Identifying Production Functions Using Restrictions from Economic Theory

Amit Gandhi, Salvador Navarro, David Rivers

University of Wisconsin-Madison

June 1, 2009

1 Introduction

As first pointed out by Marschak and Andrews [1944], using the inputs and outputs of profit maximizing firms to estimate production functions gives rise to an endogeneity problem. The endogeneity problem is caused by the presence of productive factors that are unobservable to the econometrician but that are "transmitted" to the firm's optimal choice of inputs. Traditionally, it is assumed that these unobservable factors are captured by a scalar productivity index that varies across firms and potentially evolves over time. The two standard and oldest methods of controlling for the endogeneity problem in the production function are instrumental variables and fixed effect estimation. However these solutions have proven unsatisfactory on both theoretical and empirical grounds (for a review, see Griliches and Mairesse [1998] and Ackerberg et al. [2007]).

Using an underlying model of firm dynamics, Olley and Pakes [1996] (OP for short) develop a new solution to the endogeneity problem. (This approach was extended by Levinsohn and Petrin [2003], LP for short, and Ackerberg et al. [2006], ACF for short.) Implicit in their approach are two main sets of assumptions: (1) timing assumptions on firm behavior, and

(2) an assumption that productivity is the only dimension of unobserved firm heterogeneity for which the econometrician cannot control. While (1) is a restriction from an economic model of firm behavior, (2) is purely an econometric assumption on the extent of unobserved heterogeneity in the data.

The main role of assumption (2), which we call the "scalar unobservability" assumption, is that it allows a firm's input demand function to be inverted to yield a firm's productivity as a function of only variables observable to the econometrician. The inverse of the firm's input demand thus acts as a perfect proxy, or *replacement function* [Heckman and Robb, 1985], for a firm's productivity. Conditioning on the replacement function thus controls for the endogeneity problem in the production function.

As implicitly recognized by Bond and Söderbom [2005] and Ackerberg et al. [2006], the replacement function cannot be used to identify the production function in the presence of a productive input that is both variable and static. This is an inherent limitation of scalar unobservability: it does not allow for a source of variation in such inputs to come from outside of the production function. As a consequence, the replacement function approach cannot identify gross output production functions when some inputs are variable and static.

While one solution is to focus on value-added production functions instead of gross output, value-added production functions are problematic for applied work as they rely on a number of strong restrictions (see e.g., Basu and Fernald [1997]). In general, when the assumptions of constant return to scale and perfect competition are violated, using the value-added production function to recover productivity is no longer a valid option.[1] Perhaps more importantly, gross output production functions are required to to study a number of important empirical problems, such as the problem of revenue production functions (when the econometrician does not observe quantities but rather only revenue), or analyzing productivity among exporting firms [Rivers, 2009].[2]

[1] One exception is the special case of Cobb-Douglas production functions, which is robust to non-constant returns to scale and specific forms of imperfect competition.

[2] To recover productivity when revenues come from multiple markets, such as export markets, value added presents many obstacles.

In this paper, we show that the firm's optimization problem (either maximizing profits or minimizing costs) contains enough identifying information to identify gross output production functions, both with and without the scalar unobservability assumption. Through a suitable transformation of the firm's first order condition (which we refer to in the paper as the inverse share equation for reasons that will be made apparent), we are able to exploit this information. A key feature of the transformation we exploit, i.e., the inverse share equation, is that it does not suffer from the endogeneity problem that plagues not only the production function, but also traditional input demand equations. As such, it provides a fundamentally new source of information relative to the replacement function.

Since the inverse share equation can be derived without appeal to the scalar unobservability assumption, and it does not require that the econometrician have information on variables other than those used in the replacement function, it allows us to introduce additional dimensions of unobserved heterogeneity among firms in the data besides productivity, such as firm specific demand shocks. As empirical studies have shown, heterogeneity is an important feature of firm-level data (see e.g., Bartelsman and Doms [2000], Fox and Smeets [2007]), and consequently allowing for multiple dimensions of firm heterogeneity is critical to the applicability of productivity analysis. In addition to being advantageous for its own sake, heterogeneity is a naturally useful source for identification, especially in the context of gross output production functions, as it provides a source of independent variation in static and variable inputs.

The key problem with the replacement function approach is that, while firm-level heterogeneity beyond productivity is a useful source of identification, it interferes with the role of the replacement function as a way to control for the endogeneity problem (since the replacement function relies on scalar unobservability). Our inverse share equation breaks this tension and gives us a new way to address many recent problems in the application of production functions. In particular we show that the inverse share equation can be fruitfully combined with the replacement function to identify (gross output) revenue productions without requiring any information from a parametric demand system. If a parametric demand system is introduced, then (gross output) revenue productions can be identified with firms having heterogeneous degrees of market power. Both of these results are new to the literature.

The rest of the paper is organized as follows. In Section 2 we describe the timing of production decisions that are standard in the literature and review the current structural approaches for estimating production functions based on scalar unobservability. Section 3 introduces the inverse share equation, which we show provides more information than the replacement function. In section 4 we show how our method is robust to markets with imperfect competition, and in particular can be used to estimate the production function in the presence of revenue rather than quantity data without having to simultaneously estimate a parametric demand system (which is the current state of the art of the literature). Section 5 presents results from an application to data on Chilean manufacturing firms.

2 The Timing of Production Decisions

Let Y denote a firm's output and the vector (L, K, M) denote a firm's inputs, with L denoting labor, K denoting capital, and M denoting all intermediate inputs. We assume that productivity differences among firms are driven by heterogeneity of a Hicks-neutral form. The relationship between a firm j's input and output in period t is expressed as

$$Y_{jt} = U_{jt} F(L_{jt}, K_{jt}, M_{jt}), \tag{1}$$

where U_{jt} represents the Hick's neutral shock.

Expressed in logs rather than levels (with lowercase letters denoting logs and uppercase letters denoting levels), (1) becomes

$$y_{jt} = \ln F(L_{jt}, K_{jt}, M_{jt}) + u_{jt}.$$

As we will show in the next sections, the central problem in estimating the production function F within the OP/LP/ACF tradition is the fact that the residual u_{jt} is assumed to consist of two distinct terms, i.e., $u_{jt} = \omega_{jt} + \varepsilon_{jt}$. The first term, ω_{jt}, is persistent over time and is observed before the firm makes its period t input decisions. The second term, ε_{jt}, is an unanticipated shock that the firm does not observe before making its period t input decisions. Since measurement error in output can also enter ε_{jt} we will simply refer to this term as measurement error. As ω_{jt} is anticipated but ε_{jt} is not, the endogeneity problem is generated due to the dependence between ω_{jt} and the inputs (L_{jt}, K_{jt}, M_{jt}), and we will refer to ω_{jt} as the firm's productivity.

