A Flexible Approach to Estimating Production Functions When Output Prices are Unobserved*

Dennis Epple
Carnegie Mellon University and NBER

Brett Gordon
Carnegie Mellon University

Holger Sieg
Carnegie Mellon University and NBER

July 12, 2006

*We would like to thank Andrew Chesher, Lanier Benkard, Ron Goettler, Francois Ortalo-Magne, Costas Meghir, Marc Mündler, Amil Petrin, Steven Ross, Vishal Singh and participants at workshops at Carnegie Mellon University, Humboldt University in Berlin, Koc University, University College London, the University of Munich, the SED meeting in Vancouver, and the University of Wisconsin for comments and suggestions. Special thanks to Bob Roberts from Babcox for providing us some of the data used in this paper. Financial support for this research is provided by the NSF.

Abstract

There exists significant price dispersion in many industries. As a consequence the value of output is not necessarily a good measure of the quantity of output. Estimation of production functions for these types of goods is thus challenging since quantities and prices of output are typically not separately observed by the econometrician. This paper provides a new flexible approach for estimating production functions which treats quantity and prices for output as latent variables. To illustrate the usefulness of the techniques we consider two applications. The first application focuses on estimating production functions for housing. The empirical analysis is based on a comprehensive database of recently built properties in Allegheny County, Pennsylvania. The second application focuses on the car repair service industry based on a unique survey conducted by the leading magazine that covers this industry. We find that the new methods proposed in this paper work well in these applications and provide reasonable estimates for the underlying production functions.

JEL classification: C51, L11, R12

1 Introduction

There exists significant price dispersion in many industries in the U.S. economy.[1] For example, a house in Beverly Hills or Santa Monica can easily costs five times as much as a comparable house in Riverside or San Bernadino county.[2] The cheapest room in a Hilton Hotel in New York City starts at $439. A similar room in a Hilton Hotel in Cleveland, Ohio, costs $109. Renting a Ford Taurus from Hertz at LAX costs $74 a day. Renting the same car from Hertz at the Eureka Airport costs $48 a day.[3] Tickets for "The Producers" range between $31.25 and $51.25 in Tallahassee and $36.25 to $111.25 in New York. Prices for brake repairs range between $110 and $600 in California. An oil and lube job costs between $25 and $100 in Florida. Prices for routine basic health care services, hair cuts, commercial real estate, and many other goods vary substantially across locations in the U.S.[4]

Differences in output prices are often due to differences in local factor prices.[5] For example, the 5th and 95th percentiles of land prices differ in the Pittsburgh metropolitan area by a factor of five; the 1st and 99th percentiles by a factor of fifty (Figure 1). This variation in land prices arises from differences in proximity to places of employment and commerce, differences in access, quality, and availability of public goods, and variation in amenities among locations.[6] In contrast, construction costs for housing structures are not location specific and typically do not vary within a metropolitan area. Similarly wages for experienced car mechanics range from $25 to approximately $100 in our sample, while prices for replacement parts are more or less the same within the U.S. If wages in a service industry reflect local labor market conditions and if labor is not perfectly mobile, wages

[1]Price differences for tradeable goods are often bounded by per-unit transportation costs. These types of no-arbitrage conditions do not apply for non-tradeable goods.

[2]See Sieg, Smith, Banzhaf, and Walsh (2002) for an analysis of housing price differences in the LA metropolitan area.

[3]See Singh and Zhu (2005) for an empirical analysis of price differences in the market for rental cars.

[4]There is also some evidence of price dispersion between firms in US manufacturing as discussed, for example, by Abbott (1994).

[5]Differences in output prices may also arise due to heterogeneity among firms and imperfect competition due to product differentiation. See Berry (1994) for a discussion and further reference to the literature .

[6]Epple and Sieg (1999), Bayer, McMillan, and Reuben (2004), Ferreyra (2005), Ferreira (2005), and others provide evidence that consumers value these types of urban amenities.

will differ substantially among a set of locations. [7]

Estimating production functions is challenging if quantities and prices of output are not observed separately by the econometrician. If prices differ for the same good, the value of output is not 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 shown by Klette and Griliches (1996).[8] Separating price and quantity 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 production functions which treats the quantity and price of output as latent variables unobserved by the econometrician.

It is often convenient to assume that amounts of a good can be measured in terms of homogeneous units. This type of abstraction is valuable in theoretical modeling, for tractability and simplicity. In practice, as well, this type of abstraction is valuable. Industry aggregates routinely reported by government and industry entail such an abstraction, for example, reports on "the" housing market. While it is convenient to analyze markets for goods in terms of homogeneous units, it is rare outside of agricultural commodities to observe goods that are readily measured in such homogeneous units. In addition, variation in input prices across locations tends to result in differences in quantity consumed, so that price is not readily benchmarked in terms of a "typical" unit. For example, apartments in New York City are likely to be much less spacious than apartments in areas where land prices are less stratospheric. Within markets, goods are consumed in differing quantities by different individuals, again making it difficult to benchmark prices in terms of a typical house. A wealthy household my occupy a grand house while a household of more modest means occupies a small dwelling. We observe the value of the houses, but we do not

[7]Kennan and Walker (2005) document large differences in local wage rates in the U.S.

[8]The 'omitted price bias' is distinctly different than the standard 'transmission bias' that arises to due to the endogeneity of input factors (Marschak and Andrews, 1944). This simultaneity problem is also discussed in Mundlak and Hoch (1965) and Zellner, Kmenta, and Dreze (1966). Closely related to that endogeneity problem is the dynamic selection problem studied by Olley and Pakes (1996). See also Levinsohn and Petrin (2003), Melitz (2003), and Ackerberg, Caves, and Frazer (2005).

observe separate prices and homogeneous service units.[9] Similarly, we may be interested in estimating a production function for basic car repair services ignoring the differences among routine car repairs.

We consider a standard model of production in which a single output good is produced using a variety of input factors. One input factor (such as land or labor) must be purchased locally. As a consequence the price of this factor is not constant in the population. The other factors have prices that do not depend on the location and are thus constant. These factors can, therefore, be treated as a composite input factor. In our first application, we assume that housing is produced from land and non-land factors. Our second application focuses on the car repair industry. We treat labor as the input factor with location specific prices and estimate a production function for basic repair services.

To illustrate the basic ideas behind our approach, consider the housing application. Economic theory suggests that variation in land prices induces variation in the relative proportions of land and non-land factors used in production. Broadly speaking, housing developers will use different development strategies depending on the price of land. Housing developers will build more structures on small land areas if land is expensive and vice versa. The value of a house per unit of land is then a sufficient statistic that measures how land-intensive the production process is in equilibrium. Figure 2 illustrates that there is a lot of variation in the value of housing per unit of land in the sample of houses used in our application. We can trace out the observed equilibrium relationship between land prices and housing values per unit of land. This equilibrium locus implicitly characterizes the supply of housing per unit of land. Estimating this supply function then allows us to decompose the observed house value per unit of land into a price and a quantity component. We can thus recover the production function of housing from the housing supply function per unit of land.

