure used in this paper. Section III introduces the data set on which the empirical application is based. Section IV present the main empirical findings. Section V discusses the policy implications of the paper and offers some concluding remarks.

---

[9]There is a small literature that measure the price for vacant land. Andrew Haughwout, James Orr, and David Bedoll (2008) study commercial land prices in New York City.

# I. The Theory of Housing Production

## A. *The Model*

The starting point of the analysis is the canonical model of housing production that underlies almost all theoretical models in modern urban and local public economics. This model assumes that housing can be treated as a homogeneous and perfectly divisible good denoted by $Q$. In the baseline model, housing is produced from two factors $M$ and $L$ via a production function $Q(L, M)$.[10] We can think of $L$ as land and $M$ as a composite of all mobile non-land factors. It is reasonable to treat the price of $M$, denoted by $p_m$, as constant across all locations in a metropolitan area. The price of land, $p_l$, depends on location.[11]

Since housing is a non-tradeable good, the price of housing, $p_q$, must also depend on location. It is typically assumed that the underlying production function has the following properties:[12]

ASSUMPTION 1: *The housing production function $Q(L, M)$*

a. *exhibits constant returns to scale, implying $Q(L, M) = L \cdot Q(1, M/L)$;*

b. *is strictly increasing, strictly concave, and twice differentiable.*

The constant returns to scale assumption is typically motivated by the observation that the housing construction industry is characterized by firms of varied sizes.[13] This fact is consistent with constant returns to scale since average costs are independently of firm size. In principle, it should not

---

[10]We present the basic theoretical results for a model with two input factors. We discuss the extension to a model with three and more input factors in an appendix that is available upon request from the authors.

[11]This variation in land prices largely arises from differences in proximity to places of employment and commerce, access and availability of public goods and amenities among locations. Dennis Epple and Holger Sieg (1999), Dennis Epple, Thomas Romer, and Holger Sieg (2001), Holger Sieg, V. Kerry Smith, Spencer Banzhaf, and Randall Walsh (2004), Patrick Bayer, Fernando Ferreira, and Robert McMillan (2007), Maria Ferreyra (2007) and others provide ample evidence that consumers value these types of urban amenities.

[12]Lawrits Christensen, Dale Jorgenson, and Lawrence Lau (1973) use the similar qualitative assumptions.

[13]The 2002 Economic Census reports 171,655 establishments in home building (Industry Code 2361), of which 112,012 are incorporated. Source: Industry General Summary, 2002 Economic Census, Construction Subject Series, EC02-23SG-1, October 2005.

be difficult for an efficiently operated construction firm to build similar houses on similar lots for approximately the same costs. There are few fixed factors that could be appealed to justify decreasing returns to scale. Managerial ability might be the only factor that is fixed in the short run, but clearly not in the long run.[14] Similarly, it is difficult to find a compelling reason that would justify an increasing returns to scale scenario. There is no empirical evidence that would suggest the presence of strong agglomeration or other spill-overs that might be used to justify increasing returns to scale.

We also assume that housing is produced by a large number of firms that make up the industry, i.e. the housing production industry is competitive and thus satisfies the following assumption.

ASSUMPTION 2: *Firms behave as price takers.*

Like housing producers, landowners are price takers. the price of land will be higher in more desirable locations, but we assume that no landowner is large enough to influence the market price.

As with any constant returns to scale technology, the size of the individual firm is indeterminate, but optimal input ratios are well-defined. Writing all variables on a per-unit of land basis, let $m = \frac{M}{L}$ and $q(m) = Q(1, \frac{M}{L})$. The firm's profit per unit of land can then be written:

$$(1) \qquad\qquad \pi \;=\; \frac{\Pi}{L} \;=\; p_q\, q(m) \;-\; p_m\, m \;-\; p_l$$

Since $p_m$ is constant throughout the population, and $m$ can be measured in arbitrary units, we henceforth adopt the normalization $p_m = 1$.[15]

---

[14]The assumption of constant returns to scale at the firm level, of course, does not mean that constant returns is a good assumption in housing production for an entire city or metropolitan area, due to land use constraint (Smith, 1976) and our approach does not invoke such an assumption.

[15]In some applications, $p_m$ may vary as would be the case, for example, with pooling of data from different metropolitan areas. When $p_m$ varies, all results in this paper hold with $p_q$, $p_l$, and $v$ replaced with $\frac{p_q}{p_m}$, $\frac{p_l}{p_m}$, and $\frac{v}{p_m}$ respectively.

Let $s(p_q)$ denote the normalized supply function, i.e. the supply function per unit of land. Assumption 1 implies that $s(p_q)$ is strictly increasing in $p_q$, $s(p_q) > 0$ for $p_q > 0$, and $s(p_q)$ approaches zero as $p_q$ approaches zero.[16]

Furthermore, let $m(p_q)$ denote the normalized factor demand function. We can then define the indirect profit function per unit of land as

(2)
$$\pi(p_q, p_l) = p_q s(p_q) - m(p_q) - p_l$$

By the envelope theorem, we have:

(3)
$$\frac{\partial \pi(p_q, p_l)}{\partial p_q} = s(p_q)$$

The derivative of this profit function is equal to the supply function per unit of land. Computation of $s(p_q)$ is, therefore, simple if we can compute (or estimate) the indirect profit function.[17]

A key feature of housing markets is that quantities and prices of housing are not observed separately by the econometrician. The remainder of this paper is thus motivated by the following assumption on observables:

ASSUMPTION 3: *We observe the value of housing per unit of land, denoted by $v$. We also observe $p_l$ and $L$. We do not observe $p_q$ or $Q$.*

Our first goal is to show that we can recover $s(p_q)$ under Assumptions 1-3. Once we have obtained $s(p_q)$, it is then straight forward to recover the production function $q(m)$.

---

[16]It is useful to distinguish between the supply per unit of $L$ and the total supply. It is well-known that a supply function does not exist if the production function has constant returns to scale. The supply is either zero (if per-unit profits are negative), indeterminate (if per-unit profits are zero), or infinite (if per-unit profits are positive). The supply function per unit of land is, however, well-defined since it treats $L$ as a fixed factor.

[17]The normalized indirect profit is closely related to Paul Samuelson's (1953) factor-price frontier. Lawrits Christensen, Dale Jorgenson, and Lawrence Lau (1973) refer to the factor price frontier as the price possibility frontier and show how to estimate the parameters of a trans-log function with multiple outputs based on this dual representation of the production function.

Our approach is based on duality theory. First, we show that there exists a monotonic relationship between $p_q$ and $v$. Since $p_q$ is unobserved, we focus on $v$ instead. Second, we show that there is another monotonic function that captures the equilibrium relationship between $v$ and $p_l$. Finally, we show that one can recover an alternative representation of the indirect profit function based on the observed equilibrium relationship between $p_l$ and $v$. This alternative representation of the indirect profit function gives rise to a differential equation that defines $s(p_q)$ up to a constant of integration.[18] Summarizing the first two steps of our analysis, we have the following result:

PROPOSITION 1: *The value of housing per unit of land $v$ is a monotonic function of $p_q$. As a consequence there exists a function $r(v)$ such that in equilibrium the following is true:*

$$(4) \qquad\qquad p_l \;=\; r(v)$$

Proof:

The value of housing per unit of land is defined as:

$$(5) \qquad\qquad v \;=\; p_q s(p_q) \;=\; v(p_q)$$

Since $s(p_q)$ is monotonically increasing and differentiable, it follows that $v(p_q)$ is a monotonically increasing, differentiable function of $p_q$. Hence, this function can be inverted to obtain:

$$(6) \qquad\qquad p_q \;=\; p_q(v)$$

Substituting (6) into the indirect profit function (2) and invoking the zero profit condition implies:

$$(7) \qquad\qquad p_l \;=\; p_q(v)\, s(p_q(v)) \;-\; m(p_q(v)) \;\equiv\; r(v)$$

---

[18]An appendix that illustrates the results in Propositions 1 and 2 using a Cobb-Douglas example is available upon request from the authors.