In period t, firm j takes its productivity ω_{jt} and capital stock K_{jt} as state variables that are fixed for the period. Productivity evolves according to an exogenous Markov process $\Pr\left(\omega_{jt}|\{\omega_{j\tau}\}_{\tau=1}^{t-1}\right)$, which is assumed throughout the paper to be first order (an assumption which can be easily relaxed). Capital is accumulated each period based on last period's investment decision I_{jt-1}, and thus $K_{jt} = \mathbf{K}(K_{jt-1}, I_{jt-1})$. Intermediate inputs M_{jt}, on the other hand, are static and variable inputs. They are variable in the sense that they can be adjusted each period, and they are static in the sense that they have no dynamic implications, i.e., M_{jt} does not directly affect the firm's profit in any future periods.

Regarding labor, the usual assumption used in OP and LP is that L_{jt} is also a static and variable input, and thus treated exactly like the intermediate input from the firm's point of view. While ACF also assume L_{jt} to be static and variable, they assume that it is chosen at some point after K_{jt} is chosen but before ω_{jt} is fully realized (that is, L_{jt} is chosen between t and $t-1$). As a consequence, labor becomes a state variable at time t. Since this timing assumption on labor breaks collinearity concerns between labor and materials (a point we explain later), we will also maintain it. Instead of being static, the presence of hiring/firing costs would cause labor to have dynamic implications, and thus lagged labor L_{jt-1} would enter as a state variable to the firm in the period t labor decision. Our discussion applies equally to the dynamic or static interpretation of labor.

The first thing we observe about the timing of the model is that, if the measurement error ε_{jt} were excluded from the production function, then the timing assumptions on firm behavior by themselves secure identification. To see this, express the production function as

$$y_{jt} = \ln F(L_{jt}, K_{jt}, M_{jt}) + \omega_{jt},$$

and let the first-order Markov process on ω_{jt} be expressed as $\omega_{jt} = g(\omega_{jt-1}) + \eta_{jt}$. The term η_{jt} represents an innovation to the firm's productivity that, by construction, is orthogonal to the firm's information set at period $t-1$. For simplicity, assume that the functions F and g are indexed by a finite-dimensional parameter vector β. Then we can solve for each firm j's period t innovation $\eta_{jt}(\beta)$ as a function of model parameters:

$$\eta_{jt}(\beta) = y_{jt} - \ln F(L_{jt}, K_{jt}, M_{jt}) - g[y_{jt-1} - \ln F(L_{jt-1}, K_{jt-1}, M_{jt-1})].$$

The parameters β can then be recovered using the the moment conditions implied by the fact that the firm's innovation η_{jt} is orthogonal to the firm's information set at $t-1$, i.e., $\eta_{jt} \perp K_{jt}$, $\eta_{jt} \perp L_{jt-1}$, $\eta_{jt} \perp M_{jt-1}$, $\eta_{jt} \perp K_{jt-1}$, etc.

Once measurement error, ε_{jt}, is introduced back into the production function, the above procedure breaks down. The econometrician can only recover the value of the sum $[\omega_{jt} + \varepsilon_{jt}](\beta)$ as a function of the parameter vector. However this signal extraction problem, i.e., the problem of inferring the Markov process on ω_{jt} from noisy observations on the sum $\omega_{jt} + \varepsilon_{jt}$, cannot generically, be solved unless further restrictions are invoked.

In the next subsection we present the assumptions that enable OP/LP/ACF to use a replacement function to solve this signal extraction problem. In Section 3 we introduce our approach to production function estimation that combines information from the inverse share equation with other sources of information, including the production function itself, We show that this strategy also solves the signal extraction problem and we compare it with the replacement function.

2.1 The OP/LP/ACF Approach

Consider a prototypical firm's static profit maximization problem. Given the state variables $(L_{jt}, K_{jt}, \omega_{jt})$ (recall labor is assumed to be already set at time t by the timing assumptions), the firm has a profit function at time t of the form

$$\pi(L_{jt}, K_{jt}, \omega_{jt}, \Delta_{jt}) = \max_{M_{jt}} \pi(L_{jt}, K_{jt}, M_{jt}, \omega_{jt}, \Delta_{jt}), \tag{2}$$

where Δ_{jt} is a (potentially vector-valued) state variable that captures all other aspects of the firm's economic environment, e.g., input and output prices, demand shocks, adjustment costs, idiosyncratic technological differences, etc.

In order to solve the signal extraction problem in the general model of a profit maximizing firm presented above, OP/LP/ACF impose the additional assumption that $\Delta_{jt} = \Delta_t$ which we call the "scalar unobservability" assumption. That is, from the econometrician's perspective, the only unobserved state variable that is heterogeneous at the firm level is ω_{jt}.

As recognized by Bond and Söderbom [2005] and Ackerberg et al. [2006], this assumption comes at the cost of requiring that the intermediate input M_{jt} be excluded from the production function. To see why, recall that the intermediate input M_{jt} is a static and variable input, and hence is a function of the firm's state at time t, which under scalar unobservability implies the functional relationship $M_{jt} = M_t(L_{jt}, K_{jt}, \omega_{jt})$. Thus in the production function, M_{jt} is clearly an endogenous variable as it is partly a function of the econometric error term ω_{jt}. However there is no source of cross sectional variation in M_{jt} other than the firm's remaining productive inputs $(L_{jt}, K_{jt}, \omega_{jt})$, and hence no exclusion restriction to vary the intermediate input from outside of the production function. Without such a source of variation, which we refer to as the "collinearity problem" to follow the language of ACF, the production function cannot generically be identified in the presence of M_{jt}. Thus, scalar unobservability requires the researcher to net out intermediate inputs M_{jt} from the production function (which is a potentially perilous task, a point discussed further below), and measure output as value added $\tilde{y}_{jt}$,

Given value added $\tilde{y}_{jt}$, the advantage of scalar unobservability is that it allows a firm's demand for input i (be it static or dynamic) to be expressed as $i_{jt} = d(L_{jt}, K_{jt}, \omega_{jt}, \Delta_t) = d_t(L_{jt}, K_{jt}, \omega_{jt})$ where d_t is a strictly increasing function of ω_{jt} under suitable regularity conditions.[3] This implies that the input demand equation can be inverted to yield $\omega_{jt} = d_t^{-1}(L_{jt}, K_{jt}, i_{jt})$, and the inversion mapping $\omega_{jt} = d_t^{-1}(L_{jt}, K_{jt}, i_{jt})$ can then be replaced into the production function to yield

$$\begin{aligned} \tilde{y}_{jt} &= \ln F(L_{jt}, K_{jt}) + d_t^{-1}(L_{jt}, K_{jt}, i_{jt}) + \varepsilon_{jt} \\ &= \Phi_t(L_{jt}, K_{jt}, i_{jt}) + \varepsilon_{jt}. \end{aligned} \tag{3}$$

