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

Dennis Epple
Carnegie Mellon University and NBER

Brett Gordon
Carnegie Mellon University

Holger Sieg
Carnegie Mellon University and NBER

August 18, 2006

*We would like to thank Lanier Benkard, Richard Blundell, Andrew Chesher,George-Levi Gayle, Ron Goettler, Francois Ortalo-Magne, Costas Meghir, Marc Mündler, Amil Petrin, Steven Ross, Vishal Singh, Frank Wolak, 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, SITE, 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**

If prices differ for the same good, the value of output is not necessarily a good measure of the quantity of output. Estimation of production functions is challenging if quantities and prices of output are 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 new housing construction. 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 and is 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

Production functions are identified if and only if there is sufficient variation in input factors. Firms often use different combinations of input factors since they face different local factor prices. This then implies that marginal costs and output prices differ among firms. There are many useful 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 Marschak and Andrews (1944).[1] While most applied researchers are aware of this common problem, they typically ignore it. This is primarily the case because there are no general methods that have been developed in the literature to cope with this type of omitted price problem.[2] The main objective of this paper is to develop and apply a new method 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. While it is useful to analyze markets for goods in terms of homogeneous units, it is rare outside of agricultural commodities or raw materials to observe goods that are readily mea-

[1]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).

[2]The main exception is Klette and Griliches (1996) who rely on an auxiliary pricing model due correct for the omitted price bias.

sured in such homogeneous units. The existence of a production function itself often entails a powerful abstraction. There may not be an easily measurable price associated with this theoretical construct. 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. Consider for example, housing. We observe the value of a newly constructed house, but we do not observe separate prices and homogeneous service units.[3]

We first consider a standard model of production in which a single output good is produced using two input factors. The first input factor must be purchased locally and its price varies by location. The price of the other factor does not depend on the location and is thus constant. In the housing application we assume that land and structures are the two factors.[4] Differences in output prices are due to differences in local factor prices in our model.[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 a number of sources such as 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.[7]

---

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

[4]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.

[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.

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

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 significant variation in the value of housing per unit of land in the sample of houses used in our application.

Our approach to identification and estimation is based on duality theory.[8] We assume that the production function satisfies constant returns to scale.[9] We can, therefore, normalize output in terms of one input factor (land). While we do not observe the price or quantity of output, we observe the value of output per unit of land. We show in this paper that the price of output is a monotonically increasing function of the value of output per unit of land. Since the price of output is unobserved, the attention thus focuses on value of output per unit of land instead. Constant returns to scale also implies that profits of firms must be zero in equilibrium. We exploit the zero profit condition and derive an alternative representation of the indirect profit function as a function of the price of land and value of output per unit of land.

Differentiating the alternative representation of the indirect profit function with the respect to the (unobserved) price of output gives rise to a differential equation that implicitly characterizes

Walker (2005) document large differences in local wage rates in the U.S.

[8]Duality between the production function and the price possibility frontier was introduced by Samuelson (1953), and is discussed, among others, in Burmeister and Kuga. (1970), Shepard (1970), Diewert (1973), and Jorgenson and Lau (1974).

[9]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.

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 factor unit for arbitrary functional forms. With the supply function in hand, it is then straightforward to derive the 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 our alternative representation of the indirect profit function.[10] The derivative of the indirect profit function is the key ingredient in the differential equation that characterizes supply function per unit of input. Thus, the approach proposed in this paper allows us to identify and estimate production functions with minimal functional form assumptions when output prices and quantities are unobserved. In contrast, almost all previous empirical papers assume that factor inputs and output quantities are perfectly observed. It is the lack of observability of output quantities and prices that distinguishes our approach from previous work.[11]

[10] Diewert (1971) and Christensen, Jorgenson, and Lau (1973) 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.

[11] 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.

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. It is based on a comprehensive data set that includes all recently built 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

First we present the basic theoretical results for a model with two input factors. We then discuss the extension to a model with three and more input factors.

## 2.1 Production with Two Inputs

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, $p_l$ depends on location. We assume that the good is non-tradeable. Hence the production of the good must also be local. As a consequence the price of output, $p_q$, also depends on location.[12] The underlying production function has the following properties:[13]

**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.*

We also assume that the industry is competitive.

**Assumption 2** *Firms behave as price takers.*

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 $L$ 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:

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

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

[12]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 the local factor.

[13]Christensen et al. (1973) use the similar qualitative assumptions.

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.[14]

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

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

By the envelope theorem, we have:

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

The derivative of the per unit of $L$ profit function is equal to the supply function per unit of $L$. Computation of $s(p_q)$ is, therefore, simple if we can compute (or estimate) the indirect profit function.[15]

The omitted price problem arises if quantities and prices of output are not observed separately by the econometrician. The theoretical model is thus motivated by the following assumption on observables:

**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 $Q$.*

[14] 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.

[15] The normalized indirect profit is closely related to Samuelson's (1953) factor-price frontier. Christensen et al. (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 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 $q(m)$.

Our approach is based on duality theory. The basic idea is the following. 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.

Summarizing the first two steps of our analysis, we have the following result:

**Proposition 1** *The value of output 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:*

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

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:

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

Q.E.D.