Our approach is, therefore, based on duality theory. We assume that the production

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

function satisfies constant returns to scale.[10] This implies the existence of a well-behaved supply function per unit of the price-varying input factor. If the market is competitive and entry and exit occurs at low costs, profits of firms must be zero in equilibrium. We exploit the zero profit condition to derive an alternative representation of the indirect profit function. We show that there exists (under a set of regularity conditions) a differential equation that implicitly characterizes the supply function per unit of factor input. This differential equation is based on an application of Shepard's Lemma to our alternative characterization of the indirect profit function. 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 factor unit for arbitrary functional forms. With the supply function in hand, it is then straightforward to derive the production function.

These theoretical results directly map into a flexible estimation procedure. We use semi-nonparametric and nonparametric estimation techniques to characterize the relationship between the value of output per unit of input and the price of the location specific input factor.[11] The derivative of this equilibrium locus is the key ingredient in the differential equation that characterizes supply function per unit of input. Economic theory implies that the function that relates the price of inputs to the value of output per unit of input must be monotonically increasing in the factor price and that the derivate of this function is bounded by one. Semi-nonparametric methods are convenient since it is straight-forward to impose shape restrictions in estimation.[12] The approach proposed in this paper thus allows us to identify and estimate production functions with minimal functional form assumptions when output prices are unobserved. In contrast, almost all previous empirical papers assume that

[10]This assumption is fairly standard in the literature on housing construction. For other industries, this assumption is more controversial. Basu and Fernald (1997) reports estimates for 34 industries in the U.S. that suggests that a typical industry has approximately constant returns to scale, implying at most small mark-ups over marginal costs.

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

[12]For a discussion, see, for example, Matzkin (1994) and Chen (2006).

factor inputs and outputs are perfectly observed or ignore differences in output prices. It is the lack of observability of output quantities and prices that distinguishes our approach from previous work.[13]

To illustrate the usefulness of the techniques developed in this paper, we provide two applications. The first application focuses on estimating the production function of new housing in the Pittsburgh Metropolitan Area. We have obtained access to a comprehensive data set which includes all housing units in Allegheny county. We find that the approach suggested in this paper yields plausible and robust estimates of the underlying production function. Our second application focuses on the car repair industry. We have obtained access to a unique data set that is based on a survey conducted by Underhood Service Magazine, the leading publication for the industry. The survey provides detailed information about revenues, wages, and employment for a variety of small family-owned service and repair shops throughout the U.S. We use this data set to estimate a production function for the industry. Despite the much smaller size of this data set, we find that our approach yields quite plausible results in this application.

The rest of the paper is organized as follows. Section 2 presents the main theoretical results regarding supply and production functions. Section 3 introduces the estimation procedure used in this paper. Section 4 discusses the two empirical applications. Section 5 offers some concluding remarks.

2 Theory

We assume that a homogeneous good Q can be produced from two factors M and L via a production function $Q(L, M)$. The price of mobile factors, p_M, is constant across all locations. The price of the second factor, L depends on location. As a consequence the price of Q also depends on location. Since the good is non-tradeable, the production of the

[13] Klette and Griliches (1996) propose an estimator in the context of the linear model using an auxiliary demand model for differentiated products. Our approach is quite different from their approach. In particular, it neither requires the linearity assumption nor is it based on an auxiliary demand model.

good must also be local.[14] We make the following assumption:

Assumption 1 *The 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.

and

Assumption 2 *There is free entry and firms are price takers.*

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

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

Since p_m is constant throughout the population, we henceforth adopt the normalization $p_m = 1$.

Let $s(p_q)$ denote the normalized supply function, i.e. the supply function per unit of L 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.[15]

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

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

[14] Our first application focuses on housing construction. In that case L is land and M denotes non-land factors. Our second application looks at car repairs. Here we take labor as local factor.

[15] 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 L is, however, well-defined since it treats L as a fixed factor.

By the envelope theorem, we have:

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

The derivative of the per unit of L profit function is equal to the supply function per unit of L.

The omitted price problem arises if output and prices of output are not observed separately. We, therefore, have:

Assumption 3 *We observe the value of output per unit of L, denoted by v. We also observe p_l and L. We do not observe p_q or q.*

Our goal is to show that the production is identified under Assumptions 1-3. Our approach is based on duality theory. The basic idea behind our approach is the following. Under the assumptions made above, we can show that there exists a monotonic relationship between p_q and v. Since p_q is unobserved, the attention thus focuses on v instead. Moreover, we can show that there is another monotonic function that captures the equilibrium relationship between v and p_l. We can prove the following result:

Proposition 1 *There exists a locus of market equilibrium (p_l, v) values:*

$$p_l \;=\; r(v) \tag{4}$$

Proof:
The value of output per unit of L is defined as:

$$v \;=\; p_q s(p_q) \;=\; v(p_q) \tag{5}$$

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:

$$p_q = p_q(v) \tag{6}$$

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

$$\begin{aligned} p_l &= p_q(v)\, s(p_q(v)) \;-\; m(p_q(v)) \\ &= r(v) \end{aligned} \tag{7}$$

Q.E.D.

Example: Consider a Cobb-Douglas production function $Q = M^{\alpha} L^{1-\alpha}$ which implies that $q = m^{\alpha}$. Solving the firms optimization problem yields:

$$\begin{aligned} m(p_q) &= (\alpha p_q)^{\frac{1}{1-\alpha}} \\ s(p_q) &= (\alpha p_q)^{\frac{\alpha}{1-\alpha}} \end{aligned} \tag{8}$$

as a consequence we have:

$$\begin{aligned} v(p_q) &= p_q\, s(p_q) \\ &= \alpha^{\frac{\alpha}{1-\alpha}}\, p_q^{\frac{1}{1-\alpha}} \end{aligned} \tag{9}$$

Inverting this function yields

$$p_q(v) = \alpha^{-\alpha}\, v^{1-\alpha} \tag{10}$$

Moreover, it is straightforward to verify that:

$$r(v) = (1-\alpha)\, v \tag{11}$$

Based on this equilibrium locus, we can derive an alternative characterization of the indirect profit function. Substituting equation (5) into equation (4) yields:

$$\pi^*(p_q, p_l) = r(p_q s(p_q)) - p_l = 0 \tag{12}$$

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

$$\frac{\partial \pi^*(p_q, 1, p_l)}{\partial p_q} = r'(p_q s(p_q))[s(p_q) + p_q s'(p_q)] \tag{13}$$

Moreover, in equilibrium, we must have

$$\pi^*(p_q, p_l) \quad = \quad \pi(p_q, p_l) \tag{14}$$

We thus 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 L is implicitly characterized by the solution to the following differential equation:*

$$r'(p_q s(p_q)) \cdot [s(p_q) + p_q s'(p_q)] = s(p_q) \tag{15}$$

Proposition 2 summarizes an important methodological contribution of this paper. It shows that there exists a differential equation that characterizes the normalized supply function based on the equilibrium relationship between p_l and v. Moreover, the differential equation only depends on objects that are observed.