By running a time-varying nonparametric regression of $\tilde{y}_{jt}$ on (L_{jt}, K_{jt}, i_{jt}) the measurement error ε_{jt} can be recovered from equation (3). Observe that this regression does not suffer from collinearity concerns since ω_{jt} acts as a source of variation in i_{jt} independent of (L_{jt}, K_{jt}) (recall L_{jt} was chosen in between period t and $t-1$ (say at time $t-b$), and K_{jt} was chosen at $t-1$). Since ε_{jt} is independent of (L_{jt}, K_{jt}, i_{jt}), Φ_t is identified. Since the input i_{jt} is being used to proxy for productivity ω_{jt} in (3), we will refer to d_t at times as the proxy demand equation.

Recovering ε_{jt} from equation (3) is the main econometric role of the OP/LP/ACF first stage (a point recognized by ACF), which allows one to recover $\Phi_{jt} = \Phi_t(L_{jt}, K_{jt}, i_{jt}) = \ln F(L_{jt}, K_{jt}) + \omega_{jt}$. Once Φ_{jt} is recovered, the underlying timing assumptions can be implemented without a signal extraction problem (since the measurement error ε_{jt} has already been recovered and subtracted out from the model in the first stage). That is, we can write each firm j's period t innovation $\eta_{jt}(\beta)$ as a function of model parameters:

$$\eta_{jt}(\beta) = \Phi_{jt} - \ln F(L_{jt}, K_{jt}) - g[\Phi_{jt-1} - \ln F(K_{jt-1}, L_{jt-1})]$$

[3] OP use investment as the input, whereas LP use the intermediate input.

and form moments of it with the firm's information set at $t-1$, as was shown in the previous section.

To summarize, implementing the OP/LP/ACF replacement function requires the assumption that productivity, ω_{jt}, is the only source of unobserved heterogeneity across firms. This comes at the cost of requiring the researcher to assume that one can express the value-added form of the production function. In general, when the assumptions of constant returns to scale and perfect competition are violated, using the value-added production function to recover productivity is no longer valid (see e.g., Basu and Fernald [1997]). Furthermore, for a range of applications that include revenue production functions, and the production decisions of exporting firms, value-added production functions are not the primitive of interest even if one is willing to assume that value added is well defined.

Under the replacement function approach of OP/LP/ACF, any attempt to add back intermediate inputs M_{jt} into the production functions requires either additional information, or an independent source of variation ζ_{jt} that comes from outside of the production function and shifts around M_{jt}, i.e., $M_{,jt} = M_t(L_{jt}, K_{jt}, \omega_{jt}, \zeta_{jt})$. However, the existence of any such source of variation violates the scalar unobservability assumption as ζ_{jt} would also affect the demand for the invertible input i_{jt}, i.e., $i_{jt} = d_t(L_{jt}, K_{jt}, \omega_{jt}, \zeta_{jt})$, thereby invalidating the replacement function.

3 The Inverse Share Equation

We now show that the apparent tension between identifying gross output production functions and controlling for the endogeneity problem can be resolved by exploiting information not used in the replacement function approach, but that is nevertheless contained in the firm's problem given by equation (2). We take advantage of this information by transforming the firm's FOC in such a way that it that both contains information on the underlying production function and does not suffer from an endogeneity problem. This transformation, which we refer to as the inverse share equation for reasons that will be made apparent, allows for identification of gross output production functions both with and without scalar unobservability. That is, we can use the inverse share equation to generalize scalar unobservability and allow firms to be heterogeneous in dimensions other than productivity.

We begin, however, by maintaining the assumption that all unobserved heterogeneity across firms is captured by productivity ω_{jt}, and show how to identify a gross output production function In order to explain the mechanics of the inverse share equation, we first assume that firms behave as price takers in the product market.[4]In the next section, we generalize the inverse share equation to allow for imperfect competition among firms.

Recall the econometric form of the production function,

$$Y_{jt} = e^{\omega_{jt}} e^{\varepsilon_{jt}} F(L_{jt}, K_{jt}, M_{jt}).$$

For any choice of inputs, however, the firm expects a quantity of output

$$Q_{jt} = e^{\omega_{jt}} F(L_{jt}, K_{jt}, M_{jt}). \tag{4}$$

Let ρ_{jt} equal the intermediate input price and P_{jt} equal the output price. The first-order condition with respect to M yields,

$$P_{jt} F_M(L_{jt}, K_{jt}, M_{jt}) e^{\omega_{jt}} = \rho_{jt}. \tag{5}$$

Multiplying the LHS of (9) by $\frac{F(L_{jt},K_{jt},M_{jt})}{F(L_{jt},K_{jt},M_{jt})}$ and using the definition of Q_{jt} in (4) gives

$$P_{jt} Q_{jt} \frac{F_M(L_{jt}, K_{jt}, M_{jt})}{F(L_{jt}, K_{jt}, M_{jt})} = \rho_{jt}.$$

[4]In the standard application of OP/LP/ACF to plant level data, output is measured by deflating a firm's revenue by an industry-level deflator. This assumes that all firms in the industry have the same output prices and thus assumes away product differentiation, and as the asymptotics are taken as the number of firms in the cross section (i.e., market) grow large, perfect competition is implicitly being assumed.

Observe that the first-order condition has been transformed so that ω_{jt} no longer appears in it. The firm's productivity ω_{jt} has been subsumed in the profit maximizing output, Q_{jt}. Since $Y_{jt} = e^{\varepsilon_{jt}} Q_{jt}$, we can transform the FOC further to give

$$\frac{P_{jt} Y_{jt}}{\rho_{jt} M_{jt}} = \frac{F(L_{jt}, K_{jt}, M_{jt})}{F_M(L_{jt}, K_{jt}, M_{jt}) M_{jt}} e^{\varepsilon_{jt}}.$$

Finally letting $S_{jt} = \frac{P_{jt} Y_{jt}}{\rho_{jt} M_{jt}}$ denote the intermediate input's inverse revenue share, the transformed FOC in logs is

$$s_{jt} = \ln F(L_{jt}, K_{jt}, M_{jt}) - \ln F_M(L_{jt}, K_{jt}, M_{jt}) - m_{jt} + \varepsilon_{jt} \tag{6}$$

We refer to equation (6) as the inverse share equation.