To illustrate the result in Proposition 1 it is useful to consider an example.

**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 the zero profit condition implies that:

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

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:

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

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

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

Moreover, in equilibrium, we must have

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

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 the function $r(\cdot)$ which can be consistently estimated based on observed outcomes.

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

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

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 $p$. 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 equa-*

*tion.*[16] *As a consequence the solution to equation (19) is:*

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

*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, 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. Q.E.D.

[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 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 Boyce and DiPrima (2004) for a detailed discussion.*

Having derived the normalized supply function, it is straightforward 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:

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

## 2.2 Production With Multiple Inputs

Thus far we have shown that we can recover the 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 $L$. We assume that $p_n$ and $p_l$, vary by location. $p_m$ is fixed.[17]

**Assumption 4** *We observe $V = p_q Q$, $N$, $L$, $p_n$, and $p_l$. We do not observe $p_q$ and $Q$.*

As before let

$$\pi(p_q, p_n, p_l) \;=\; p_q\, s(p_q, p_n) \;-\; m(p_q, p_n) \;-\; p_n\, n(p_q, p_n) \;-\; p_l \tag{23}$$

[17] 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.

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

$$\begin{aligned} \frac{\partial \pi}{\partial p_q} &= s(p_q, p_n) \\ \frac{\partial \pi}{\partial p_n} &= -n(p_q, p_n) \end{aligned} \tag{24}$$

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

$$v = p_q\, s(p_q, p_n) \tag{25}$$

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 $p_n$:

$$p_l = r(v, p_n) \tag{26}$$

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 (26) then implies the following alternative representation of the indirect profit function:

$$\pi^*(v(p_q, p_n), p_n) = r(p_q\, s(p_q, p_n), p_n) - p_l \tag{27}$$

Taking partial derivatives, we have:

$$\frac{\partial \pi^*}{\partial p_q} = \frac{\partial r}{\partial v} \frac{\partial v}{\partial p_q} \tag{28}$$

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

and

$$\begin{aligned}\frac{\partial \pi^*}{\partial p_n} &= \frac{\partial r}{\partial v}\frac{\partial v}{\partial p_n} + \frac{\partial r}{\partial p_n} \\ &= \frac{\partial r}{\partial v}p_q\frac{\partial s(p_q,p_n)}{\partial p_n} + \frac{\partial r}{\partial p_n}\end{aligned} \tag{29}$$

Note that in equilibrium

$$\begin{aligned}\pi(p_q,p_n) &= \pi^*(p_q,p_n) \\ n(p_q,p_n) &= n^*(p_q s(p_q,p_n),p_n)\end{aligned} \tag{30}$$

We thus obtain the generalization of Proposition 2:

**Proposition 4** *The supply function is the solution to the following fundamental system of partial differential equations:*

$$\begin{aligned}s(p_q,p_n) &= \frac{\partial r}{\partial v}\left[s(p_q,p_n) + p_q\frac{\partial s(p_q,p_n)}{\partial p_q}\right] \\ -n^*(p_q s(p_q,p_n),p_n) &= \frac{\partial r}{\partial v}p_q\frac{\partial s(p_q,p_n)}{\partial p_n} + \frac{\partial r}{\partial p_n}\end{aligned} \tag{31}$$

Note that the functions $r$ and $n^*$ only depend on observables.

**Example: cont** In the three factor Cobb-Douglas Case, we have

$$q = m^\alpha\, n^\beta \tag{32}$$

The FOC's of the firms maximization problem imply that

$$\begin{aligned} r(v,p_n) &= (1-\alpha-\beta)v \\ n^*(v,p_n) &= \beta\frac{v}{p_n} \end{aligned} \tag{33}$$

Hence we can write the fundamental system of partial differential equations as:

$$\begin{aligned} s(p_q,p_n) &= \frac{1-\alpha-\beta}{\alpha+\beta}\, p_q\, \frac{\partial s(p_q,p_n)}{\partial p_q} \\ s(p_q,p_n) &= -\frac{1-\alpha-\beta}{\beta}\, p_n\, \frac{\partial s(p_q,p_n)}{\partial p_n} \end{aligned} \tag{34}$$

It is straight forward to verify that the solution to this problem is

$$s(p_q,p_n) = c\, p_n^{\frac{-\beta}{1-\alpha-\beta}}\, p_q^{\frac{\alpha+\beta}{1-\alpha-\beta}} \tag{35}$$

In general, 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

$$\begin{aligned} n(p_q,p_n) &= n^*(p_q s(p_q,p_n),p_n) \\ m(p_q,p_n) &= m^*(p_q s(p_q,p_n),p_n) \end{aligned} \tag{36}$$

By varying $p_q$ and $p_n$ We can therefore trace out the relationship between $n, m$ and $q = s(p_q,p_n)$.[18]

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

It is of interest to note the relationship of our approach to the standard approach based on duality theory.[19] 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 function can be estimated without output price data.[20] Below we present results for these cases as part of our empirical analysis.