Example: (cont) Suppose the relationship between p_l and $r(v)$ is linear:

$$p_l \quad = \quad r(v) \ = \ (1 - \alpha)\, v \tag{16}$$

Equation (15) implies the following differential equation for the supply function:

$$(1-\alpha)\,[s+p_q s'] = s \tag{17}$$

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

$$s = c p_q^{\frac{\alpha}{1-\alpha}} \tag{18}$$

where c is the constant of integration. We conclude that 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.

We can show that a unique solution to our differential equation 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. However, absence of such a closed-form solution is not of great importance, as our subsequent applications demonstrate. To derive the general solution, rewrite equation (15) as:

$$(r'(p\,s)-1)\,s\,dp \;+\; r'(p\,s)\,p\,ds \;=\; 0 \tag{19}$$

We have the following result:

Proposition 3 *The integrating factor $\mu(p,s) = p\,s$ converts (19) into an exact differential equation.*[16] *As a consequence the solution to equation (19) is:*

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

[16]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 4. 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 Boyce and DiPrima (2004) for a detailed discussion.

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 (19) can be written:

$$M(p,s)dp + N(p,s)ds = 0 \tag{20}$$

where

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

Straightforward differentiation then establishes that the necessary and sufficient condition for (20) to be exact is satisfied $\partial M/\partial s = \partial N/\partial p$. The second result follows from the first result by invoking the solution of an exact differential equation. Q.E.D.

Having derived the normalized supply function, it is straight forward to derive the underlying production function. Let

$$m^*(p_q) = p_q s(p_q) - r(p_q s(p_q)) \tag{21}$$

Points on the production function $q(m)$ are then given by $(m^*(p_q), s(p_q))$. Let the inverse of (21) be $p_q^*(m)$. Then the production function is equivalently written:

$$q(m) = s(p_q^*(m)) \tag{22}$$

3 Estimation

The theoretical results presented in the previous section directly translate into algorithms that can be used to estimate the production function. We adopt a semi-nonparametric approach that does not require restrictive assumptions to obtain a consistent estimator of the production function. In our application it is convenient to approximate the unknown $r(v)$ function with a polynomial of arbitrary order k:

$$p_l = \sum_{i=1}^{k} \frac{r_i}{i} v^i \ + \ \epsilon \tag{23}$$

where the error term ϵ may, for example, reflect the fact that p_l is measured with error. Assuming that $E(\epsilon|v) = 0$, we can estimate this equilibrium locus using least squares based on a sample of observations with size N.[17] 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. [18] The main drawback of this approach 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. Using polynomials in estimation, however, has the advantage that we can easily characterize the normalized supply function once an appropriate order of the polynomial has been determined. We have the following general result:

Proposition 4 *Substituting equation (23) into (15) and normalizing such that* $s(1) = 1$*, the implicit solution to the differential equation gives the supply function:*

$$\sum_{i=2}^{k} \frac{r_i}{i-1} \left[(ps)^{i-1} - 1 \right] + (r_1 - 1)\log(p) + r_1 \log(s) = 0 \tag{24}$$

Proof:

Applying Proposition 2, we obtain

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

[17] We also explore a variety of instrumental variable estimators in our application.

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

or

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

Normalizing such that $s(1) = 1$ implies $c = \sum_{i=2}^{k} \frac{r_i}{i-1}$. Q.E.D.

After estimating $r(v)$, the equation in Proposition 4 could be numerically solved for $s(p)$. The next Proposition provides a simpler approach to calculate the supply function.

Proposition 5 *A closed-form expression for the supply function in the general polynomial case, expressed solely in terms of* v *and* $\{r_i\}$*, is*

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

Proof of Proposition 5:

Rearrange equation (24) such that

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

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

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

Q.E.D.

One 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.[19] However, we find in our applications that the unconstrained estimator satisfies the shape restrictions discussed above. There is therefore no need to discuss these issues in detail here.

Alternatively, we can estimate $r(v)$ using a fully nonparametric estimator, such as a kernel estimator.[20] After we have obtained an unrestricted kernel estimate of $r(v)$, we also need to check whether the derivative restrictions are satisfied everywhere.[21] With the bandwidth set equal to the standard deviation of v, we find in our housing application that the derivative conditions are met. Other reasonable values for the bandwidth were tested, and yielded similar results. We thus conclude, that, at least in the housing application studied below, the unrestricted kernel estimator satisfies the restrictions that economic theory imposes on $r(v)$. Once we have obtained a nonparametric estimate of the function $r(v)$, we solved the ordinary differential equation in Proposition 2 using the boundary-value condition $s(1) = 1$. We used piecewise polynomials to interpolate $\hat{r}'$ during the process of solving the differential equation.

Finally, we also explore some simple parametric specification for estimating (23), such as log-linear models, and investigate whether these simple parametric forms capture the observed variation in the data.[22]

In summary, the semi-nonparametric approach based on polynomial approximations of the underlying equilibrium function $r(v)$ yields a closed form solution to the estimated normalized supply function. We have seen that it is straightforward to estimate the underlying production function once we have estimated the supply function.

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

[20] A review of the literature in applied non-parametric regression analysis is given by Haerdle and Linton (1994).

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

[22] An appendix that discusses the properties of the log-linear case is available upon request from the authors.

4 Empirical Applications

4.1 New Housing Construction

Our first empirical application focuses on new housing construction in Allegheny County in Pennsylvania, which contains the greater metropolitan area of Pittsburgh. Our main data source is the Allegheny County web site, which is maintained by the Office of Property Assessments.[23] The web site provides access to a database with detailed information about all properties, both residential and commercial, in the entire county. The database is updated on a yearly basis, and contains a wide array of information about the property. In contrast to most other publicly available data set, this data set contains separate data for the value of land and the value of housing structures of a dwelling. We also observe the land area for each house as well as a large vector of structural characteristics of the house.

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 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. 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. Figure 3 illustrates the location of the properties within Allegheny County.