The first thing to notice about the inverse share equation (6) is that, when viewed as a regression of s_{jt} on (L_{jt}, K_{jt}, M_{jt}), there is no endogeneity problem present. This is due to the fact that the transformation of the FOC used above has incorporated ω_{jt} into the dependent variable s_{jt}, which is an observable variable in the data since both the nominal intermediate input bill and nominal revenue are both primitives in plant level data. Moreover, the regression is identified as ω_{jt} acts as an independent source of variation in M_{jt} separate from (L_{jt}, K_{jt}) (by the timing assumptions and the fact that ω_{jt} does not appear on the RHS of (6)).

The second thing to notice is that the inverse share equation contains more information than the replacement function from the first stage of the OP/LP/ACF procedure. To see why, observe that we could in principle treat (6) nonparametrically as

$$s_{jt} = \phi(L_{jt}, K_{jt}, M_{jt}) + \varepsilon_{jt}.$$

Thus a nonparametric regression of s_{jt} on (L_{jt}, K_{jt}, M_{jt}) identifies ε_{jt}. However, as discussed in Section 2.1, recovering ε_{jt} is the main econometric function of the replacement function.

Thus if we treated ϕ nonparametrically, the inverse share equation would be as informative as the replacement function. However, by treating (6) purely as a nonparametric regression, we are ignoring information. Notice that ϕ is actually a known function (a partial differential equation) of the production function F itself, and thus the regression (6) can directly identify features of F in addition to "just" recovering ε_{jt}.

Take, for example, the case of a Cobb-Douglas production function (in logs):

$$y_{jt} = \alpha_l l_{jt} + \alpha_k k_{jt} + \alpha_m m_{jt} + \omega_{jt} + \varepsilon_{jt}.$$

In this case, m_{jt} can be written as a linear function of $l_{jt}, k_{jt}, \omega_{jt}$ and the lack of identification of the gross output production function is most apparent because m_{jt} literally drops out of the production function. Thus α_m cannot be identified using the OP/LP/ACF replacement function. However, when one writes the inverse share equation

$$s_{jt} = -\ln \alpha_m + \varepsilon_{jt}$$

it is obvious that one can first recover α_m from it. Notice that, by using the inverse share equation (as opposed to, for example, the conditional demand for M_{jt}) the estimating equation does not contain an endogeneity problem due to the presence of ω_{jt} on the right hand side.[5]

The general form of the inverse share equation (i.e., equation (6)) makes it apparent that it can be used as a source of identifying information for any production function F (not only Cobb-Douglas). It is this additional information that restores the ability to work

[5]Motivated by concerns other than endogeneity in the production function, the use of shares in the Cobb-Douglas case dates to (at least) Solow's 1957 growth accounting exercise. While different variations of the idea exist, typically the goal of using shares is to calibrate all of the parameters of the production function (sometimes with the ultimate goal of recovering productivity) using only the information contained in the shares. These index methods require imposing (sometimes implicitly) potentially strong assumptions, including assumptions about returns to scale and/or perfect competition and/or assumptions about the nature of the approximation errors introduced when computing parameters this way. In contrast, as we show, the inverse share equation can be used to structurally estimate the production function and relax a number of these strong assumptions.

with gross output production functions even under the scalar unobservability assumption. This feature, that the FOC contains information about the production function itself, is sometimes exploited via the estimation of input demands arising from the FOC (see for example Hamermesh [1993] for a basic reference and Doraszelski and Jaumandreu [2006] for a recent application). On the other hand, the use of the inverse share transformation that we propose in this paper is free of the endogeneity problem present in both the conditional input demand function for M_{jt} and the production function itself. Furthermore, the inverse share equation does not require the analyst to observe input and output prices (observing prices is required to estimate input demand functions).[6] Finally, the inverse share transformation gives rise to a closed-form estimating equation (6) that is applicable to any production function F as opposed to the input demands which, in a number of important cases, can only be defined implicitly (as occurs for example with a CES technology).

To see the more general applicability of the inverse share equation outside of Cobb-Douglas, consider, for example, a CES technology

$$Y_{jt} = e^{\omega_{jt}} e^{\varepsilon_{jt}} \left(\alpha_l L_{jt}^{\delta} + \alpha_k K_{jt}^{\delta} + \alpha_m M_{jt}^{\delta} \right)^{\frac{r}{\delta}} \tag{7}$$

and the log inverse share equation it generates:

$$s_{jt} = -\ln\left(\alpha_m r\right) - \delta m_{jt} + \ln\left(\alpha_l L_{jt}^{\delta} + \alpha_k K_{jt}^{\delta} + \alpha_m M_{jt}^{\delta} \right) + \varepsilon_{jt}. \tag{8}$$

Notice that, in principle, one can identify all of the parameters of a CES production function from the inverse share equation alone. The fact that M_{jt} is a function of $(L_{jt}, K_{jt}, \omega_{jt})$ (i.e., the collinearity problem in the production function) is not a problem in the inverse share equation because ω_{jt} is not on the right hand side of the inverse share equation, and as such

[6] If such observable price variation existed and could be observed by the analyst, then controlling for such prices in the proxy demand equation would allow for identification of gross output production functions, as prices act as an independent source of variation. The problem is that input and output prices are usually not reliably observed - if they were, then IV estimation of the production function becomes possible.

it is a valid source of variation to identify the coefficient on M_{jt}.

As a final example,[7] consider the case of the translog production function:

$$\begin{aligned} q_{jt} &= \alpha_l l_{jt} + \alpha_k k_{jt} + \alpha_m m_{jt} + \beta_l l_{jt}^2 + \beta_k k_{jt}^2 + \beta_m m_{jt}^2 \\ &\quad + \gamma_1 l_{jt} k_{jt} + \gamma_2 l_{jt} m_{jt} + \gamma_3 k_{jt} m_{jt} + \omega_{jt} + \varepsilon_{jt}. \end{aligned}$$

In this case, the inverse share equation is

$$s_{jt} = -\ln\left(\alpha_m + 2\beta_m m_{jt} + \gamma_2 l_{jt} + \gamma_3 k_{jt}\right) + \varepsilon_{jt}.$$

As before, this equation does not suffer from endogeneity problems because ω_{jt} is not on the right hand side, and thus ω_{jt} can act as an independent source of variation in m_{jt}. Furthermore, the information contained in the inverse share equation is enough to identify all of the coefficients associated with M_{jt} in F, and so collinearity of the variable and static input m_{jt} is no longer a problem for the use of gross output production functions.