## B. *An Alternative Characterization of the Supply Function*

Based on the equilibrium locus $p_l = r(v)$ in equation (7), we can derive an alternative characterization of the indirect profit function. Substituting equation (5) into equation (4) yields:

$$(8) \qquad \pi^*(v(p_q), p_l) = r(p_q s(p_q)) - p_l = 0$$

Note that this alternative representation of the indirect profit function (and thus the production function) is not the price space $(p_q, p_l)$. Instead it is in the 'hybrid' space characterized by $(v, p_l)$. The key advantage of using this 'hybrid' space is, of course, that the outcomes are observed by the econometrician, while the pure price and quantity spaces contain latent variables.

Differentiating this alternative characterization of the indirect profit function with respect to the price of housing, we obtain:

$$(9) \qquad \frac{\partial \pi^*(v(p_q), p_l)}{\partial p_q} = r'(p_q s(p_q))[s(p_q) + p_q s'(p_q)]$$

Moreover, in equilibrium, we must have

$$(10) \qquad \pi^*(v(p_q), p_l) \;\; = \;\; \pi(p_q, p_l)$$

From equations (3), (9), and (10), we then have the following key result that provides the basis of our approach to estimating $s(p_q)$:

PROPOSITION 2: *The supply function per unit of land is implicitly characterized by the solution to the following differential equation:*

$$(11) \qquad r'(p_q s(p_q)) \cdot [s(p_q) + p_q s'(p_q)] = s(p_q)$$

10

Proposition 2 summarizes an important methodological contribution of this paper. It shows that there exists a differential equation that characterizes the normalized housing supply function based on the equilibrium relationship between $p_l$ and $v$. Moreover, the differential equation only depends on the function $r(\cdot)$ which can be consistently estimated based on observed outcomes. Finally, a direct corollary of Proposition 2 is that the equilibrium locus $p_l = r(v)$ must satisfy the condition $0 < r'(v) < 1$ for all $v$.[19] Intuitively, $r'(v) > 0$ simply implies that, in equilibrium, the price per unit land rises as the value per unit land rises. If $r'(v)$ were greater than one, then a marginal increase in $v$ would be accompanied by an even larger increase in $p_l$, which would imply an increase in the price per unit land greater than the increase in the combined value of land and structures on the land.

Note that equation (11) can also be written as $r' = 1/(1 + \epsilon_s)$ were $\epsilon_s$ is the elasticity of the supply function per unit of land. If supply were infinitely elastic, we should see no relationship between between $p_l$ and $v$, whereas if it is inelastic the relationship would be one to one.

## C.   *Uniqueness*

We can show that a unique solution to the differential equation (11) exists. This solution expresses the supply relationship as an implicit function of $s$ and $p$. Depending on the form of $r(v)$, this solution may sometimes be expressed in closed-form with $s$ a function of $p$. To derive the general solution, rewrite equation (11) as:

(12) $$(r'(p\,s) - 1)\, s\, dp \;+\; r'(p\,s)\, p\, ds \;=\; 0$$

We have the following result:

---

[19]We test this restriction in our application.

PROPOSITION 3: *The integrating factor* $\mu(p, s) = p\,s$ *converts (12) into an exact differential equation.*[20] *As a consequence the solution to equation (12) is:*

$$\int M(p,s)dp + \int [N(p,s) - \frac{\partial \int M(p,s)dp}{\partial s}]ds = c$$

*or*

$$\int \frac{r'(ps)}{p}dp + \int \left[ \frac{r'(ps)}{s} - \frac{\partial \int \frac{r'(ps)}{p}dp}{\partial s} \right] ds = c + \ln(p)$$

Proof:

Dividing by the integrating factor, equation (12) can be written:

(13) $$\qquad\qquad\qquad\qquad M(p,s)dp + N(p,s)ds = 0$$

where

$$M(p,s) \;\; = \frac{r'(ps)-1}{p} \qquad N(p,s) = \frac{r'(ps)}{s}$$

Straightforward differentiation then establishes that the necessary and sufficient condition for (13) to be exact is satisfied, i.e. $\partial M/\partial s = \partial N/\partial p$. The second result follows from the first result by invoking the solution of an exact differential equation.

---

[20] *A differential equation of the form $M(x,y)dx + N(x,y)dy = 0$ is exact if and only if $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$. The general form of solution to exact differential equations is known and is employed in Proposition 3. Often, a differential equation that is not in exact form can be made exact by multiplying or dividing the equation by an integrating factor. Finding such a factor is not always easy, but if such a factor can be found as we have done for our application, then the general solution is available. See Section 2.6 of William Boyce and Richard DiPrima (2004) for a detailed discussion.*

## D.  Recovering the Production Function

Having derived the normalized housing supply function, it is straightforward to derive the housing

production function. Let

$$(14) \qquad\qquad m^*(p_q) = p_q s(p_q) - r(p_q s(p_q))$$

Points on the production function $q(m)$ are then given by $(m^*(p_q), s(p_q))$. Let the inverse of (14)

be $p_q^*(m)$. Then the production function for housing is equivalently written:

$$(15) \qquad\qquad q(m) = s(p_q^*(m))$$

## E.  The Polynomial Case

In empirical applications it is convenient to approximate the unknown $r(v)$ function with a poly-

nomial of order $k$:

$$(16) \qquad\qquad r_k(v) = \sum_{i=1}^{k} \frac{r_i}{i} v^i$$

Using polynomials has the advantage that we can easily characterize the normalized supply function

once an appropriate order of the polynomial has been determined.[21] We have the following general

result:

PROPOSITION 4: *Substituting equation (16) into (11) and normalizing such that $s(1) = 1$, the*

*implicit solution to the differential equation gives the supply function:*

$$(17) \qquad\qquad \sum_{i=2}^{k} \frac{r_i}{i-1} \left[ (ps)^{i-1} - 1 \right] + (r_1 - 1)\log(p) + r_1 \log(s) = 0$$

---

[21] The result in this section thus provides the theoretical justification for adopting a semi-nonparametric estima-
tion approach that does not require restrictive functional form assumptions to obtain a consistent estimator of the
production function.

*A closed-form expression for the supply function in the general polynomial case, expressed solely in*

*terms of $v$ and $\{r_i\}$, is*

$$
\begin{aligned}
(18) \qquad s &= \frac{v^{1-r_1}}{\exp\left\{\sum_{i=2}^{k} \frac{r_i}{i-1}(v^{i-1}-1)\right\}} \\
p &= v^{r_1}\exp\left\{\sum_{i=2}^{k} \frac{r_i}{i-1}(v^{i-1}-1)\right\}
\end{aligned}
$$

Proof:

Applying Proposition 2, we obtain

$$
(19) \qquad \sum_{i=2}^{k} \frac{1}{i-1}(ps)^{i-1}r_i + (r_1-1)\log(p) + r_1\log(s) = c
$$

Normalizing such that $s(1)=1$ implies $c = \sum_{i=2}^{k} \frac{r_i}{i-1}$ . Rearrange the equation above such that

$$
(20) \qquad \sum_{i=2}^{k} \frac{r_i}{i-1}\left[(ps)^{i-1}-1\right] + r_1\log(ps) = \log(p)
$$

We can solve for $s$ using the fact that $v = ps(p)$,

$$
(21) \qquad s = \frac{v}{p} = \frac{v^{1-r_1}}{\exp\left\{\sum_{i=2}^{k} \frac{1}{i-1}v^{i-1}r_i - c\right\}}
$$

## II.    Estimation

### A.    A Semi-nonparametric Approach

The theoretical results presented in the previous section directly translate into algorithms that can

be used to estimate the housing production function. The equation

$$
(22) \qquad p_l = r(v) \equiv r_k(v) = \sum_{i=1}^{k} \frac{r_i}{i}v^i
$$

14

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, $v$. These vary across location because of variation in the attributes of locations (accessibility to places of employment and other location-specific amenities). Households bid more for housing at locations with desirable attributes, leading to higher $v$ for more desirable locations. This in turn results in bidding up of the price per unit of land, $p_l$, at more desirable locations. The amount by which $p_l$ increases when $v$ increases is determined by the production function for housing.