In July of 2004, the Pennsylvania State Tax Equalization Board performed a study of all

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

Table 1: Descriptive Statistics

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

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

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. For these reasons, we chose the assessed market values since they appear to be accurate and give us a measure of housing value for all properties. We do not have comparable validation of assessed land prices. However, prior to 2001, values of land and structures were taxed separately. thus, there was in incentive for assessors to provide meaningful estimates for land values. Since we focus on new construction, assessors may have ad access to transaction prices for properties purchased for new construction. Thus while measurement error for land values is likely greater than for property values, there is no reason to expect systematic errors in land values.

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.

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.

Table 2: OLS and IV estimates of $r(v)$

OLS Estimates				
	Log-linear	Linear	Quadratic	Cubic
v		0.1394***	0.1685***	0.1622***
v^2			-0.0002***	-0.0001
v^3				3.9e−7*
Constant	-1.6051***			
log(v)	0.9090***			
R^2	0.8649	0.8014	0.8382	0.8391
N	6,362	6,362	6,362	6362
2SLS Estimates				
	Log-linear	Linear	Quadratic	Cubic
v		0.1440 ***	0.1631***	0.1732***
v^2			-0.0002***	-0.0005***
v^3				1.1e−6*
Constant	-1.6129***			
log(v)	0.9119***			
R^2	0.8649	0.7992	0.8360	0.8135
N	6,362	6,362	6362	6,362

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

Comparing the fit of the log-linear equation to the polynomial approximations, we see that the log-linear form compares well with the polynomial forms.

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 reestimated 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.

The OLS estimates are based on the assumption that the error in our model reflects measurement error in price per unit of land. Since v is an endogenous variable that may also be measured with error, we also estimate our models using instrumental variable estimation. 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. Two-stage least squares results can be found in the bottom panel of Table 2. While results for the log-linear cases are quite similar, 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.

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 fairly similar. Figure 4 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 fairly 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. We also estimated the supply function using the OLS and IV versions on the post-1995 data set and found similar results.

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 0.139. This implies that the Cobb-Douglas production function is given by $Q(L, M) = 1.138 * L^{0.139} * M^{0.861}$. As before, we find that the different econometric specifications of the $r(v)$ function yields similarly shaped production functions. Figure 5 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 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.[24]

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).

[24]These findings are broadly consistent with early empirical studies on housing supply. 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.

4.2 The Car Repair Service Industry

The second 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.[25] 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 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).

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

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 documents 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 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 overfit, 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.

4.3 Output Price Dispersion and Price Index Construction

Having estimated the production function, we can decompose the observed value of output into a price and a quantity component. Note that the inputs m are "observed" by the econometrician and given by:[26]

$$m = v - p_l \tag{25}$$

The quantity of output per unit of L is given by $q(m)$. By definition, the price of output must satisfy:

$$p_q = \frac{v}{q(m)} \tag{26}$$

The normalization imposed in computing the supply function per unit of L imposes a normalization on the measurement of output units. Prices are thus measured in dollars per implied unit.

Suppose we have a sample of N observations from J different regions. Let N_j denote the number of observations from region j. The average price of output in region j is then given by:

$$P_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{v_i}{q(m_i)} \tag{27}$$

A simple price index of region j relative to a base line region 1 is then, for example, given by P_j/P_1.

To illustrate the usefulness of this procedure we consider our sample of car repair service providers. Figure 6 shows the density of prices that are implied by the estimated production function reported in the previous section. We find that the vast majority of the price

[26] Instead of using p_l in the computation, we can impose the restrictions implied by our model and use the predicted price $\hat{p}_l = \hat{r}(v)$ that is consistent with optimal firm behavior in the construction of the price index. Not surprisingly, we find that the two alternatives give quite similar results.

estimates range between 0.5 and 4. The estimate output price dispersion among the service providers is thus quite substantial. This finding, therefore, supports the central conjecture that underlies this paper. It is also broadly consistent withe evidence of price dispersion for specific services reported in Table 3.

We then assign each car repair service provider to one of the following 4 regions in the U.S.: Northeast , Southeast, Midwest, and West. Finally we construct price indices for the 4 regions. To facilitate the comparison of results we normalize the price index for the northeast to be equal to one. The estimates for the others are: 0.7668 in the Southeast, 0.9711 in the Midwest, and 0.6505 in the West. Given the relative small sample size of our subsamples by region, these results are best taken as illustrative of an approach that be readily applied by, for example, government data collecting agencies such as the Bureau of Labor Statistics.

We have thus seen that we can construct simple price indices once we have estimated the underlying production for output. To highlight the main identifying assumptions, it is useful to compare our approach to more traditional approaches for price index construction. The main alternative that is commonly used by the Bureau of Labor Statistics is based on hedonic regressions (Rosen, 1974). Following that approach, we would regress the value of output of observed measures of product characteristics and a full set of regional dummy variables. The basic idea is that the product characteristics pick up differences in quality among products. The regional dummy variables capture price differences among the et of localities. A hedonic approach thus treats a good as a differentiated product that is primarily valued for its associated characteristics. As long as we observe all relevant product characteristics, a very demanding assumption, we can consistently estimate the underlying price function and thus decompose values into a price and a quantity component. In contrast, our approach assumes that differences in observed products can be measured by a one-dimensional quantity index. Our approach is otherwise agnostic about product characteristics and can be implemented without observing any product characteristics.

5 Conclusions

In this paper, we have demonstrated how to estimate production functions when output prices are unobserved. We have developed a new approach that allows us to identify and estimate the underlying production function without relying on strong functional form assumptions or auxiliary demand models.[27] We have shown that the observed variation in input prices and output values per unit of input is sufficient to identify and estimate production functions. We have illustrated our approach with two applications. One application is based on data housing production in Allegheny County. We 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. Another application is based on unique data for the car repair service industry and documents the applicability of our method in quite a different context.

Our research has some important implications for policy analysis. Consider, for example, the application that focuses on new housing construction. Quantifying new housing supply is an important ingredient for 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 is largely met by an increase in housing supply. Hence we would expect that 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.[28]

Our research also has some broad implications for understanding the industrial organization of many industries. Much recent research in economics focuses on differences in

[27] A similar problem arises when the output measure is quantity measure, but ignores quality differences. See, for example, the discussion in Van Biesebroeck (2003).

[28] Glaeser and Gyourko (2005) and Brueckner and Rosenthal (2005) have recently argued that understanding housing supply is key to understand the growth and decline of urban areas.

productivity and their impact on economic growth and welfare. Having reliable estimates of production functions provides the micro foundation for this important literature. The recent literature in industrial organization has highlighted the fact that there exists a significant amount of heterogeneity among firms in the same industry. Some of this heterogeneity is clearly driven by the fact that firms operate in a variety of local markets and face a different sets of factor prices. We have shown in this paper how to exploit these differences in factor prices to estimate production functions, even if output prices are not observed by the econometrician. The approach developed in this paper and the associated applications are encouraging for future research to improve understanding of production technologies, productivity, and technological progress.