4 Imperfect Competition

We have motivated the inverse share equation on the basis of the assumption that firms act as price takers in the product market. We now show that the inverse share equation can also be used to think more generally of models of imperfect competition. As we show in this section, by progressively incorporating the information used previously in the literature along with the inverse share equation, we can allow for more general forms of imperfect competition (and/or weaken the requirements about the data available to the analyst) than the existing literature can do for each set of assumptions. To do so, we consider the firm's cost minimization problem.[8]

[7] The same approach works for other technologies like, for example, the generalized quadratic (also know as a Diewert technology).

[8] The cost minimization approach is appealing because the construction of the inverse share equation does not depend on assumptions about product market competition. Consequently, one can use the inverse share

Recall that Q is the firm's expected output and $Y = e^{\varepsilon}Q$ is output as observed by the econometrician. At time t, firm $j's$ problem is

$$\min_{M_{jt}} \rho_{jt} M_{jt} + \Lambda_{jt} \left(Q_{jt} - F\left(L_{jt}, K_{jt}, M_{jt} \right) e^{\omega_{jt}} \right),$$

where, from standard envelope arguments, it follows that the multiplier Λ_{jt} is the marginal cost of firm j at the optimum. The first-order condition with respect to M yields

$$\Lambda_{jt} F_M(L_{jt}, K_{jt}, M_{jt}) e^{\omega_{jt}} = \rho_{jt}. \tag{9}$$

Multiplying the LHS of (9) by $\frac{M_{jt} F(L_{jt}, K_{jt}, M_{jt})}{F(L_{jt}, K_{jt}, M_{jt})}$ and using the definition of Q_{jt} gives

$$\Lambda_{jt} Q_{jt} \frac{M_{jt} F_M(L_{jt}, K_{jt}, M_{jt})}{F(L_{jt}, K_{jt}, M_{jt})} = \rho_{jt} M_{jt}.$$

Observe that the first-order condition has been transformed so that ω_{jt} no longer appears in it (and is subsumed in Q_{jt}). Multiplying the LHS by $\frac{P_{jt}}{P_{jt}}$ and using the definition of Y_{jt}, we transform the FOC further to yield

$$s_{jt} = \theta_{jt} + \ln F\left(L_{jt}, K_{jt}, M_{jt} \right) - \ln F_M\left(L_{jt}, K_{jt}, M_{jt} \right) - m_{jt} + \varepsilon_{jt}, \tag{10}$$

where we have replaced the log of the markup, $\ln\left(\frac{P_{jt}}{\Lambda_{jt}}\right)$, with θ_{jt}. Observe that equation (11) nests the inverse share equation (6) that was derived under perfect competition, the only difference being the addition of the markup term θ_{jt}, which is 0 under perfect competition.

We now develop the use of the inverse share equation (10) to estimate production functions among imperfectly competitive firms. It is critical whether the analyst has access to direct data on quantities or only revenues as the measure of output of the plant. We will first proceed under the assumption that the analyst has access to quantity data, and then in

equation within the cost minimization approach to analyze non-profit maximizing firms and other forms of competition where the profit maximization assumption is not appropriate.

Section 4.3 consider the problem of revenue production functions. As we will see, quantity data allows us to be more flexible in allowing heterogeneity in market power amongst firms.

4.1 Breaking Scalar Unobservability: An illustration with Constant Markups

In order to fix ideas, we first show how to incorporate a simple restriction on imperfect competition that arises when firms have constant elasticity residual demand curves. This is the restriction that all firms have the same markup, i.e., $\theta_{jt} = \theta_t$. This restriction can be incorporated into both the OP/LP/ACF framework and the inverse share equation. We then show that while the inverse share equation remains valid in the presence of other sources of unobserved heterogeneity, even in this simple setting adding such heterogeneity breaks scalar unobservability and the replacement function approach. This basic fact, that the inverse share equation can handle sources of heterogeneity, is the driving force behind the extensions we develop in the remainder of the paper.

To understand why the OP/LP/ACF approach can allow for potentially time-varying markups that are constant across firms, simply notice that the proxy demand equation can be written as

$$\begin{aligned} i_{jt} &= \tilde{d}_t\left(L_{jt}, K_{jt}, \omega_{jt}, \theta_t\right) \\ &= d_t\left(L_{jt}, K_{jt}, \omega_{jt}\right). \end{aligned}$$

That is, since the nonparametric demand equation is allowed to be time-varying it absorbs the time-varying markups and the procedure is completely unchanged.

On the other hand, under the constant markups restriction, the inverse share equation is given by

$$s_{jt} = \theta_t + \ln F\left(L_{jt}, K_{jt}, M_{jt}\right) - \ln F_M\left(L_{jt}, K_{jt}, M_{jt}\right) - m_{jt} + \varepsilon_{jt}. \tag{11}$$

Notice that once we control for time-varying intercepts in equation (11), ε_{jt} can be recovered and used to solve the signal extraction problem exactly as in Section 3.

While both methods can allow for time varying markups, the fact that we can work with gross output while the OP/LP/ACF approach cannot remains. A natural way to escape the collinearity problem that forces them to work with value added is to introduce unobserved firm-level heterogeneity, other than ω_{jt}, that shifts input demand. However, this is the very form of heterogeneity ruled out by scalar unobservability. An additional advantage of the inverse share equation is that it allows for such additional heterogeneity to be exploited.

As an example, now consider introducing a firm-level demand shock (Ξ_{jt}), which is observable to the firm but not the econometrician. As the constant markup assumption is consistent with firms having identical constant elasticity residual demand curves in equilibrium, we take the demand shock Ξ_{jt} to be one that affects the level of demand but not its curvature. That is, letting D_{jt} denote firm j's residual demand curve in equilibrium at time t, we have $D_{jt}(p_{jt}) = \Xi_{jt} D(p_{jt}, \Delta_t)$, where D_t is an iso-elastic demand function.

Observe that the demand shock Ξ_{jt} will be a second unobservable (in addition to ω_{jt}) entering into the firm's period t input decisions. The input i_{jt} is now determined as $i_{jt} = \tilde{d}_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$ violating the scalar unobservability assumption, i.e., this function can no longer be inverted to express ω_{jt} as a function only of observables. The presence of Ξ_{jt} invalidates the application of the replacement function (absent additional assumptions[9]), although it is an inherently useful source of variation in the intermediate input M_{jt} that comes from outside of the production function. The inverse share, on the other hand, remains a valid transformation of the firm's FOC despite the presence of Ξ_{jt}. Indeed (11) could be treated nonparametrically as $s_{jt} = \phi_t(L_{jt}, K_{jt}, M_{jt}) + \varepsilon_{jt}$ to recover ε_{jt}. Therefore, $y_{jt} - \varepsilon_{jt} =$

[9] An alternative approach to solving the scalar unobservability problem is to use the idea of a bijection introduced in Ackerberg et al. [2007]. They propose that, if there exists a variable (such as advertising expenditures) which together with the firm's input used in the proxy demand equation, form a bijection for the two unobservables, then this system can be inverted to solve and control for both unobservables as a function of observables. One problem with this approach is that it requires that the researcher have access to such a variable in the data. Furthermore, to the best of our knowledge, no one has shown equilibrium conditions under which such a valid bijection exists.