To estimate equation (22) 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. In particular, let the measured price of land be $\tilde{p}_l$ with:

$$(23) \qquad\qquad \tilde{p}_l \;=\; p_l + \epsilon_p$$

This measurement model and equation (22) imply:

$$(24) \qquad\qquad \tilde{p}_l \;=\; r_k(v) \;+\; \epsilon_p$$

If we adopt the assumption that $\epsilon_p$ uncorrelated with $p_l$, this relationship can be estimated by OLS based on a sample of observations with size $N$. If we treat $k$ as a function of the sample size $N$, i.e. assume that $k = k(N)$, we can reinterpret the model above as a semi-nonparametric model. We can use standard econometric techniques to determined the number of expansion terms in the polynomial and thus approximate arbitrary functions with minimal functional form assumptions.[22]

---

[22]Xiaohong Chen (2006) provides an overview of semi-nonparametric estimation techniques.

The value of housing per unit of land may also be measured with error:

$$\tilde{v} \;\; = \;\; v \;\; + \;\; \epsilon_v \tag{25}$$

We can show in detail that $\epsilon_v$ may also be a consequence of productivity shocks. For example, 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.[23] Under either interpretation of $\epsilon_v$ we have:

$$\tilde{p}_l = r_k(\tilde{v} - \epsilon_v) + \epsilon_p \tag{26}$$

If we adopt the classical measurement error assumptions that $\epsilon_v$ and $\epsilon_p$ are mutually independent and independent of $v$ and $p_l$, we can estimate this relationship using methods developed by Jerry Hausmann, Hideo Ichimura, Whitney Newey, James Powell (1991) and Jerry Hausmann, Whitney Newey, James Powell (1995).[24] Implementing these estimation approaches requires instruments. Suitable instruments are locational attributes that affect the desirability of alternative locations.[25]

Variation in land prices across locations, and associated variation in property value per unit land, arise from differences in proximity to places of employment and commerce, as well as variation across locations in access and availability of public goods and amenities. Hence, systematic differences in such variables can be exploited when instrumenting for idiosyncratic measurement error in values of individual properties. For example, distances of properties from the city center are a systematic source of variation in property values that can be used to instrument for idiosyncratic errors in

---

[23]An appendix that provides a more detailed discussion is available from the authors.

[24]See also Whitney Newey (2001) and Susanne Schennach (2007)

[25]Semi-nonparametric IV techniques are also discussed in Whitney Newey and James Powell (2003), Chunrong Ai and Xiaohong Chen (2003) or Richard Blundell, Xiaohong Chen, and Dennis Kristensen (2004).

valuation of individual properties. Fixed effects for municipalities, reflecting systematic differences across municipalities in quality of public goods and amenities, can also be used to instrument for idiosyncratic errors in valuation of individual properties.[26]

One drawback of the approach outlined above is that polynomials do not form an orthonormal basis for the class of functions in which we are interested. Hence, they are not optimal from a purely econometric perspective. It is straightforward to extend our approach to use different types of series estimators. Alternatively, we can estimate $r(v)$ using a fully nonparametric estimator, such as a kernel estimator as along as errors conform to the model in equation (24).[27]

An additional problem encountered in estimation is that the estimated function must satisfy the condition that $0 < r'(v) < 1$ for all $v > 0$. One advantage of semi-nonparametric approaches discussed above is that it is relatively straightforward to invoke these types of shape restrictions.[28] If we use fully nonparametric techniques we obtain an unrestricted kernel estimate of $r(v)$. We then need to check whether the derivative restrictions are satisfied everywhere.[29] With the bandwidth set equal to the standard deviation of $v$, we find in our application that the derivative conditions are met.[30]

---

[26]A valid estimator can also be constructed if researchers have access to multiple measurements of housing values. Assessors sometime provide independent valuation of houses which may not coincide with sales prices.

[27]A review of the literature in applied non-parametric regression analysis is given by Wolfgang Härdle and Oliver Linton (1994).

[28]Arie Beresteanu (2005) and Xiaohong Chen (2006) discuss techniques for semi-nonparametric estimation under general shape restrictions.

[29]Without the upper bound restriction, the restricted estimation problem is equivalent to nonparametric monotone regression as developed by Haw Mukerjee (1988) and Enno Mammen (1991), who combine isotonic regression with a nonparametric kernel. See also the review by Rosa Matzkin (1994).

[30]We also explored other reasonable values for the bandwidth and obtained similar results.

Before we proceed with the empirical analysis, it is useful to compare our approach to identification and estimation of the production function to other approaches in the modern IO literature. There are many other empirical applications in which researchers do not have access to reliable price and quantity measures of output. Instead researchers rely on value-based output measures. If prices differ for the same good, the value of output is not necessarily a good measure for the quantity of output. Ignoring the heterogeneity in output prices will generally lead to inconsistent estimators of the production function as noted by Jacob Marschak and William Andrews (1944). It is the lack of observability of output quantities and prices that distinguishes our approach from most previous studies.[31]

The 'omitted price bias' is distinctly different than the standard 'transmission bias' that arises to due to endogeneity of input factors.[32] Our approach differs from most previous approaches since we focus on the dual problem. The 'transmission bias' arises in our context if the errors capture productivity shocks in addition to measurement error. The use of instruments in estimation largely deals with this endogeneity problem. Our identifying assumption is then that there are no spatial patterns in productivity shocks within the metropolitan area. For example, productivity shocks must be uncorrelated with proximity to the city center.

The recent literature has also focused on dynamic selection problems. Endogeneity arises due to entry and exit in a dynamic oligopoly model. Our approach is static and thus ignores the selection

---

[31]This point is also made in Tor Klette and Zvi Griliches (1996) who directly estimate the production function relying on an auxiliary pricing model to correct for the omitted price bias. See also Hajime Katayama, Shihua Lu and James Tybout (2003).

[32]This problem is also discussed in Yair Mundlak and Irving Hoch (1965) and Arnold Zellner, Jan Kmenta, and Jacques Dreze (1966).

problems discussed in (1996).[33] We view it as interesting future research to investigate whether the methods developed in this paper can be extended to account for selection due to entry and exit.

Yair Mundlak (1996) studies the efficiency of dual and primal estimators. He demonstrates that "...estimates based on duality, unlike direct estimators of the production function, do not utilize all the available information and therefore are statistically inefficient and the loss in efficiency may be sizable." (p. 431) He concludes that "...the most efficient way of estimating product supply or factor demand is to derive them from the empirical production function, rather than the reverse." (p. 437) Mundlak's formulation assumes that the econometrician has access to measures of the price and the quantity of output as well as the quantities of inputs. We appeal to duality to overcome the problem that arises when the value of output is observed, but output price and quantity are not observed.

It is of interest to note the relationship of our approach to the standard approach based on duality theory.[34] Econometric analyses using duality generally rely on availability of price data for outputs and inputs (e.g., Christensen, Jorgensen, and Lau, 1973), whereas our approach is specifically designed for situations in which output price data are not available. A more restrictive but complementary approach is to utilize functional forms that can be estimated without price data for output. Perhaps the simplest case is the Cobb-Douglas production function. The indirect profit function for the Cobb-Douglas can be estimated without output prices. Also, the first order conditions can be written as factor shares—which can be estimated without use of output price data. Similarly, the first-order conditions implied by the single-output translog indirect profit

---

[33]See also James Levinsohn and Amil Petrin (2003), Mark Melitz (2003), and Dan Ackerberg, Kevin Caves, and Garth Frazer (2005).