References

Abbott, T. (1994). Observed price Dispersion: Product heterogeneity, Regional Markets, or Local Market Power?. *Journal of Economics and Business*, *46*, 21–37.

Ackerberg, D., Caves, K., and Frazer, G. (2005). Structural Estimation of Production Functions. Working Paper.

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

Bayer, P., McMillan, R., and Reuben, K. (2004). The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting. Working Paper.

Beresteanu, A. (2005). Nonparametric Estimation of Regression Functions Under Restrictions on Partial Derivatives. Rand Journal of Economics.

Berry, S. (1994). Estimating Discrete-Choice Models of Product Differentiation. *Rand Journal of Economics*, *25(2)*, 242–262.

Boyce, W. and DiPrima, R. (2004). *Elementary Differential Equations*. Wiley.

Brueckner, J. and Rosenthal, S. (2005). Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich?. Working Paper.

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

Christensen, L., Jorgenson, D., and Lau, L. (1973). Trancendental logarithmic production frontiers. *Review of Economics and Statistics*, *55*, 28–45.

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

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

Ferreira, F. (2005). You Can Take It with You: Proposition 13 Tax Benefits, Residential Mobility, and Willingness to Pay for Housing Amenities. Working Paper.

Ferreyra, M. (2005). Estimating the Effects of Private School Vouchers in Multi-District Economies. Working Paper.

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

Glaeser, E. and Gyourko, J. (2005). Urban Decline and Durable Housing. *Journal of Political Economy*, *113*, 345–75.

Haerdle, W. and Linton, O. (1994). Applied Nonparametric Methods. In *Handbook of Econometrics IV*. Elsevier North Holland.

Kennan, J. and Walker, J. (2005). The Effect of Expected Income on Individual Migration Decisions. Working Paper.

Klette, T. and Griliches, Z. (1996). The Inconsisteny of Common Scale Estimators When Output Prices are Unobserved and Endogeneous. *Journal of Applied Economics*, *11*, 343–361.

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

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

Marschak, J. and Andrews, W. (1944). Random Simulataneous Equations and the Theory of Production. *Econometrica*, *12*, 143–205.

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

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

Melitz, M. (2003). Estimating Firm-level Productivity in Differentiated Product Industries. Working Paper.

Mukerjee, H. (1988). Monotone Nonparametric Regressions. *Annals of Statistics*, *16*, 741–750.

Mundlak, Y. and Hoch, I. (1965). Consequences of Alternative Specification of Cobb-Douglas Production Functions. *Econometrica*, *33*, 814–28.

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

Olley, S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*, 1263–98.

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

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*, 34–55.

Sieg, H., Smith, V. K., Banzhaf, S., and Walsh, R. (2002). Interjurisdictional Housing Prices in Locational Equilibrium. *Journal of Urban Economics*, *52*, 131–153.

Singh, V. and Zhu, T. (2005). An Empirical Analysis of Pricing in the Market for Rental Cars. Working Paper.

Van Biesebroeck, J. (2003). Productivity Dynamics with Technology Change: An Application to Automobile Assembly. *Review of Economic Studies*, *70 (1)*, 167–198.

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

Zellner, A., Kmenta, J., and Dreze, J. (1966). Specification and Estimation of Cobb-Douglas Production Functions. *Econometrica*, *34*, 784–95.