$\log F(L_{jt}, K_{jt}, M_{jt}) + \omega_{jt}$ can be identified off the timing moments without concern for the collinearity problem as Ξ_{jt} is acting as the independent source of variation in M_{jt}.

An additional implication of equation (11) is that, by using the inverse share equation, the growth in the markups can be recovered without further assumptions. If the intercept of equation (11) can be identified, then the level of the markups can also be recovered. While it may seem that separating the markups from the intercept requires an arbitrary normalization, this is only because of the nonparametric way in which we present the problem. To understand why, in general, this is not a problem, consider the following examples.

In the case of a CES production function, the inverse share equation under symmetric differentiation is given by

$$\begin{aligned} s_{jt} &= \theta_t - \ln(\alpha_m r) - \delta m_{jt} + \ln\left(\alpha_l L_{jt}^{\delta} + \alpha_k K_{jt}^{\delta} + \alpha_m M_{jt}^{\delta}\right) + \varepsilon_{jt} \\ &= \mu_t - \delta m_{jt} + \ln\left(\alpha_l L_{jt}^{\delta} + \alpha_k K_{jt}^{\delta} + \alpha_m M_{jt}^{\delta}\right) + \varepsilon_{jt}, \end{aligned}$$

where $\mu_t = \theta_t - \ln(\alpha_m r)$. Adding the identifying information from the production function allows us to recover r separately from the markups θ_t. If the production function is translog instead, the inverse share equation alone is enough to identify the markups since we have

$$s_{jt} = \theta_t - \ln(\alpha_m + 2\beta_m m_{jt} + \gamma_2 l_{jt} + \gamma_3 k_{jt}) + \varepsilon_{jt}.$$

4.2 Firm-Specific Markups

We now extend the analysis from the previous section by continuing to assume that the analyst can observe quantities directly, but no longer assuming that all firms have the same markup. In the previous section we assumed that markups were constant as an illustration of how the inverse share equation can take advantage of unobserved firm level heterogeneity while at the same time avoid gaining a new econometric error term. However, because the inverse share equation allows us to break the scalar unobservability assumption, we can now combine the OP/LP/ACF idea of the proxy demand equation with the inverse share equation to relax the assumption that markups do not vary across firms. This is an important extension since, to the best of our knowledge, the previous literature has only been able to allow for common markups across firms.

In terms of our earlier example, we now let a firm's equilibrium residual demand function take the more flexible form $D_{jt} = D(p_{jt}, \Xi_{jt}, \Delta_t)$, and drop the iso-elastic requirement on D. Instead we assume that the demand shock Ξ_{jt} increases a firm's market power, i.e., controlling for a firm's cost, a firm that receives a higher Ξ_{jt} earns a higher markup θ_{jt}. We can formalize this restriction by assuming that the demand shock and supply shock (ω_{jt}, Ξ_{jt}) are the only two unobservable (to the econometrician) state variables at the firm level. That is, we relax scalar unobservability by introducing the second econometric unobservable Ξ_{jt}. Thus the markup can be expressed as $\theta_{jt} = \mathcal{F}_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$, and we use the restriction that $\mathcal{F}_t$ is increasing in Ξ_{jt}. Hence we can invert this function to yield $\Xi_{jt} = \mathcal{F}_t^{-1}(L_{jt}, K_{jt}, \omega_{jt}, \theta_{jt})$, where $\mathcal{F}_t^{-1}$ is increasing in θ_{jt}.

Moreover, the firm's proxy demand i_{jt} can be expressed as the outcome of the policy $\tilde{d}_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$ For the same reason that the proxy demand $\tilde{d}_t$ is increasing and hence invertible in ω_{jt} (high productivity today signals more profitability tomorrow all else being equal), we have that proxy demand $\tilde{d}_t$ is also increasing and hence invertible in the demand shock Ξ_{jt}. It follows that

$$i_{jt} = \tilde{d}_t(L_{jt}, K_{jt}, \omega_{jt}, \mathcal{F}_t^{-1}(L_{jt}, K_{jt}, \omega_{jt}, \theta_{jt})) = d_t(L_{jt}, K_{jt}, \omega_{jt}, \theta_{jt}), \tag{12}$$

where the function d_t is increasing and hence invertible in θ_{jt}.[10]

Now, combining the proxy demand equation (12) with the inverse share equation (10) and the production function, yields the following system of equations:

[10]There is a slight abuse of notation here in that previously d_t had only three arguments, whereas now it also includes θ_{jt} as a fourth agrument. We do this so that the primary function of interest in each section is always just d_t.

$$
\begin{aligned}
\ln\left(\frac{P_{jt}Y_{jt}}{\rho_{jt}M_{jt}}\right) &= \theta_{jt} + \ln F\left(L_{jt}, K_{jt}, M_{jt}\right) - \ln F_M\left(L_{jt}, K_{jt}, M_{jt}\right) - m_{jt} + \varepsilon_{jt}, \\
y_{jt} &= \ln F\left(L_{jt}, K_{jt}, M_{jt}\right) + \omega_{jt} + \varepsilon_{jt}, \\
i_{jt} &= d_t\left(L_{jt}, K_{jt}, \omega_{jt}, \theta_{jt}\right).
\end{aligned}
$$

Given that the proxy demand equation is an increasing function of the firm's markup, this ensures that this system of equations has a unique solution for the econometric unobservables: $\varepsilon_{jt}, \omega_{jt}, \theta_{jt}$. Solving for these econometric unobservables for any guess of the parameter vector (F, g, d_t) (where recall g denotes the first-order Markov process on ω_{jt}), we can then implement the timing restrictions on productivity (i.e., $\eta_{jt}(\beta) = \omega_{jt}\left(\beta\right) - g\left(\omega_{jt-1}\left(\beta\right)\right)$), use the orthogonality of ε_{jt} to the covariates, form moments and identify the model). Additionally (or alternatively) one can impose timing assumptions on the markups, say $\theta_{jt} = G\left(\theta_{jt-1}\right) + \zeta_{jt}$ and also form moments with respect to ζ_{jt} to aid in identification.