[34]For an excellent, comprehensive treatment of the dual approach to analysis of production, see Melvin Fuss and Daniel McFadden (1978).

function can be estimated without output price data.[35] Below we present results for these cases as part of our empirical analysis.

Finally, we would like to point out the approach to identification and estimation of the production function is general and can be adopted to study other industries.[36].

## III.  Data

Our empirical application focuses on new housing construction in Allegheny County in Pennsylvania, which contains the greater metropolitan area of Pittsburgh.[37] Property assessment falls into the domain of the Allegheny county government. The vast majority of municipalities in Allegheny County have used a standard property tax system which taxes structures and land at the same rate. In contrast, the City of Pittsburgh (and three other communities in Allegheny County) used a two-rate property taxation from 1913-2001. A key feature of this system was that the tax on land value was higher and the tax on improvement value is lower. Thus, there were strong incentives to value both land and properties.[38] After years of under-assessment, Common Pleas Judge R. Stanton Wettick ordered that Allegheny county conducted a complete reassessment of all properties. Allegheny county hired Sabre Systems of Ohio to prepare the 2001 assessment roll. The total cost of the revaluation was $30 million.

Sabre Systems created approximately 1,800 different neighborhoods within Allegheny County on which the assessment was based. It used computer assisted methods of assessment which employ

---

[35]The translog indirect profit function cannot be estimated without price data for output. However, estimation of the first-order conditions from the single-output translog indirect profit function does not require output prices. This useful feature of the translog appears not to have been fully appreciated.

[36]An appendix that provides an illustration of our techniques for the car service repair industry is available from the authors.

[37]Another application that analyzes construction in Wake County, North Carolina, is available upon request from the authors.

[38]Pittsburgh's tax on land was about 5.77 times the tax on improvements.

a variety of techniques. Sabre used the comparable sales method to value land which entails analysis of vacant land sales within the neighborhood to establish base land rates. Where there are sufficient quantities of vacant land sales, the comparable sales method is a simple correlation of unit comparison, such as square foot divided by sale price. However, the absence of adequate vacant land sales in a given neighborhood required Sabre to use other methods, primarily the land residual option. In order to arrive at an estimated land value, Sabre subtracted depreciated building value estimates from sale prices. As detailed in the documentation provided by Sabre Systems (2001), these valuations were based on extensive information about the size, characteristics, age, and condition of each dwelling. Finally, the individual base lot value was often adjusted by appraisers, in the review and appeals process, for numerous possible influences to produce the final land value estimate.[39]

There has not been a subsequent major reassessment in Allegheny county.[40] As a consequence, the property appraisals are currently used to determine the tax base for school districts and municipalities in the county. In July of 2004, the Pennsylvania State Tax Equalization Board performed a study of all properties that had been sold to determine how close the assessed values were to sale prices for properties that had sold recently. They concluded that, on average, the assessments were within 2.5 percent of the sale price. We chose the assessed values since they give us a measure of housing value for all properties.

Our data set is based on the appraisals conducted by Sabre Systems and uses land value appraisals as well as property appraisals. The data come from the Allegheny County web site, which is maintained by the Office of Property Assessments.[41] The web site provides access to a database

---

[39]Details are available in Sabre System (2001).

[40]The county government has been appealing the requirement of annual reassessment, and, in the interim, has used the results of the 2001 assessment process.

[41]The web site is http://www2.county.allegheny.pa.us/RealEstate/

with detailed information about all properties, both residential and commercial, in the entire county.

The complete database lists 561,174 properties. After eliminating all non-residential properties and those that are listed as condemned or abandoned, we are left with 423,556 observations. We successfully geocoded – matched to longitude and latitude coordinates – 370,178 of these properties. We used the coordinates to assign each property to its corresponding travel zone, and retrieved the travel times to the designated city center traffic zone (for use as an instrument.) Eliminating properties that did not have positive lot area sizes and market values listed and those that we were unable to match with travel time data reduces the sample to 358,677 observations. We implement our estimation procedure using the subsample of housing units that were built after 1995. Despite the fact that there has been little population growth in the Pittsburgh metropolitan area during the past decades, we observe a large amount of new housing construction in that time period. There are 6,362 houses that have been built since 1995 in our sample. The upper panel of Table 1 provides descriptive statistics of our data set for residences.

**Table 1**. Descriptive Statistics

| Sample of Residential Real Estate | | | | | |
|---|---|---|---|---|---|
| Variable | Mean | Median | Stdev | Min | Max |
| value per unit of land, $v$ | 21.44 | 14.29 | 26.91 | 0.15 | 366.62 |
| price of land, $p_l$ | 3.32 | 2.28 | 3.86 | 0.05 | 41.75 |
| lot area (sq ft) | 26756 | 15507 | 52197 | 540 | 1207483 |
| travel time (minutes) | 29.12 | 30 | 9.47 | 1 | 59 |
| Sample of Commercial Real Estate | | | | | |
| Variable | Mean | Median | Stdev | Min | Max |
| value per unit of land, $v$ | 41.55 | 10.56 | 116.870 | 0.0687 | 1807.62 |
| price of land, $p_l$ | 6.76 | 2.72 | 9.939 | 0.0108 | 68.20 |
| lot area (sq ft) | 139393.84 | 39437 | 481327.12 | 10038 | 6827594 |

The size of the residential (commercial) sample is 6,362 (992).

We observe a large amount of variation in land prices within the Pittsburgh metropolitan area. The 5th and 95th percentiles of land prices differ by a factor of five; the 1st and 99th percentiles

by a factor of fifty. Figure 1 provides a nonparametric estimate of the density of land prices. The variation in land prices induces variation in the relative proportions of land and non-land factors used in housing production. Broadly speaking, housing developers will use different development strategies depending on the price of land. Housing developers will build a lot of structures on small land areas if land is expensive and vice versa. The value of a house per unit of land is a measure of how land-intensive the production process is in equilibrium. Figure 2 illustrates a nonparametric estimate of the density of the value of a house per unit of land.

We also construct a second sample that consists of commercial properties located in downtown Pittsburgh which corresponds with the central business direct and contains most high-rise office buildings. We restricted out sample to commercial properties with a lot area of at least 10,000 sq feet. This left us with 992 observations. The lower part of Table 1 provides some descriptive statistics about this sample. We do not have year built for these properties. So we are unable to restrict the sample to new structures. Hence we primarily focus on the residential estimates.

## IV.    Empirical Results

We estimate the function $r(v)$ which relates land price and home value per unit land. Table 2 summarizes the results using OLS for log-linear, linear, quadratic, and cubic models. We also tested higher-order polynomials, and while additional terms were significant, they were not quantitatively important. All p-values were calculated using heteroskedasticity-robust standard errors. With the exception of the log-linear form, there are no constant terms in the equations we estimate because $p_l$ must go to zero as $v$ goes to zero.[42]

---

[42]We find that the constant is typically insignificant or rather small in all unconstrained regressions which is broadly consistent with our approach.

**Table 2**. Estimates of Equilibrium Locus

| | OLS Estimates | | | |
|---|---|---|---|---|
| | Log-linear | Linear | Quadratic | Cubic |
| $v$ | | 0.1394*** | 0.1685*** | 0.1622*** |
| $v^2$ | | | -0.0002*** | -0.0001 |
| $v^3$ | | | | 0.00000039* |
| Constant | -1.6051*** | | | |
| $\log(v)$ | 0.9090*** | | | |
| $N$ | 6,362 | 6,362 | 6,362 | 6362 |

| | HNIP Estimates | | | |
|---|---|---|---|---|
| | Log-linear | Linear | Quadratic | Cubic |
| $v$ | | 0.1440 *** | 0.1631*** | 0.1732*** |
| $v^2$ | | | -0.0002*** | -0.0005*** |
| $v^3$ | | | | 0.000004** |
| Constant | -1.6129*** | | | |
| $\log(v)$ | 0.9119*** | | | |
| $N$ | 6,362 | 6,362 | 6362 | 6,362 |

Notation: * indicates significance at the 90% level, ** at the 95% level, and *** at the 99% level.