Figure 1: Density of Land Prices

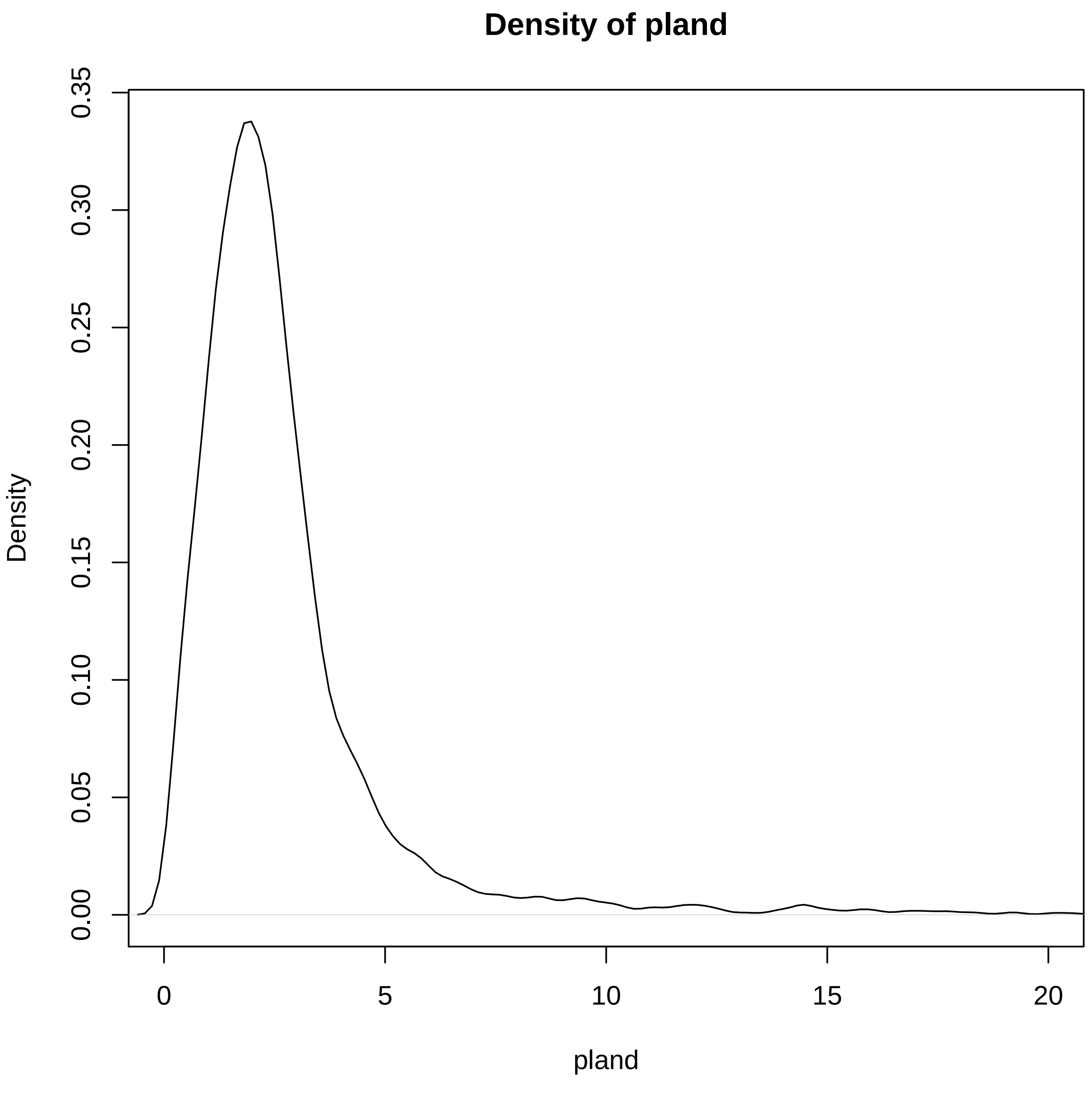


Figure 2: Density of Housing Values per Unit of Land

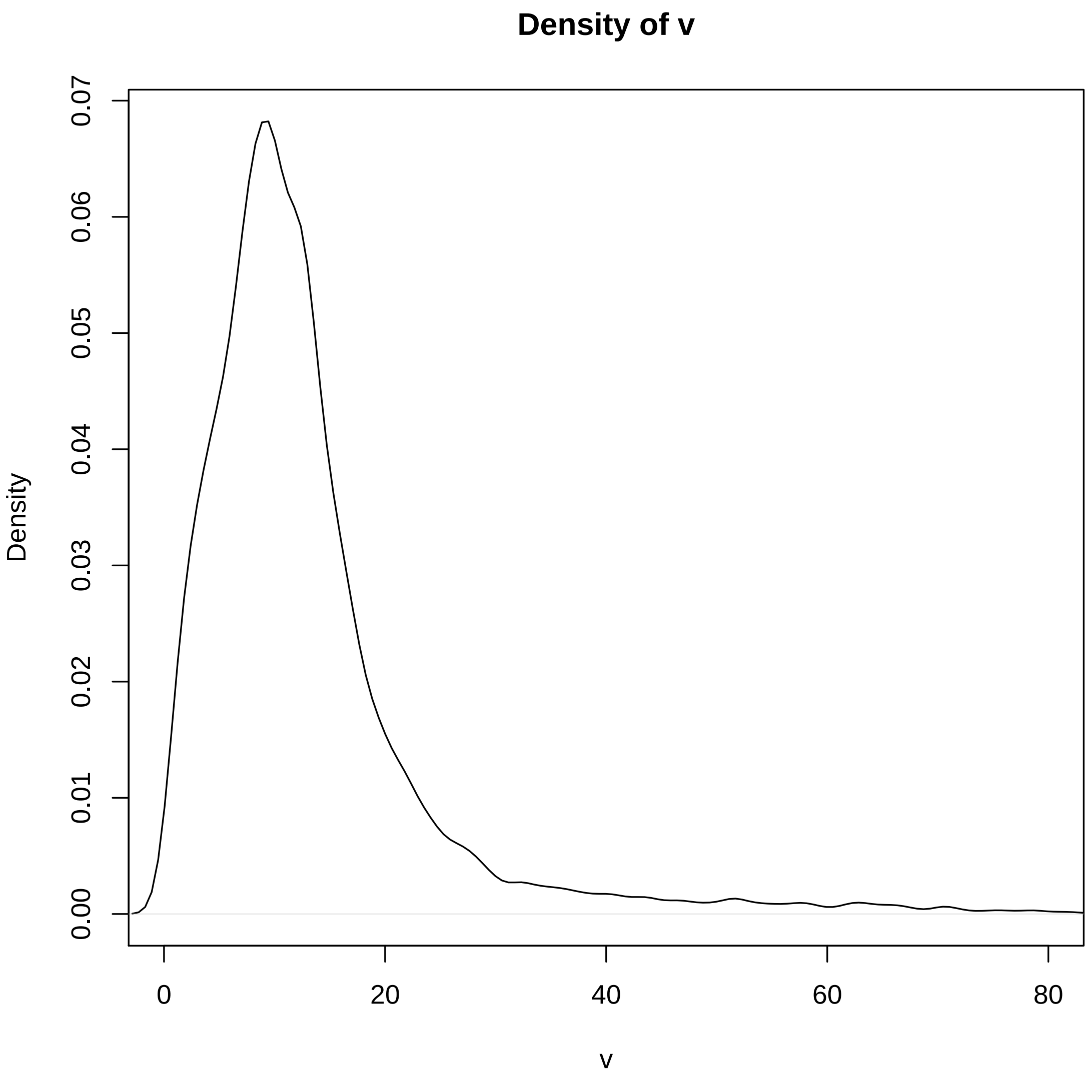


Figure 3: Map of Properties in the Data Set