## 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 functional form assumptions to obtain a consistent estimator of the production function. In our application it is convenient to approximate the unknown $r(v)$ function

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

[20] 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.

with a polynomial of arbitrary order $k$:

$$r_k(v) = \sum_{i=1}^{k} \frac{r_i}{i} v^i \tag{37}$$

Using polynomials 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 5** *Substituting equation (37) into (15) and normalizing such that $s(1) = 1$, the implicit solution to the differential equation gives the supply function:*

$$\sum_{i=2}^{k} \frac{r_i}{i-1} \left[ (ps)^{i-1} - 1 \right] + (r_1 - 1)\log(p) + r_1 \log(s) = 0 \tag{38}$$

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

$$\begin{aligned} s &= \frac{v^{1-r_1}}{\exp\left\{\sum_{i=2}^{k} \frac{r_i}{i-1}(v^{i-1}-1)\right\}} \\ p &= v^{r_1} \exp\left\{\sum_{i=2}^{k} \frac{r_i}{i-1}(v^{i-1}-1)\right\} \end{aligned} \tag{39}$$

Proof:

Applying Proposition 2, we obtain

$$\sum_{i=2}^{k} \frac{1}{i-1} (ps)^{i-1} r_i + (r_1 - 1)\log(p) + r_1 \log(s) = c \tag{40}$$

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

$$\sum_{i=2}^{k} \frac{r_i}{i-1} \left[ (ps)^{i-1} - 1 \right] + r_1 \log(ps) = \log(p) \tag{41}$$

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

$$s = \frac{v}{p} = \frac{v^{1-r_1}}{\exp\left\{\sum_{i=2}^{k} \frac{1}{i-1} v^{i-1} r_i - c\right\}} \tag{42}$$

Q.E.D.

Based on the polynomial specification, we then obtain the following econometric model

$$p_l \;=\; \sum_{i=1}^{k} \frac{r_i}{i} v^i \;+\; \epsilon \tag{43}$$

where $\epsilon$ is an additively separable error associated with the indirect profit function.[21]

Assuming that $E(\epsilon|v) = 0$, we can estimate $r(v)$ using least squares 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] Depending on how one interprets the error term in equation (43), the assumption that $E(\epsilon|v) = 0$ may not be plausible. For example, if the error reflects productivity shocks, one may expect that $v$ may be correlated with $\epsilon$. As long as we have a sufficient number of instruments $z$ that satisfy $E(\epsilon|z) = 0$, we can estimate the model above using a parametric IV estimator. Moreover, we can also use the semi-nonparametric IV techniques suggested by Newey and Powell (2003), Ai and Chen (2003) or Blundell, Chen, and Kristensen (2004) and treat $k$ as a function of $N$. Since the choice of instruments is application specific, we discuss potential instruments in the next section

[21] Alternatively, we could assume that $\epsilon$ reflects the fact that $p_l$ is measured with error.

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

Before we proceed with our applications, we offer three additional observations. First, 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.[23] Second, 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.[24] 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.[25] With the bandwidth set equal to the standard deviation of $v$, we find in our housing application that the derivative conditions are met.[26] We thus conclude, that, at least in the housing application studied below, the unrestricted kernel estimator satisfies the restrictions that economic theory imposes on $r(v)$. Finally, we also explore some simple parametric specification for estimating (37), such as log-linear models, and investigate whether these simple parametric forms capture the observed variation in the data.[27]

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

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

[25] 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).

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

[27] 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.[28] 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

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

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.

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).

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. 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 had access to transaction prices for properties purchased for new construction. 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.

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.

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.

The OLS estimates are based on the assumption that $E(\epsilon|v) = 0$. Since $v$ is determined in equilibrium, it may be correlated with the error term. Hence 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 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 output price data. In our housing application, the first-order condition for the translog indirect profit function is: $p_l/v = \alpha + \beta \ln(p_l)$.[29] 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

[29]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 output.

.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.[30]

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 0.144. 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 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

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

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.[31]

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).

Our research has some important implications for policy analysis. 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.[32]

[31] 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.

[32] 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.

## 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.[33] 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

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

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).

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.

### 4.3 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:[34]

$$m = v - p_l \tag{44}$$

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

$$p_q = \frac{v}{q(m)} \tag{45}$$

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.

[34] 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.

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:

$$P_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{v_i}{q(m_i)} \tag{46}$$

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 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 set 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

We have demonstrated how to estimate production functions when output prices and quantities 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.[35] We have shown that the observed variation in input prices and normalized output values 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

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

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 also has some broad implications for understanding the industrial organization of many industries. Much recent research in economics focuses on differences in 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 and quantities are not observed by the econometrician.

Other research in industrial organization has stressed the importance of product differentiation.[36] In this paper we have abstracted from these issues. In a differentiated product model, products are typically described by a vector of characteristics. One potential approach is then to assume that each characteristic is produced via a separate production function. In that case, the methods used in this paper can be fairly easily extended to estimate production functions for characteristics. Alternatively, one could model the joint production of all product specific characteristics.[37] An

[36]The pioneering early papers on characteristics models with horizontal product differentiation are due to Gorman (1980) and Lancaster (1966). Recent empirical studies include, for example, Breshnahan (1987), Berry (1994), Berry, Levinsohn, and Pakes (1995, 2004), Nevo (2001), Petrin (2002), Blundell, Browning, and Crawford. (2003) Ekeland, Heckman, and Nesheim (2004), and Bajari and Benkard (2005).