The table reports estimates of the equilibrium locus which characterizes the relationship between the price of land and the value of housing per unit of land.

In general, we find that all our models fit the main features of the residential data reasonably well. We also performed a variety of robustness checks to validate our empirical approach. One may be concerned that our results might be sensitive to extreme values of $v$. To test this hypothesis, we re-estimated all models of $r(v)$ excluding the smallest one percent of observations ($v < 0.9282$), the largest one percent of observations ($v > 65.9924$), and both. We find that the results are robust to the exclusion of extreme values of $v$. Also, heteroskedasticity-weighted regression results were similar to those obtained with OLS. Finally, we find that the monotonicity condition of $r(v)$ is satisfied under all the polynomial estimation cases for the range of values of $v$ observed in the data

Next we estimate our models using the instrumental variable estimator suggested by Hausmann et al. (1991). We choose commuting time to the city center as an instrument, since it is natural to expect that property values tend to decline as commuting time rises. We use travel time data from the Southwestern Pennsylvania Commission (SPC) for Allegheny County. The SPC divided the county into 995 traffic zones of varying size, roughly distributed according to traffic and population density. The city of Pittsburgh is covered by 465 zones. The SPC provided us with estimated travel times from each zone to another, under both congested and uncongested conditions. We also include as instruments fixed effects for each municipality in the metropolitan area and for the 32 wards in the city of Pittsburgh. These dummy variables serve to capture locational amenities that can be expected to vary widely given the topography of the Pittsburgh area. The HNIP estimation results can be found in the bottom panel of Table 2. While results for the log-linear cases are quite similar to the OLS results, the estimates are slightly different in the cubic case, with the coefficient on the quadratic term now significant. As we discuss below, these differences have, however, little impact on the estimated supply and production functions.

As we noted earlier, the translog indirect profit function yields first-order conditions that can be

estimated without housing price data. In our housing application, the first-order condition for the translog indirect profit function is: $p_l/v = \alpha + \beta \ln(p_l)$.[43] Instrumental variable estimates for this equation yield coefficients (est. standard deviation) $\alpha = .165$ (.015) and $\beta = .0022$ (.0015). The estimate of coefficient $\beta$ is quantitatively small relative to the magnitude of $\ln p_l$ (which has mean .8 and standard deviation .9). In addition, the estimate of $\beta$ is statistically insignificant. Thus, the translog estimates do not reject the null hypothesis that the factor share of land $(p_l/v)$ is constant (i.e., Cobb-Douglas), which is in accord with our more general analysis above.[44]

Given an estimate of $r(v)$, we can estimate the supply function per unit land. We find that the supply functions for parametric, semi-nonparametric, and nonparametric estimates of $r(v)$ are relatively similar. Figure 3 plots the supply function for the log-linear case as well as 95 % confidence bands. The plots suggest that the supply function per unit of land is relatively price elastic. Since the specifications estimated in this paper typically do not yield constant price elasticities, we compute the elasticity for each observation in the sample and average across observations. We find that the average price elasticity ranges from 4.31 in the cubic case to 6.6 in the fully nonparametric case.

After obtaining $r(v)$ and $s(p)$, we can estimate the production function $q(m)$. Consider, for example, the Cobb-Douglas case in which $r(v) = kv$. The estimated slope coefficient is 0.144. This implies that the Cobb-Douglas production function is given by $Q(L, M) = 1.38 * L^{0.144} * M^{0.856}$. As before, we find that the different econometric specifications of the $r(v)$ function yields similarly shaped production functions. Figure 4 plots the production function and 95 % confidence bands that corresponds to the log-linear case. One important feature of the production function is the elasticity

---

[43]Note that all results in Table 2 are for the normalized indirect profit function, while the translog estimates are for first order conditions determining land input per unit housing.

[44]The OLS estimate of $\beta$ is .0044 (.007); statistically significant, but, again, quantitatively small.

of substitution between land and non-land factors. As with supply elasticities, the specifications estimated in this paper typically do not yield constant elasticities of substitution. Hence we compute weighted averages of the elasticities of substitution based on the sample frequencies. We find that the elasticity of substitution between land and non-land factors ranges between 1 in the linear case to 1.16 in the log-linear case.[45]

Finally, we applied our approach using the data for commercial properties in the central business district. Not surprisingly, we find that the estimates are substantially different from the residential property case. Consider the log-linear case. We estimate a constant term of -0.7230 (0.0398) and an intercept of 0.7440 (0.0152). The mean supply elasticity for commercial property is 3.9854 (1.4320), and the mean substitution elasticity is 1.39 (0.04).

## V.    Conclusions

We have discussed in this paper how to estimate the production function for housing. The main problem encountered in estimation is that prices and quantities for housing are never observed separately. We have developed a new approach, treating prices and quantities as latent variables, that allows us to identify and estimate the underlying production function without relying on strong functional form assumptions. The main insight behind our approach is that the observed variation in land prices and housing values per unit of land is sufficient to identify the housing supply function per unit of land if the production function exhibits constant returns to scale. Given the supply function per unit of land it is straight forward to recover the underlying production function. We have considered an application that is based on data from Allegheny County in Pennsylvania. We

---

[45]These findings are broadly consistent with early empirical studies on housing supply. John McDonald (1981) surveys 13 studies and report estimates of the elasticity of substitution ranging between 0.36 and 1.13 with a majority obtaining estimates significantly less than one.

have seen that the approach suggested in this paper yields plausible estimates for the price elasticity of the housing supply per unit of land and the elasticity of substitution between land and non-land factors.

Our research has some important implications for policy analysis. The production function for housing plays an important role in conducting applied general equilibrium policy analysis. Many urban policies – such as school voucher programs, property tax reforms, housing vouchers, welfare reform, urban development policies, or policies aimed at improving access of poor and minority households to economic opportunity – are likely to affect the demand for housing and residential sorting patterns. If the supply of new housing is price elastic, an increase in the demand for housing can be met by an increase in housing supply as rising land prices induce producers to increase the amount of structure per unit of land. Hence, even large policy changes may only have a small impact on housing prices if the supply is elastic. As such welfare implications, will largely be driven by household adjustments and changes in housing quantities, and not so much by price changes.[46]

---

[46]In contrast to Glaeser and Gyourko (2005), we do not focus on total housing supply. Instead we only consider the supply of new housing. Our results for new housing supply is broadly consistent with the evidence reported in Glaeser and Gyourko.

# References

Ackerberg, Daniel, Caves, Kevin, and Frazer, Garth. 2005. Structural Estimation of Production Functions. Working Paper.

Ai, Chunrong and Chen, Xiaohong. 2003. Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, *71*, 1795–1843.

Alonso, William. 1964. *Location and Land Use: Towards a General Theory of Land Rent*. Cambridge University Press. Cambridge.

Basu, Susanto and Fernald, John. 1997. Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy*, *105*, 249–83.

Bayer, Patrick., Ferreira, Fernando., and McMillan, Robert. 2007. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115 (4)*, 588–638.

Beresteanu, Arie. 2005. Nonparametric Estimation of Regression Functions Under Restrictions on Partial Derivatives. *Rand Journal of Economics*, *36 (4)*, 870–89.

Blundell, Richard, Chen, Xiaohong, and Kristensen, Dennis. 2004. Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. Working Paper.

Boyce, William and DiPrima, Richard. 2004. *Elementary Differential Equations*. Wiley.

Brueckner, Jan and Rosenthal, Stuart. 2005. Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich?. Working Paper.

Burmeister, Edwin and Kuga, Kiyoshi. 1970. The Factor-Price Frontier Duality, and Joint Production. *The Review of Economic Studies*, *37*, 11–19.

Chen, Xiaohong. 2006. Large Sample Sieve Estimation of Semi-Nonparametric Models. In *Handbook of Econometrics 6*. North Holland.