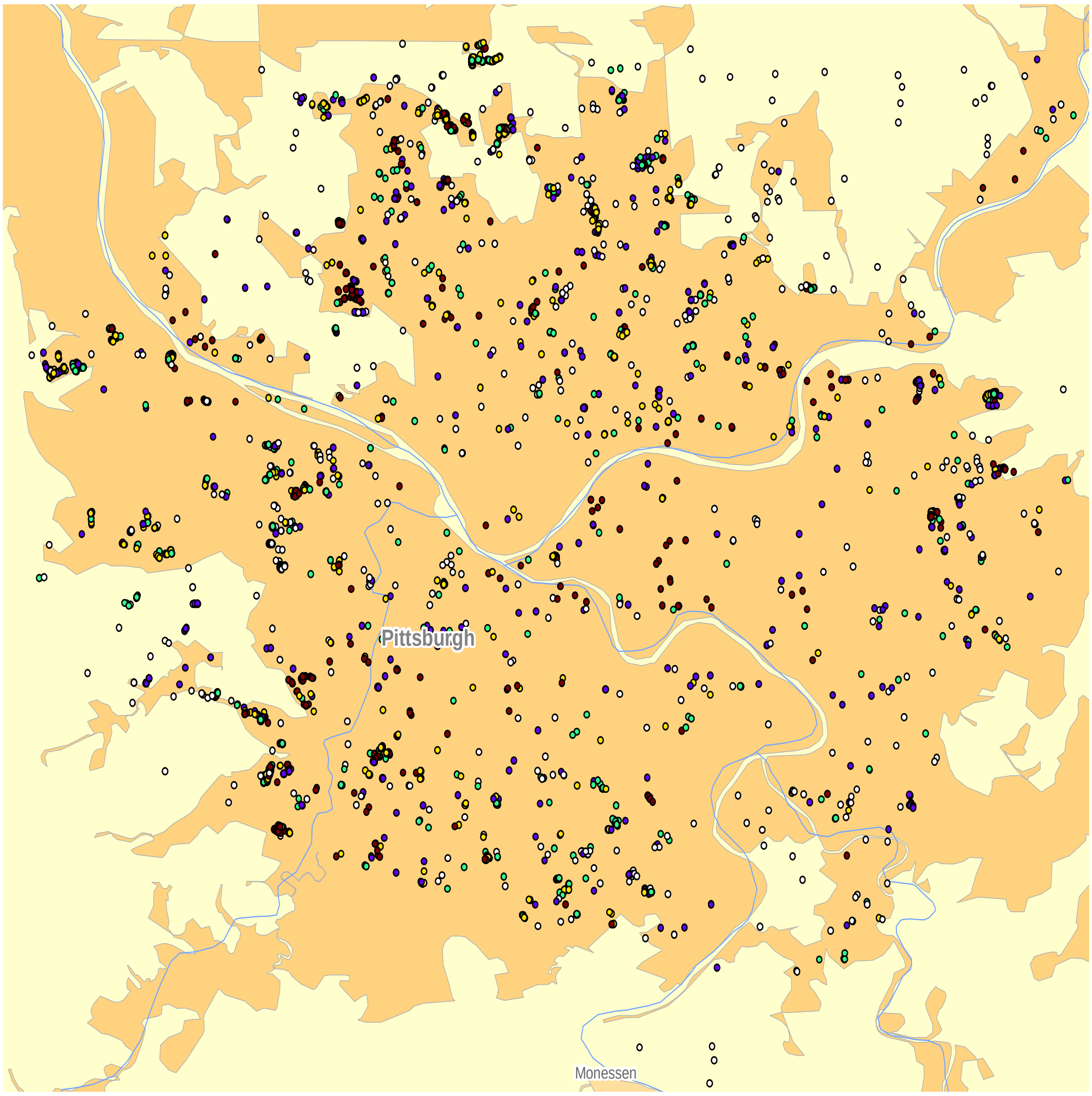
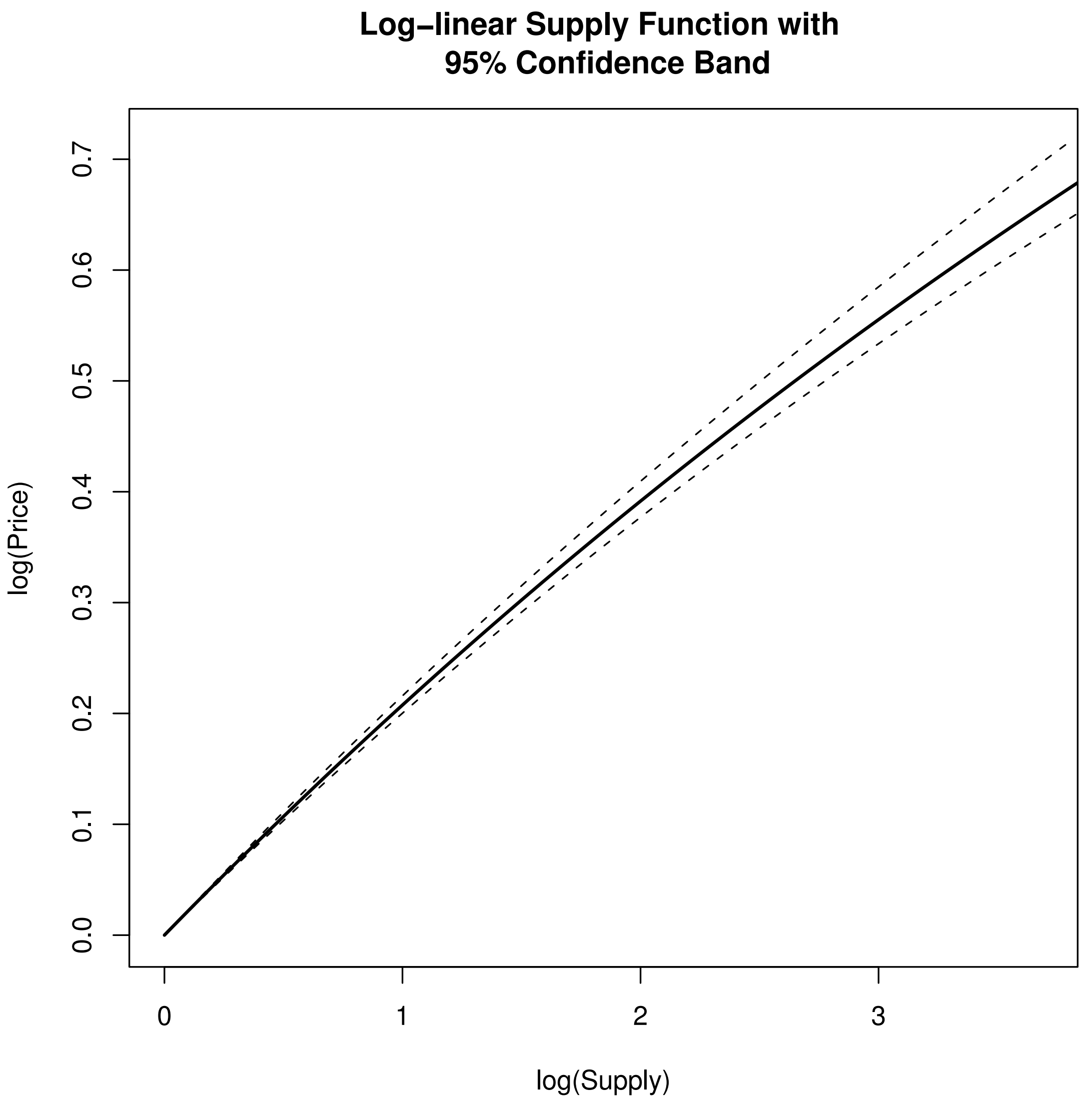


Figure 4:



Log−linear Supply Function with
95% Confidence Band
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
log(Price)
0
1
2
3
log(Supply)

Figure 5:

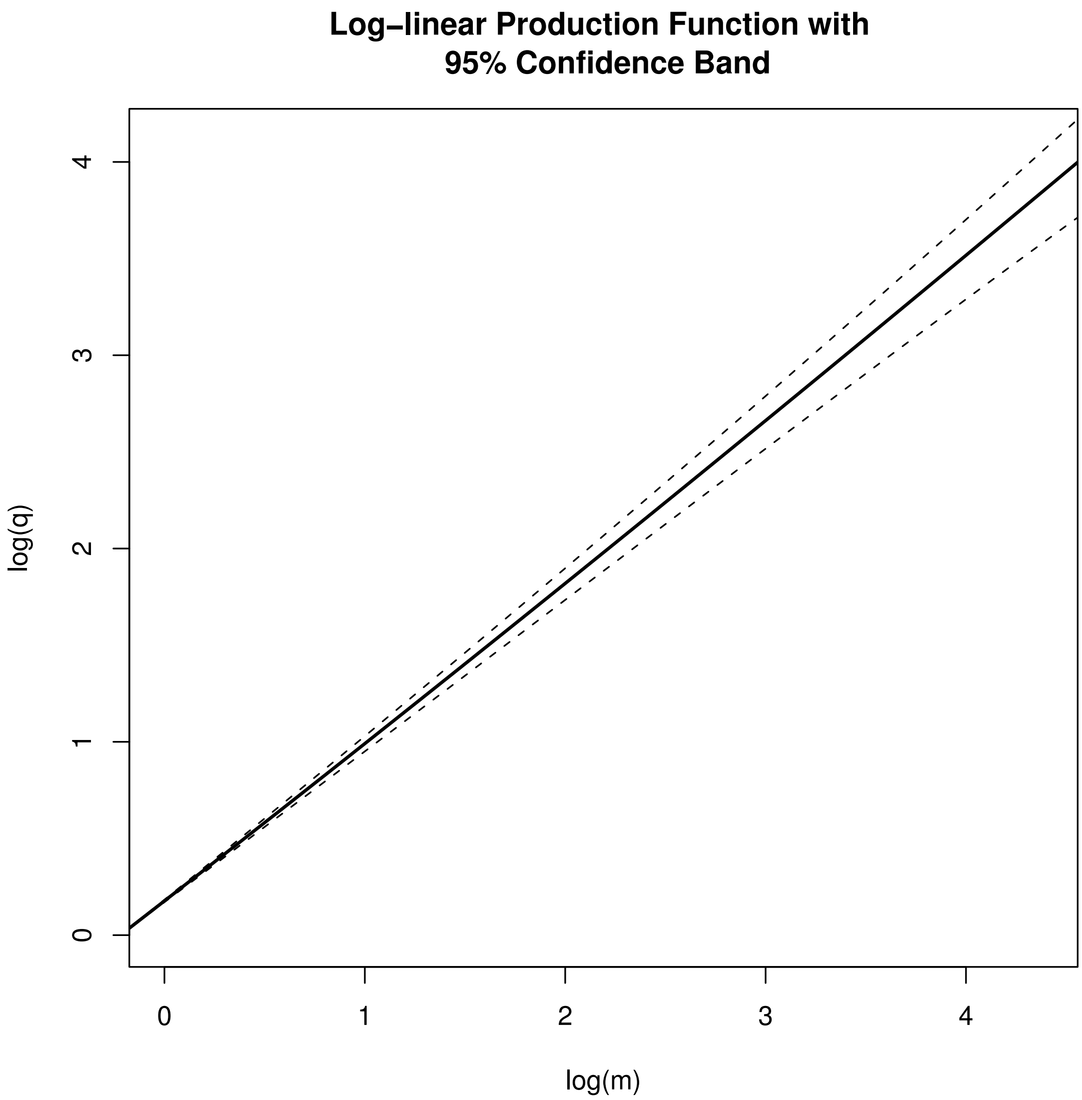
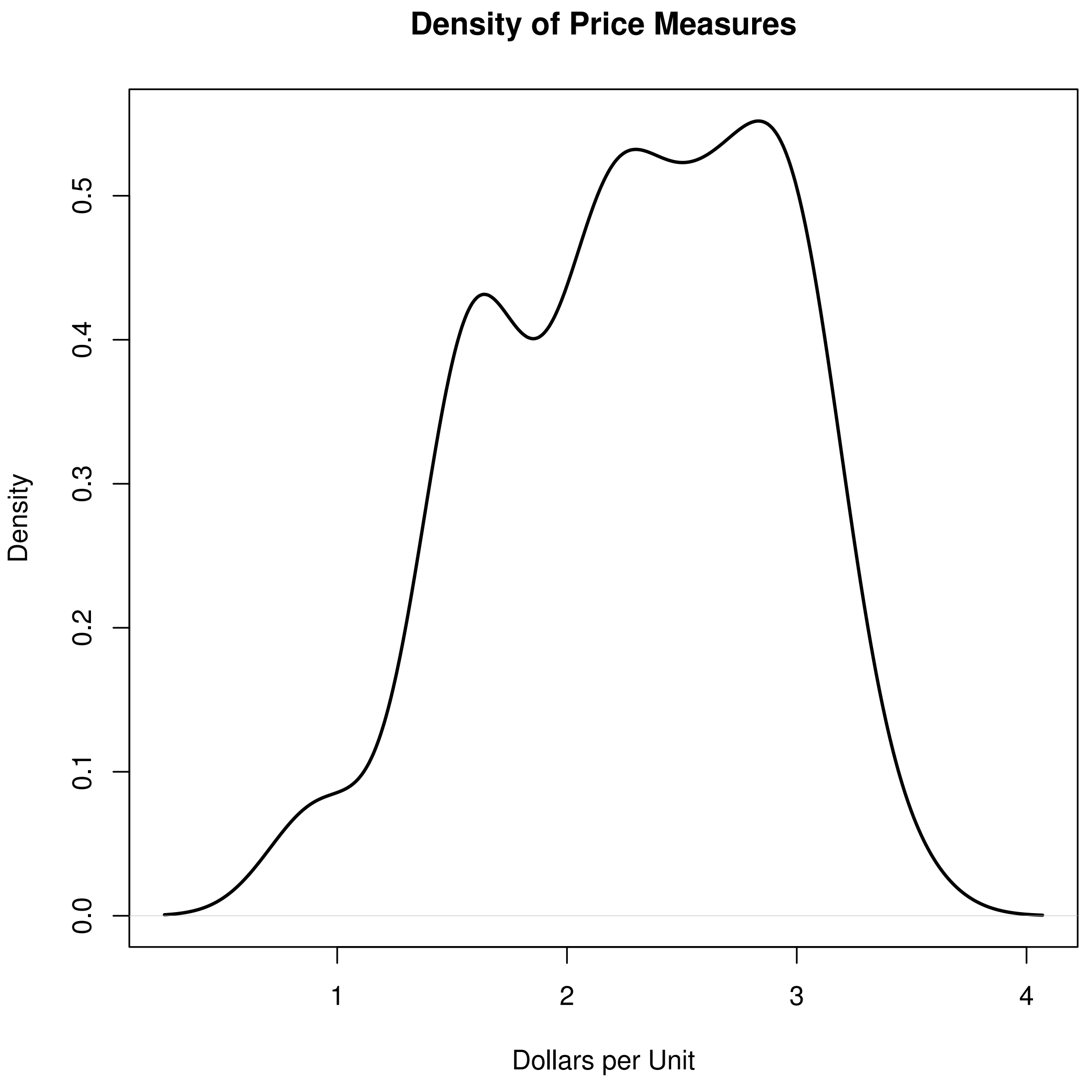

Log−linear Production Function with
95% Confidence Band
log(q)
0
1
2
3
4
log(m)
0
1
2
3
4

Figure 6:


Density of Price Measures
Density
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
Dollars per Unit