[37]This is similar to the problem of estimating a production function for a multi-product firm allowing for comple-

interesting question then is to investigate the extend to which our approach can be extended to such an environment when separate prices for individual characteristics are unobserved. Product differentiation also leads to reduced competition and deviations from simple marginal cost pricing. The approach approach used in this paper is based on the notion that markets are competitive and firms behave as price takers. More research is clearly needed to deal with imperfect competition and market power. However, we view the results reported in this paper as encouraging for future research that will help improve our understanding of production technologies, productivity, and technological progress.

mentarities in the production process as discussed, for example, in Christensen et al. (1973).

## References


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

Ai, C. and Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, *71*, 1795–1843.

Bajari, P. and Benkard, L. (2005). Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, *113*, 1239–77.

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.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, *63 (4)*, 841–890.

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market. *Journal of Political Economy*, *112*, 68–104.

Blundell, R., Browning, M., and Crawford., I. (2003). Nonparametric Engel Curves and Revealed Preferences. *Econometrica*, *71*, 205–240.

Blundell, R., Chen, X., and Kristensen, D. (2004). Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. Working Paper.

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

Breshnahan, T. (1987). Competition and Collusion in the American Auto Industry: The 1955 Price War. *Journal of Industrial Economics*, *35*, 457–482.

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

Burmeister, E. and Kuga., K. (1970). The Factor-Price Frontier Duality, and Joint Production. *The Review of Economic Studies*, *37*, 11–19.

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.

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

Ekeland, I., Heckman, J., and Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, *112(1)*, S60–S101.

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.

Fuss, M. and McFadden, D. (Eds.). (1978). *Production Economics: A Dual Approach to Theory and Applications*. North Holland. Amsterdam.

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.

Gorman, W. (1980). A Possible Procedure for Analyzing Quality Differentials in the Egg-market. *Review of Economic Studies*, *47*, 843–856.

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

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

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.

Lancaster, K. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*, 132–157.

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.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, *69(2)*, 307–342.

Newey, W. and Powell, J. (2003). Instrumental Variables Estimation for Nonparametric Models. *Econometrica*, *71*, 1565–78.

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.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy*, *110(4)*, 705–729.

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

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

Shepard, R. (1970). *The Theory of Cost and Production Functions*. Princeton University Press. Princeton.