Christensen, Lawrits, Jorgenson, Dale, and Lau, Lawrence. 1973. Transcendental logarithmic production frontiers. *Review of Economics and Statistics*, *55*, 28–45.

De Leeuw, Frank and Ekanem, Nkanta. 1971. The Supply of Rental Housing. *American Economic Review*, *61 (5)*, 806–817.

Diewert, William. 1971. An Application of Shepart's Duality Theorem: A Generalized Leontief Production Function. *Journal of Political Economy*, *79*, 461–507.

Diewert, William. 1973. Functional Forms for Profit and Transformation Functions. *Journal of Economic Theory*, *6 (3)*, 284–316.

Epple, Dennis, Romer, Thomas, and Sieg, Holger. 2001. Interjurisdictional Sorting and Majority Rule: An Empirical Analysis. *Econometrica*, *69*, 1437–1465.

Epple, Dennis and Sieg, Holger. 1999. Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy*, *107 (4)*, 645–681.

Ferreyra, Maria. 2007. Estimating the Effects of Private School Vouchers in Multi-District Economies. *American Economic Review*, *97*, 789–817.

Fuss, Melvin and McFadden, Daniel (eds). 1978. *Production Economics: A Dual Approach to Theory and Applications.* North Holland. Amsterdam.

Gallant, Ronald. 1981. On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form. *Journal of Econometrics*, *15*, 211–245.

Glaeser, Edward and Gyourko, Joseph. 2005. Urban Decline and Durable Housing. *Journal of Political Economy*, *113*, 345–75.

Glaeser, Edward., Gyourko, Joseph, and Saks, Ravin. 2005. Urban Growth and Housing Supply. Working Paper.

Härdle, Wolfgang and Linton, Oliver. 1994. Applied Nonparametric Methods. In *Handbook of Econometrics IV*. Elsevier North Holland.

Haughwout, Andrew, Orr, James, and Bedoll, David. 2008. The Land Prices in the New York Metropolitan Area. *Federal Reserve Bank of New York Economic Policy Review*, *14 (3)*, 1–7.

Hausmann, Jerry, Newey, Whitney, Ichimura, Hideo, and Powell, James. 1991. Identification and Estimation of Polynomial Errors in Variables Models. *Journal of Econometrics*, *61*, 273–295.

Hausmann, Jerry, Newey, Whitney, and Powell, James. 1995. Nonlinear Errors in Variables Estimation of Some Engle Curves. *Journal of Econometrics*, *65*, 205–233.

Jorgenson, Dale and Lau, Lawrence 1974. The Duality of Technology and Economic Behavior. *Review of Economic Studies*, *41 (2)*, 181–200.

Katayama, Hajime, Lu, Shihua, and Tybout, James. 2003. Why Plant-Level Productivity Studies are of Misleading, and an Alternative Approach to Inferences. NBER Working Paper 9617.

Klette, Tor and Griliches, Zvi. 1996. The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous. *Journal of Applied Economics*, *11*, 343–361.

Levinsohn, James and Petrin, Amil. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, *70*, 317–42.

Lucas, Robert and Rossi-Hansberg, Esteban. 2002. On the Internal Structure of Cities. *Econometrica*, *70 (4)*, 1445–76.

Mammen, Enno. 1991. Nonparametric Regression under Qualitative Smoothness Assumptions. *Annals of Statistics*, *19*, 741–759.

Marschak, Jacob and Andrews, William. 1944. Random Simultaneous Equations and the Theory of Production. *Econometrica*, *12*, 143–205.

Matzkin, Rosa. 1994. Restrictions of Economic Theory in Nonparametric Methods. In *Handbook of Econometrics IV*. Elsevier North Holland.

McDonald, John. 1981. Capital-land Substitution in Urban Housing: A Survey of Empirical Estimates. *Journal of Urban Economics*, *9*, 190–211.

Melitz, Mark. 2003. Estimating Firm-level Productivity in Differentiated Product Industries. Working Paper.

Mills, Edwin. 1967. An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review*, *57*, 197–210.

Mukerjee, Haw. 1988. Monotone Nonparametric Regressions. *Annals of Statistics*, *16*, 741–750.

Mundlak, Yair. and Hoch, Irving. 1965. Consequences of Alternative Specification of Cobb-Douglas Production Functions. *Econometrica*, *33*, 814–28.

Mundlak, Yair. 1996. Production Function Estimation: Reviving the Primal. *Econometrica*, *64 (2)*, 431–38.

Muth, Richard. 1960. The Demand for Non-Farm Housing. In Harberger, A. (Ed.), *The Demand for Durable Goods*. University of Chicago Press.

Muth, Richard. 1969. *Cities and Housing*. University of Chicago Press. Chicago.

Newey, Whitney. 2001. Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models. *Review of Economics and Statistics*, *83 (4)*, 616–27.

Newey, Whitney and Powell, James. 2003. Instrumental Variables Estimation for Nonparametric Models. *Econometrica*, *71*, 1565–78.

Olley, Steven and Pakes, Ariel. 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*, 1263–98.

Olson, Edward. 1969. A Competitive Theory of the Housing Market. *American Economic Review*, *59*, 612–619.

Sabre Systems. 2001. Overview of Land Appraisal Methodology. Discussion Paper.

Samuelson, Paul. 1953. Prices of Factors and Goods in General Equilibrium. *The Review of Economic Studies*, *21*, 1–20.

Schennach, Susanne. 2007. Estimation of Nonlinear Errors-in-Variables Models. *Econometrica*, *75*, 201–39.

Shepherd, Ronald. 1970. *The Theory of Cost and Production Functions*. Princeton University Press. Princeton.

Sieg, Holger, Smith, Kerry., Banzhaf, Spencer., and Walsh, Randall. 2004. Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods. *International Economic Review*, *45 (4)*, 1047–1077.

Smith, Barton. 1976. The Supply of Urban Housing. *Quarterly Journal of Economics*, *90 (3)*, 389–405.

Smith, Lawrence, Rosen, Kenneth, and Fallis, George. 1988. Recent Developments in Economic Models of Housing Markets. *Journal of Economic Literature*, *26*, 29–64.

Vinod, Hrishikesh and Ullah, Annan. 1988. Flexible Production Function Estimation by Nonparametric Kernel Estimators. In *Advances in Econometrics*. JAI Press.

Zellner, Arnold, Kmenta, Jan, and Dreze, Jacques. 1966. Specification and Estimation of Cobb-Douglas Production Functions. *Econometrica*, *34*, 784–95.