4.3 Revenue Production Functions: Constant Expected Markups

Revenue production functions arise because in practice we often do not observe output in quantities but instead only observe revenues. The typical solution in the literature is to deflate revenues using an industry-level deflator,

$$
r_{jt} - p_t = y_{jt} = \ln F\left(L_{jt}, K_{jt}, M_{jt}\right) + \omega_{jt} + \left(p_{jt} - p_t\right) + \varepsilon_{jt}. \tag{13}
$$

where r_{jt} and p_t denote revenue and the deflator respectively. As pointed out by Klette and Griliches [1996], this leads to an endogeneity problem, because inputs will likely be correlated with a firm's output price. A recent solution to this problem is to model the unobserved prices by specifying a demand system, typically a simple parametric CES demand system. For example, De Loecker [2007] uses a one parameter constant elasticity demand system:

$$Q_{jt} = Q_t \left(\frac{P_{jt}}{P_t} \right)^{\frac{1}{\eta}} \Xi_{jt} \Upsilon_{jt} \tag{14}$$

where Ξ_{jt} is an observed (to the firm) and potentially persistent demand shock, and Υ_{jt} is an independent and unobserved demand shock. This strategy, introduced by Klette and Griliches [1996], crucially depends on the ability of the researcher to form a valid quantity index Q_t (to use as a demand shifter) which, when paired with the production function, identifies the unique demand parameter η. Moreover, as we showed in Section 4.1, the presence of Ξ_{jt} breaks scalar unobservability.

In this section we show that, by combining the information given by the inverse share equation and the proxy demand equation of OP/LP/ACF (for example investment), we can control for unobserved prices without imposing the strong parametric assumption that iso-elastic demand curves take the particular functional form in (14). Rather we use the weaker restriction implicit in (14) that firms have constant markups in equilibrium. This approach offers two basic advantages over the literature to date. First, we can let the the elasticity of demand η vary over time since we don't rely on variation in the aggregate quantity index q_t to identify it. Second, and perhaps more importantly, we are able to control for other sources of unobserved heterogeneity (in addition to ω_{jt}), without having to bring in outside data or impose additional assumptions (that is, without having to rely on the high level assumption of a bijection between (ω_{jt}, Ξ_{jt}) and the two inputs (i^1_{jt}, i^2_{jt}), as discussed in Footnote 9).

We thus abstract from (14) and express demand as:

$$Q_{jt} = \Upsilon_{jt} \Xi_{jt} D(P_{jt}, \Delta_t). \tag{15}$$

where $D()$ is a constant elasticity demand system, $\Xi_{jt} > 0$ is the (potentially persistent) demand shock observed by the firm at the time inputs are being chosen, $\Upsilon_{jt} > 0$ is an independent demand shock that is realized after inputs are chosen, and Δ_t consists of market-level variables which affect demand, but do not vary across firms. In this case, *expected* *markups* will be equalized across firms, i.e., $E(\Theta_{jt}) = \Theta_t$ (where Θ_t is determined by the Lerner index and the elasticity of $D(\cdot, \Delta_t)$) and Ξ_{jt} will enter into the firm's period t input decisions. That is, while actual markups $\Theta_{jt} = \frac{P_{jt}}{\Lambda_{jt}}$ will be firm specific due to the Υ_{jt} demand shocks, firms will have ex-ante symmetric markups.

To see how we can use this information to estimate revenue production functions, observe that the expected demand function implies that, in equilibrium, the firm's expected price $E(P_{jt})$ can be expressed as $E(P_{jt}) = P_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$. Furthermore, by definition, $E(P_{jt}) = \Theta_t \Lambda_{jt}$ where Λ_{jt} is the firm's short-run marginal cost (at the firm's optimum in equilibrium) and Θ_t is the expected markup. Hence, we can write $\Lambda_{jt} = \tilde{P}_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$. As long as the firm's short-run marginal cost is increasing in output (as it will be so long as the production function is concave in M_{jt}) and increased levels of the demand shock Ξ_{jt} induce the firm to expand output, $\tilde{P}_t$ will be an increasing function of Ξ_{jt}. This monotonicity implies that $\tilde{P}_t$ can be inverted to yield $\Xi_{jt} = V_t(L_{jt}, K_{jt}, \omega_{jt}, \Lambda_{jt})$. The fact that we can express the demand unobservable as a function of the marginal cost implies that the firm's proxy demand equation $i_{jt} = \tilde{d}_t(L_{jt}, K_{jt}, \omega_{jt}, \Xi_{jt})$ can now be expressed as

$$\begin{aligned} i_{jt} &= \tilde{d}_t(L_{jt}, K_{jt}, \omega_{jt}, V_t(L_{jt}, K_{jt}, \omega_{jt}, \Lambda_{jt})) \\ &= d_t(L_{jt}, K_{jt}, \omega_{jt}, \Lambda_{jt}). \end{aligned}$$

As $\tilde{d}_t$ is increasing in Ξ_{jt}, then d_t is increasing in Λ_{jt}, and hence can be inverted to yield

$$\lambda_{jt} = \psi_t(L_{jt}, K_{jt}, i_{jt}, \omega_{jt}) \tag{16}$$

where ψ_t is *decreasing* in ω_{jt} and $\lambda_{jt} = \ln(\Lambda_{jt})$.

We can then rewrite the revenue production function as

$$\begin{aligned} r_{jt} &= \ln F(L_{jt}, K_{jt}, M_{jt}) + p_{jt} + \omega_{jt} + \varepsilon_{jt} \\ &= \ln F(L_{jt}, K_{jt}, M_{jt}) + \omega_{jt} + \theta_t + \lambda_{jt} + \omega_{jt} + (\varepsilon_{jt} + \upsilon_{jt}), \end{aligned} \tag{17}$$

where $\theta_t = \ln E(\Theta_{jt}) = \ln(\Theta_t)$ and $\upsilon_{jt} = \ln(\Upsilon_{jt})$. Finally, the inverse share equation can be rewritten as

$$\ln\left(\frac{P_{jt}Y_{jt}}{\rho_{jt}M_{jt}}\right) = \theta_t + \ln F(L_{jt}, K_{jt}, M_{jt}) - \ln F_M(L_{jt}, K_{jt}, M_{jt}) - m_{jt} + (\varepsilon_{jt} + \upsilon_{jt}). \quad (18)$$

Equations (16), (17) and (18) form a system of three equations in three econometric unobservables $(\omega_{jt}, (\varepsilon_{jt} + \upsilon_{jt}), \lambda_{jt})$. For any guess of the parameters of (F, g, ψ_t) and the expected markups θ_t, $(\varepsilon_{jt} + \upsilon_{jt})$ can be recovered directly from the inverse share equation. Then, from fact that ψ_t is decreasing in ω_{jt}, $(\omega_{jt}, \lambda_{jt})$ can be solved for uniquely from the other two equations. Since we can solve for all of the econometric unobservables, we can form moments as before to identify the model.