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

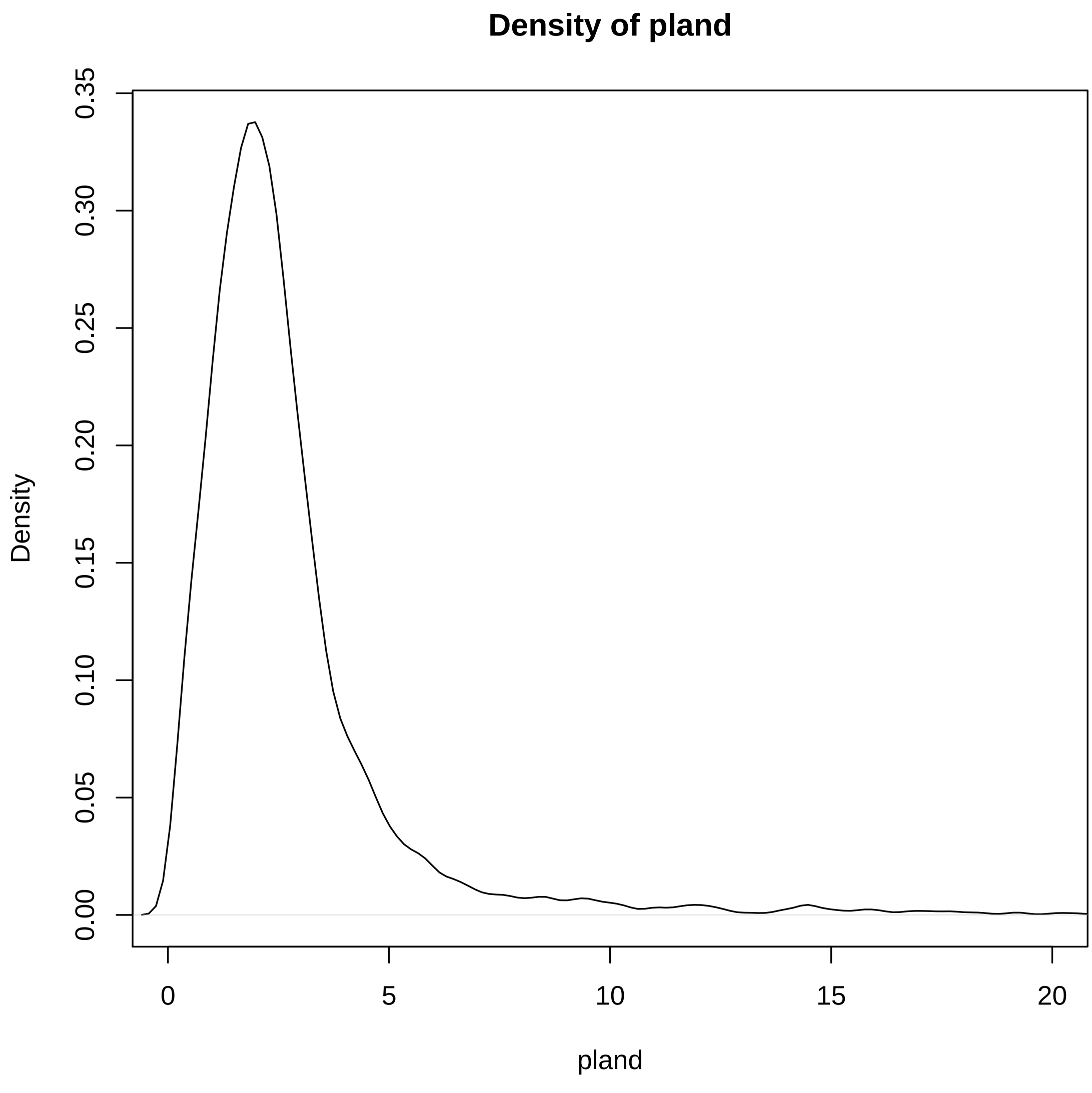


Figure 2: Density of Housing Values per Unit of Land