**Figure 1**. Density of Land Prices
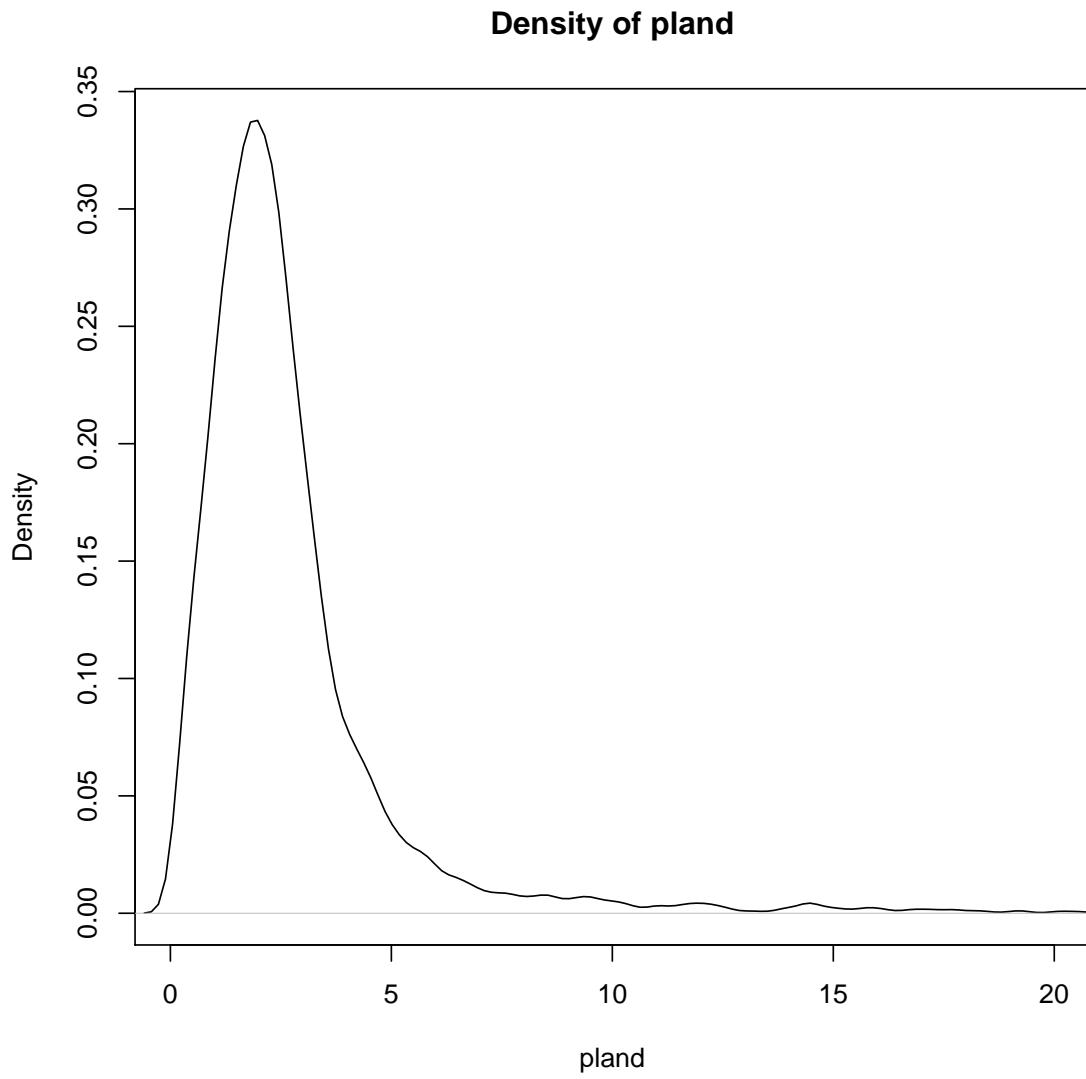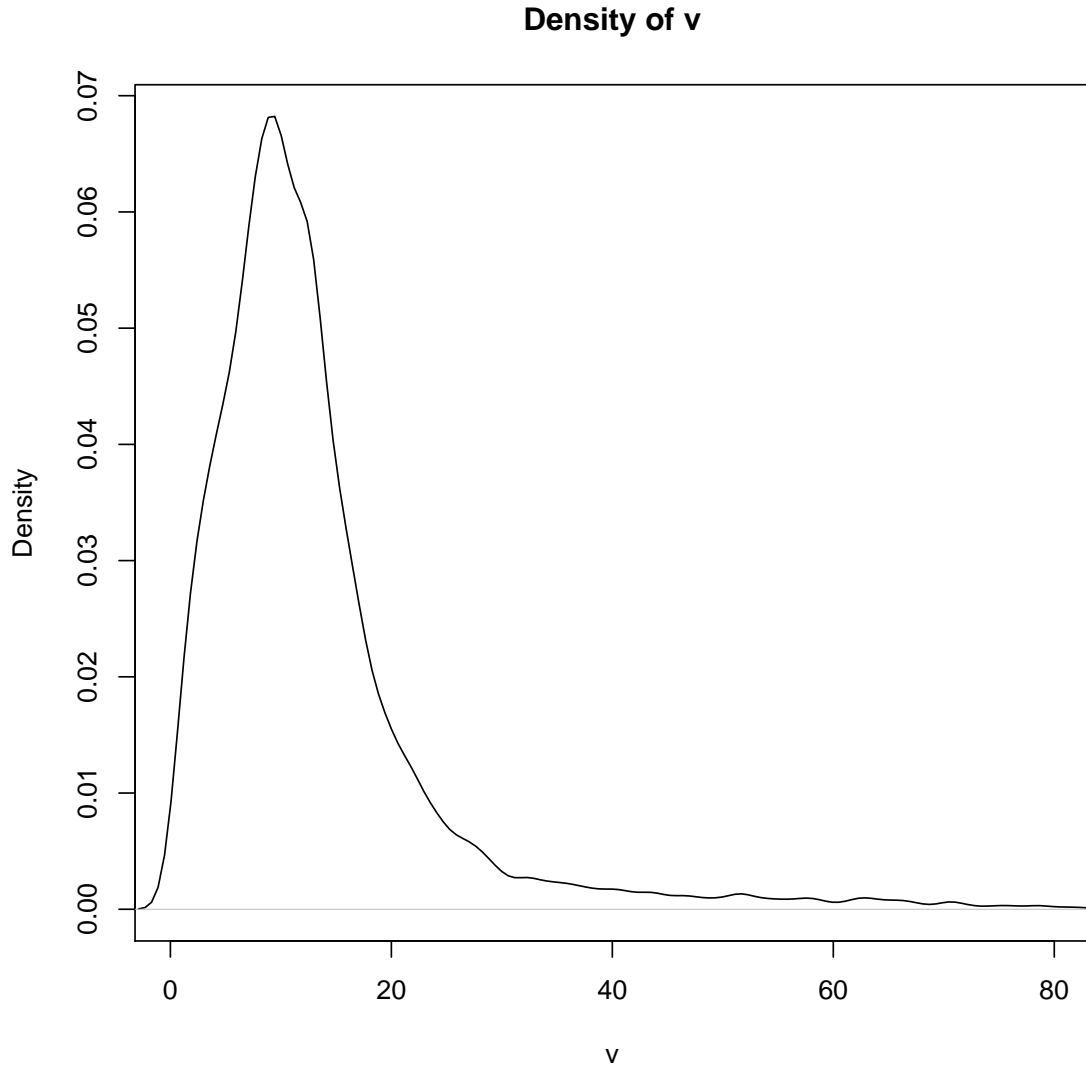
## Density of pland

## Density of v

Figure 3.