4.4 Revenue Production Functions: Firm-Specific Markups

Lastly, in this section we show that by using a general parametric form of the iso-elastic demand system, we can use the inverse share equation to let firms have heterogeneous markups even in the presence of revenue production functions. Let the inverse demand system be given by the following:

$$P_{jt} = \left(\frac{Q_{jt}}{Q_t}\right)^{\eta_{jt}} P_t$$

where η_{jt} is the inverse of the elasticity of demand, and $\frac{1}{1+\eta_{jt}}$ is the resulting markup.

In this case, the resulting system of equations is the following:

$$\begin{aligned}
\ln\left(\frac{P_{jt}Y_{jt}}{\rho_{jt}M_{jt}}\right) &= -\ln(1+\eta_{jt}) + \ln F(L_{jt}, K_{jt}, M_{jt}) - \ln F_M(L_{jt}, K_{jt}, M_{jt}) - m_{jt} + \varepsilon_{jt} \\
r_{jt} - p_t &= (1+\eta_{jt}) \ln F(L_{jt}, K_{jt}, M_{jt}) + (1+\eta_{jt})\omega_{jt} - \eta_{jt} q_t + \varepsilon_{jt} \\
i_{jt} &= d_t(L_{jt}, K_{jt}, \omega_{jt}, \eta_{jt}).
\end{aligned}$$

When a unique solution to this system of equations exists (which will be the case under standard monotonicity conditions and provided a bound condition in the data holds), one can solve for $(\varepsilon_{jt}, \omega_{jt}, \theta_{jt})$ and implement the timing restrictions on ω_{jt} and/or θ_{jt} to estimate the model.

5 Empirical Illustration

To illustrate our methodology, we apply a simple version of our empirical strategy based on the inverse share equation to the same Chilean manufacturing data used in Levinsohn and Petrin [2003]. In particular, the dataset we use is identical to that employed by Greenstreet [2007], which is constructed using the most precise industry deflators and capital variables (the appendix to Greenstreet [2007] contains an extensive description). We focus on the largest industry, namely 311 (food products). We employ a three input gross output production function in labor, capital and an aggregated intermediate input formed from materials, fuels, electricity, and services).

For our data, the output measure is constructed as deflated revenue. As this is a valid measure under the assumption that all firms in the data are competing in the same homogeneous goods industry, we use the assumption of price-taking behavior in the output market for consistency with the interpretation of deflated revenues as quantity. We can use the theory of the inverse share equation to reject Cobb-Douglas production functions under this specification - if we regress s_{jt} on (L_{jt}, K_{jt}, M_{jt}), an implication of Cobb-Douglas would be that only the intercept should predict s_{jt}. This is easily rejected, as the firm's share, s_{jt}, exhibits persistence over time.

We instead use the CES production function because it can explain the persistence in s_{jt}. In particular, we use the inverse share equation (8) along with the CES production function itself (7) to solve for the two econometric errors $(\omega_{jt}(\alpha_l, \alpha_k, \alpha_m, r, \delta), \varepsilon_{jt}(\alpha_l, \alpha_k, \alpha_m, r, \delta))$ as a function of the parameter vector $(\alpha_l, \alpha_k, \alpha_m, r, \delta)$. We then implement the timing moments

Table 1: CES Point Estimates

Parameter	Point Estimate
α_l	.816
α_k	.049
α_m	.134
δ	.406
r	.995
Avg. Labor Elas.	.216
Avg. Capital Elas.	.136
Avg. Intermediate Input Elas.	.643

to estimate the parameters of the production function. We take a dynamic view of labor and treat L_{jt} as a state variable in period t and hence orthogonal to the period t innovation η_{jt}. Note that the scale of α_l, α_k, and α_m is generically not identified separately from the mean of productivity. As a result, one normalization is necessary. We normalize them such that they add to one. The results are reported in Table 1.

Our estimates imply constant returns to scale ($r = .995$). The implied elasticity of substitution of 1.68 signifies a departure from Cobb-Douglas (which has an elasticity of substitution of 1), but not an extreme departure. Also note that, since the elasticity of output with respect to the inputs now depends on the level of the inputs, it is heterogeneous across observations in the sample. Therefore, we report the average elasticity among firms.

References

Daniel Ackerberg, Lanier Benkard, Steven Berry, and Ariel Pakes. Econometric tools for analyzing market outcomes. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 6. Elsevier Science, 2007. Forthcoming.

Daniel A. Ackerberg, Kevin Caves, and Garth Frazer. Structural identification of production functions. Unpublished Manuscript, UCLA Economics Department, 2006.

Eric J. Bartelsman and Mark Doms. Understanding productivity: lessons from longitudinal microdata. Finance and Economics Discussion Series 2000-19, Board of Governors of the Federal Reserve System (U.S.), 2000.

Susanto Basu and John G Fernald. Returns to scale in u.s. production: Estimates and implications. *Journal of Political Economy*, 105(2):249 – 283, April 1997.

Stephen Bond and Måns Söderbom. Adjustment costs and the identification of cobb douglas production functions. Unpublished Manuscript, The Institute for Fiscal Studies, Working Paper Series No. 05/4, 2005.

Jan De Loecker. Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity. NYU Stern, Unpublished Manuscript, 2007.

U. Doraszelski and J. Jaumandreu. R&d and productivity: Estimating production functions when productivity is endogenous. Technical working paper, Harvard University, 2006.

Jeremy Fox and Valerie Smeets. Do input quality and structural productivity estimates drive measured differences in firm productivity? 2007.

David Greenstreet. Exploiting sequential learning to estimate establishment-level productivity dynamics and decision rules. Economics Series Working Papers 345, University of Oxford, Department of Economics, 2007.

Zvi Griliches and Jacques Mairesse. Production functions: The search for identification. In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pages 169–203. Cambridge University Press, New York, 1998.

Daniel S. Hamermesh. *Labor Demand*. Princeton University Press, Princeton, NJ, 1993.

James J. Heckman and Richard Robb. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2):239–267, October-November 1985.

Tor Jacob Klette and Zvi Griliches. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics*, 11(4): 343–361, July 1996.

James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(243):317–342, April 2003.

Jacob Marschak and W.H. Andrews. Random simultaneous equations and the theory of production. *Econometrica*, 12:143–205, 1944.

G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, November 1996.

David A. Rivers. A structural approach to relating exporting and productivity. Working paper, University of Wisconsin-Madison, 2009.

Robert M. Solow. Technical change and the aggregate production function. *Review of Economics and Statistics*, 39(3):312–320, 1957.