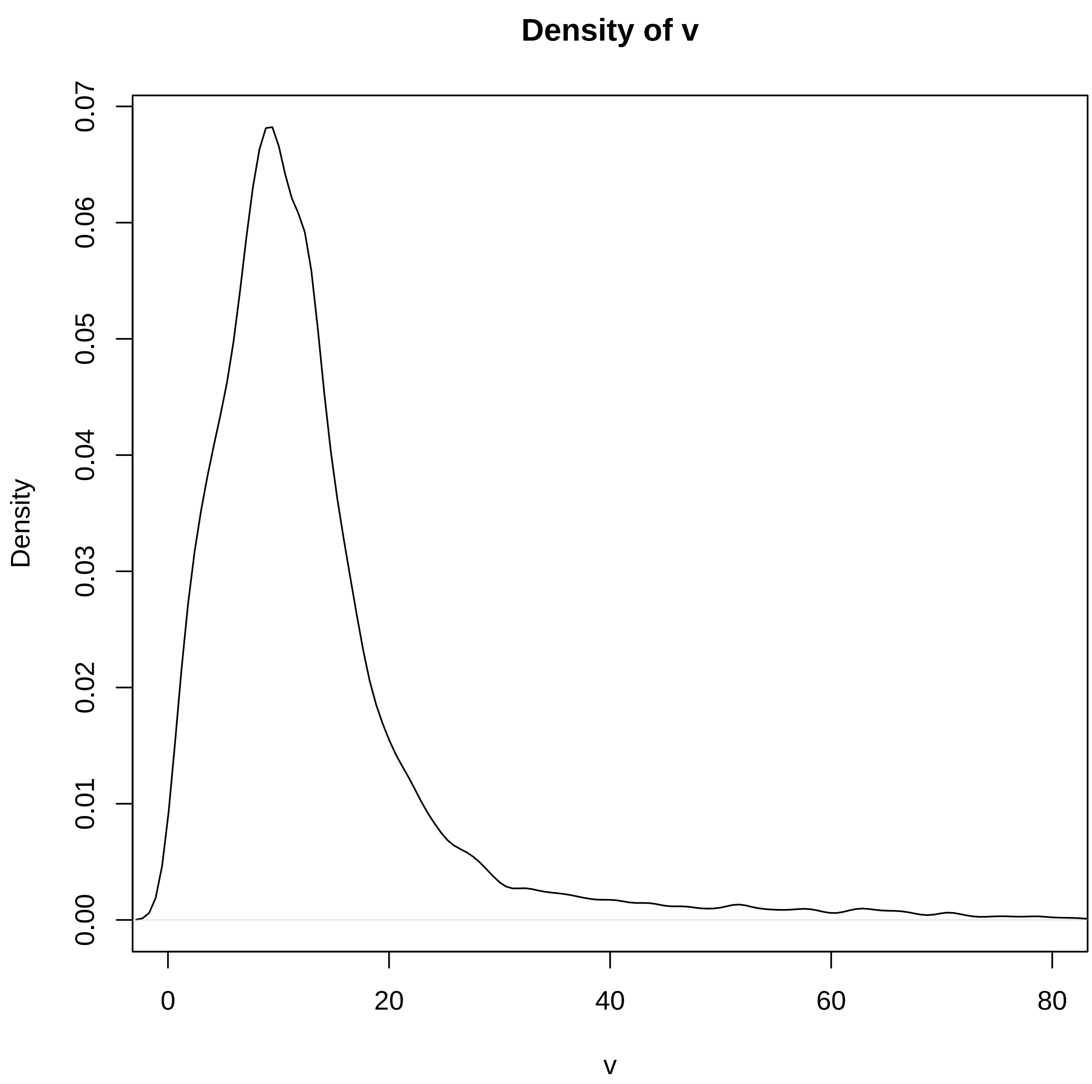


Figure 3: Map of Properties in the Data Set

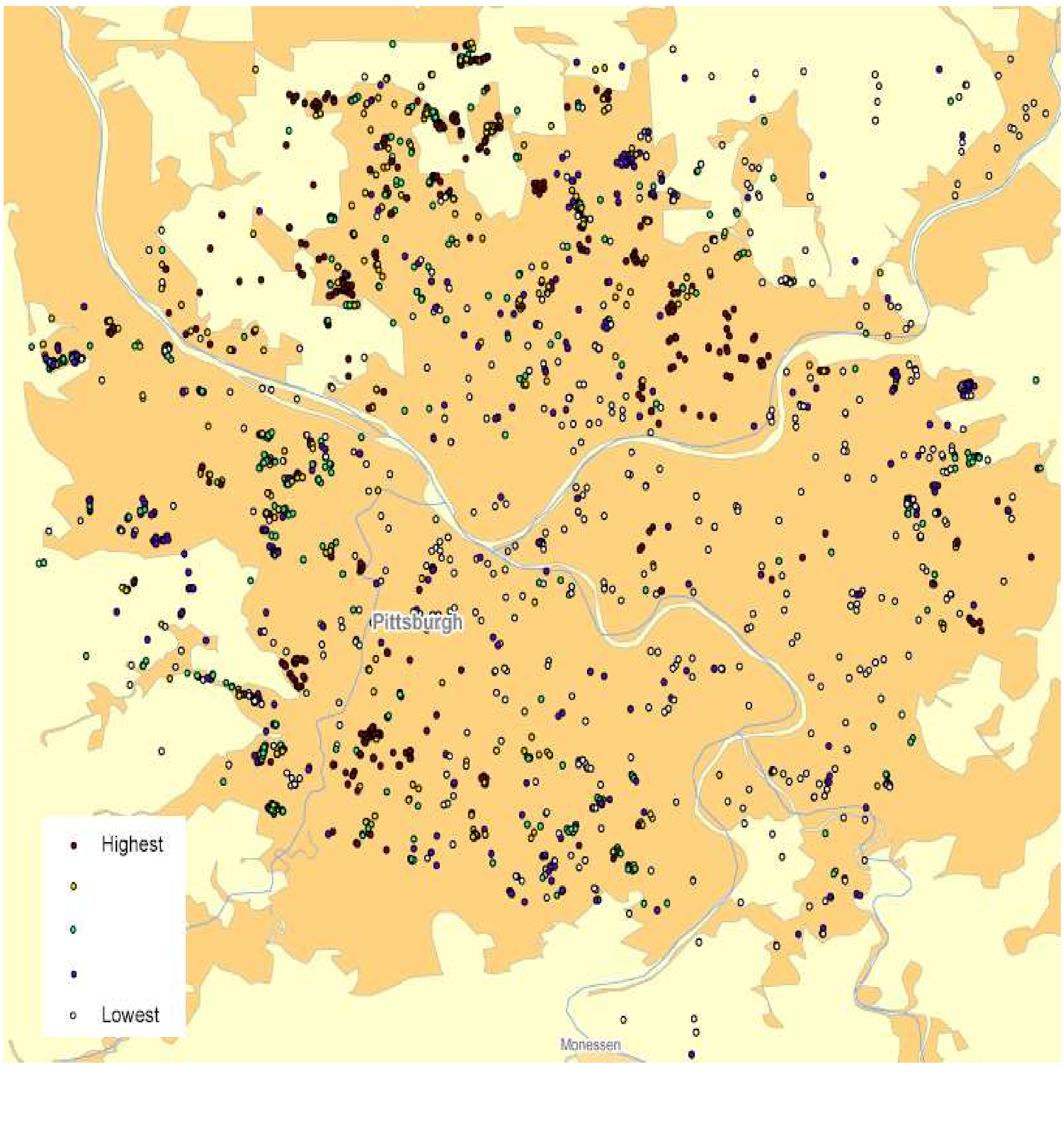


Figure 4:

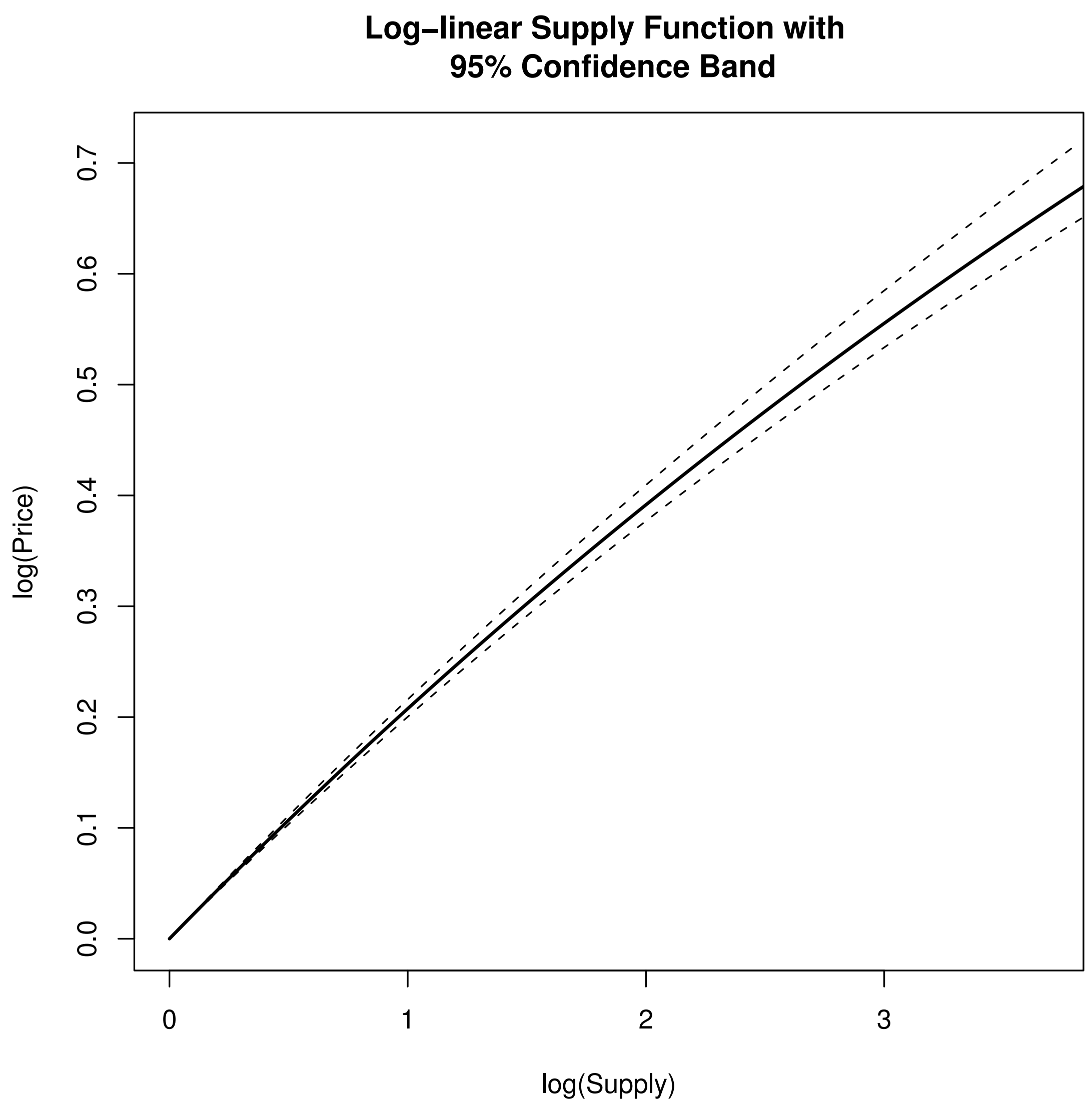
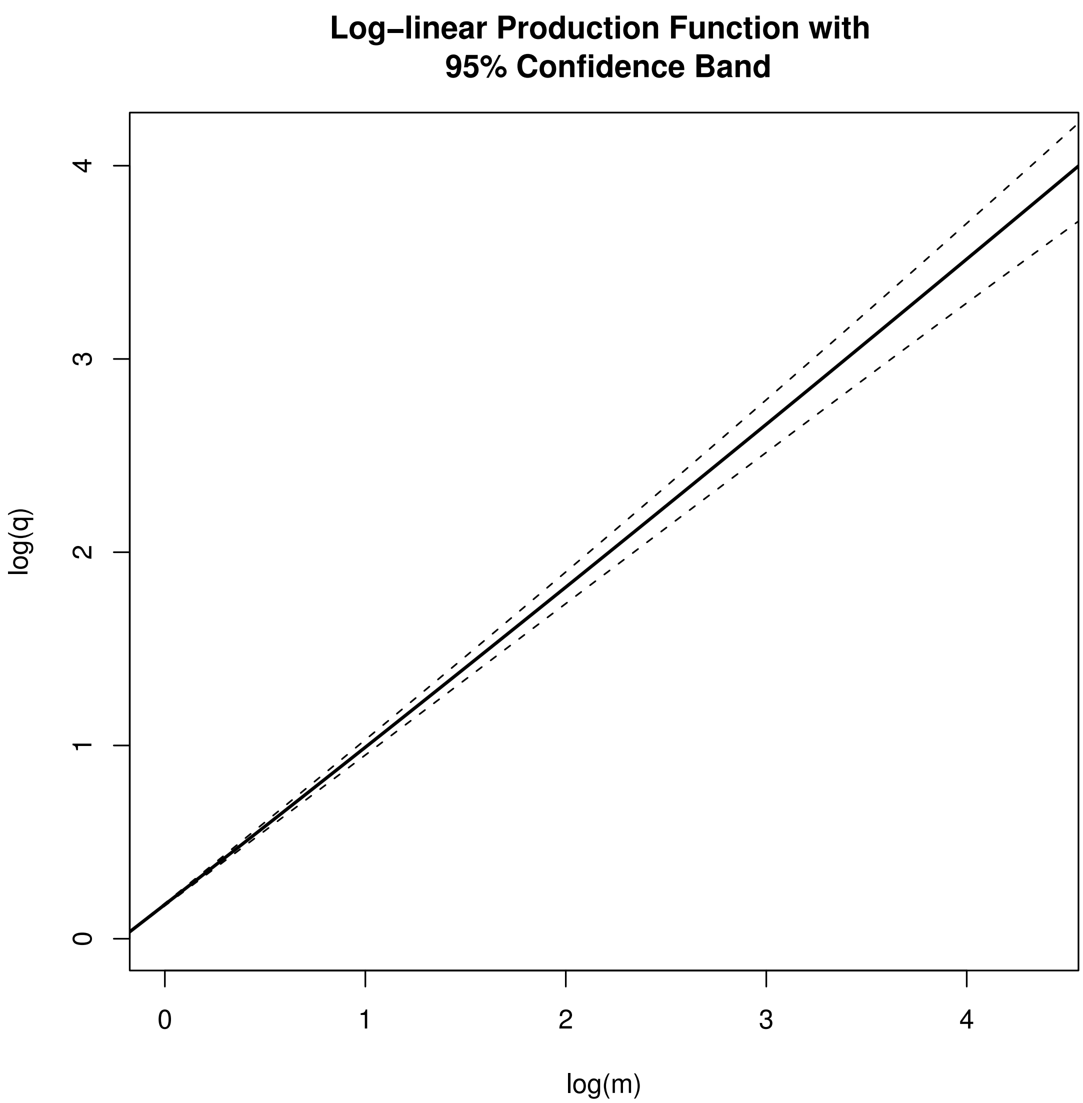
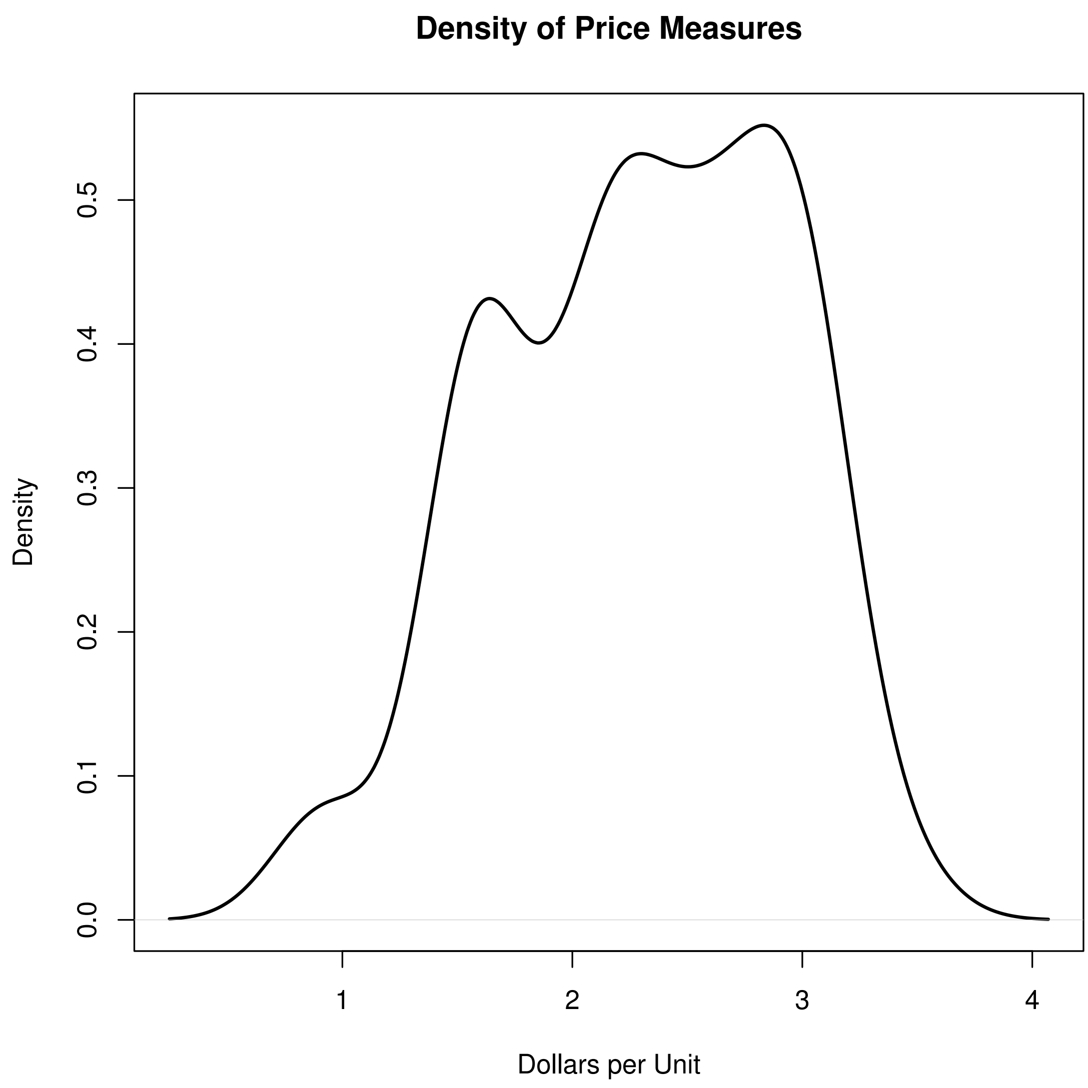

Log−linear Supply Function with
95% Confidence Band
log(Price)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
log(Supply)

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