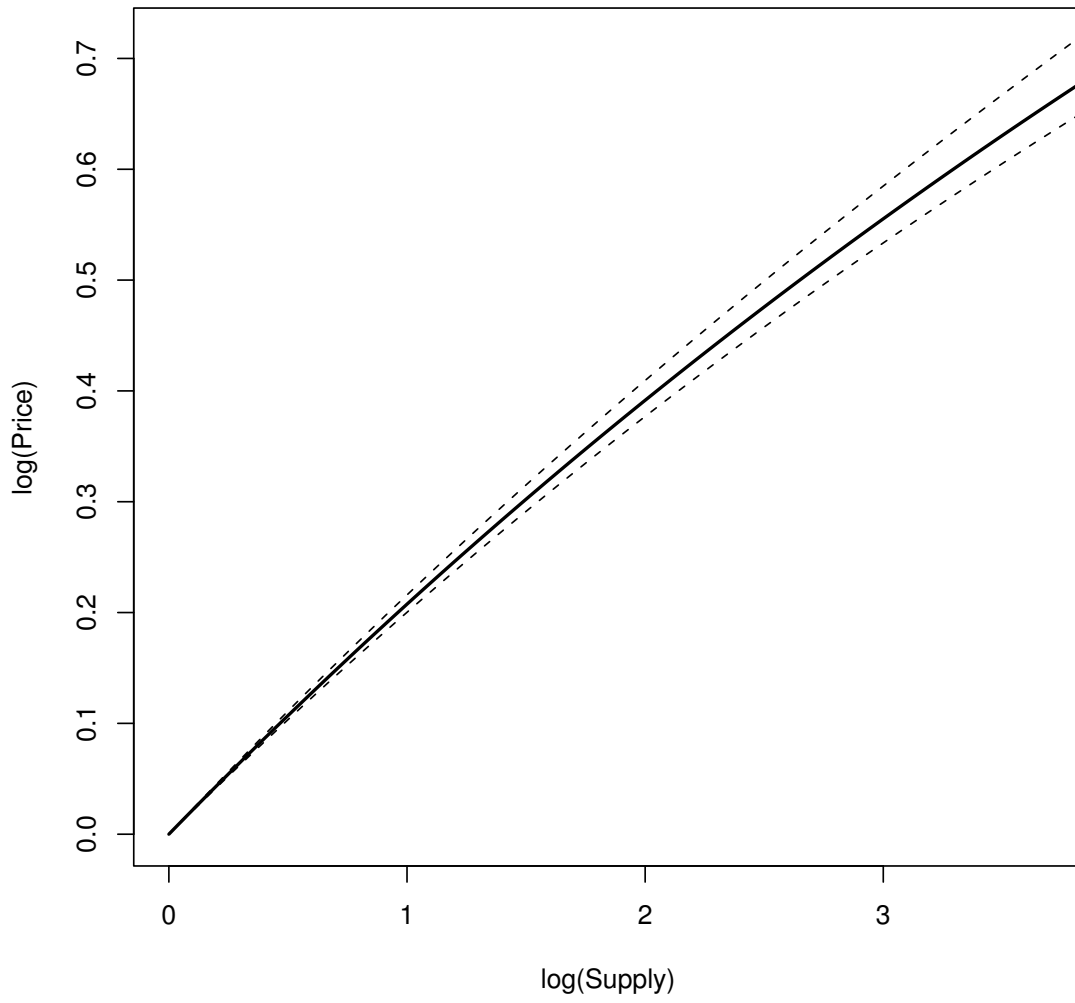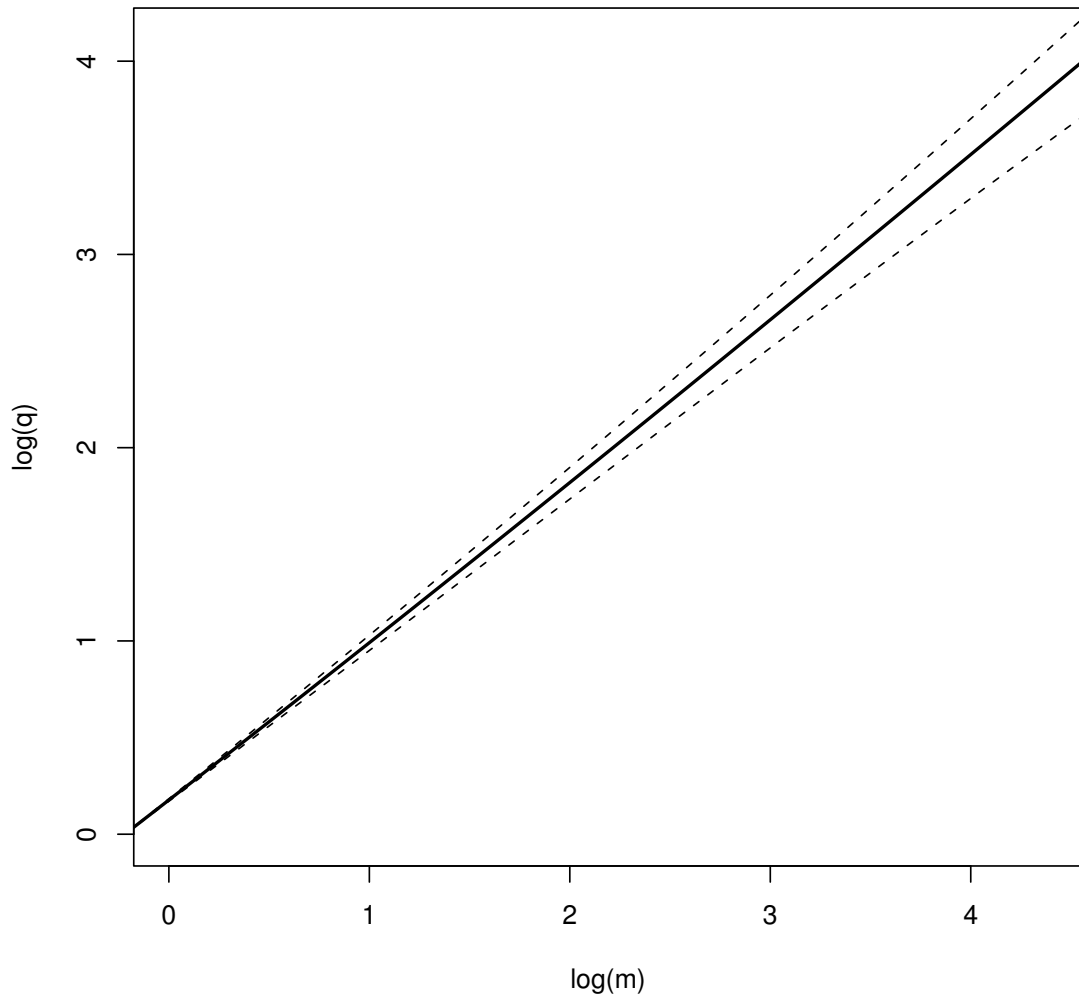**Log−linear Supply Function with
95% Confidence Band**

Figure 4.

## Log–linear Production Function with
## 95% Confidence Band

# 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:

$$
\begin{aligned}
(1) \qquad m(p_q) &= (\alpha p_q)^{\frac{1}{1-\alpha}} \\
s(p_q) &= (\alpha p_q)^{\frac{\alpha}{1-\alpha}}
\end{aligned}
$$

as a consequence we have:

$$
\begin{aligned}
(2) \qquad v(p_q) &= p_q \, s(p_q) \\
&= \alpha^{\frac{\alpha}{1-\alpha}} \, p_q^{\frac{1}{1-\alpha}}
\end{aligned}
$$

Inverting this function yields

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

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

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

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:

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

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

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

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

$$(8) \qquad \begin{aligned} \frac{\partial \pi}{\partial p_q} &= s(p_q, p_n) \\ \frac{\partial \pi}{\partial p_n} &= -n(p_q, p_n) \end{aligned}$$

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

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

---

[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 $p_n$:

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

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:

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

Taking partial derivatives, we have:

$$
(12) \qquad\qquad
\begin{aligned}
\frac{\partial \pi^*}{\partial p_q} &= \frac{\partial r}{\partial v}\frac{\partial v}{\partial p_q} \\[2ex]
&= \frac{\partial r}{\partial v}\left[ s(p_q, p_n) + p_q\frac{\partial s(p_q, p_n)}{\partial p_q}\right]
\end{aligned}
$$

and

$$
(13) \qquad\qquad
\begin{aligned}
\frac{\partial \pi^*}{\partial p_n} &= \frac{\partial r}{\partial v}\frac{\partial v}{\partial p_n} + \frac{\partial r}{\partial p_n} \\[2ex]
&= \frac{\partial r}{\partial v}p_q\frac{\partial s(p_q, p_n)}{\partial p_n} + \frac{\partial r}{\partial p_n}
\end{aligned}
$$

Note that in equilibrium

$$
(14) \qquad\qquad
\begin{aligned}
\pi(p_q, p_n) &\equiv \pi^*(p_q, p_n) \\[2ex]
n(p_q, p_n) &\equiv n^*(p_q s(p_q, p_n), p_n)
\end{aligned}
$$

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)
$$p_l = r(v)$$

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:

$$\tilde{p}_l = p_l + \epsilon_{p_l}$$
$$\tilde{v} = v + \epsilon_v$$

The $\epsilon_v$ 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)
$$\tilde{p}_q = p_q + \epsilon_p$$

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

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

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

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

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:

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

where $\epsilon_v = q \cdot \epsilon_p$

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

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

If we assume that $\epsilon_v$ and $\epsilon_p$ 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.