Economics Letters 104 (2009) 112–114

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On estimating firm-level production functions using proxy variables to control for unobservables

Jeffrey M. Wooldridge *

Department of Economics, Michigan State University, 110 Marshall-Adams Hall, East Lansing, MI 48824-1038, United States

ARTICLE INFO

Article history:
Received 2 January 2008
Received in revised form 15 April 2009
Accepted 23 April 2009
Available online 6 May 2009

Keywords:
Production function
Panel data
Proxy variables
Generalized method of moments

JEL classification:
C23
C33

ABSTRACT

In the common case where polynomial approximations are used for unknown functions, I show how proxy variable approaches to controlling for unobserved productivity, proposed by Olley and Pakes [Olley, S. and Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1298.] and Levinsohn and Petrin (Levinsohn, J. and Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70, 317–341.], can be implemented by specifying different instruments for different equations and applying generalized method of moments. Studying the parameters within a two-equation system clarifies some key identification issues, and joint estimation of the parameters leads to simple inference and more efficient estimators.

1. Introduction

For estimating production functions using firm-level panel data, Olley and Pakes (1996) (OP for short) showed how, under certain assumptions, investment can be used as a proxy variable for unobserved, time-varying productivity. Specifically, OP show how to invert an investment rule to express productivity as an unknown function of capital and investment (when investment is strictly positive). OP present a two-step estimation method where, in the first stage, semiparametric methods are used to estimate the coefficients on the variable inputs. In a second step, the parameters on capital inputs can be identified under assumptions on the dynamics of the productivity process.

Levinsohn and Petrin (2003) (LP for short) propose a modification of the OP approach to address the problem of lumpy investment. LP suggest using intermediate inputs to proxy for unobserved productivity. Their paper contains assumptions under which productivity can be written as a function of capital inputs and intermediate inputs (such as materials and electricity). As with OP, LP propose a two-step estimation method to consistently estimate the coefficients on the variable inputs and the capital inputs.

In implementing the OP or LP approaches, it is convenient to assume that unknown functions are well approximated by low-order polynomials. Petrin et al. (2004) (PPL for short) suggest using third-degree polynomials, and LP note that such a choice leads to estimated parameters that are very similar to locally weighted estimation. Because PPL's paper includes Stata commands to implement LP's approach using third-degree polynomials, it seems likely that the polynomial approximation approach will gain favor among empirical economists.

Because of the complicated two-step nature of the LP estimation method, the authors suggest using bootstrapping methods to obtain standard errors and test statistics. In this note, I show that the moment conditions used by LP, as well as important extensions, can be easily implemented in a generalized method of moments (GMM) framework. The key is to write the moment restrictions in terms of two equations – with the same dependent variable – where the set of instruments differs across equation, as in Wooldridge (1996).

The GMM setup has several advantages over two-step approaches. First, Ackerberg et al. (2006) (ACF for short) have recently highlighted a potential problem with identification of the parameters in the LP first-stage estimation problem. Namely, if the variable input (labor), determined optimally by the firm, is also a deterministic function of unobserved productivity and state variables, then the coefficient on the variable input is nonparametrically unidentified. ACF show that specifying popular functional forms for the production process does not help. In fact, in the Cobb–Douglas case (and some others), labor disappears after substituting unobserved productivity as a function of inputs. ACF propose a hybrid of the OP and LP approaches, along with assumptions on the timing of decisions concerning input choice. ACF resolve the potential lack of identification by using a two-step

* Tel.: +1 517 353 5972; fax: +1 517 432 1068.
E-mail address: wooldri1@msu.edu.

0165-1765/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2009.04.026

estimation method that does not attempt to identify any production parameters in the first stage. My unified approach allows for the possibility that the first stage of OP or LP actually contains identifying information for parameters on the variable inputs, such as labor.

Another benefit of joint GMM estimation of the system is that fully robust standard errors are easy to obtain. Also, GMM efficiently uses the moment conditions implied by the OP and LP assumptions. Two-step estimators are inefficient for two reasons: (i) they ignore the contemporaneous correlation in the errors across two equations; and (ii) they do not efficiently account for serial correlation or heteroskedasticity in the errors. GMM uses the cross-equation correlation to enhance efficiency, and the optimal weighting matrix efficiently accounts for serial correlation and heteroskedasticity.

2. The model and assumptions

The point I make in this paper is fairly general, but I will use the language of the production function literature because that is the primary motivation. For firm i in time period t, write

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + v_{it} + e_{it},\ t = 1, \ldots, T, \tag{2.1}$$

where y_{it} is typically the natural logarithm of the firm's output, $\mathbf{w}_{it}$ is a $1\times J$ vector of variable inputs – such as labor – and $\mathbf{x}_{it}$ is a $1\times K$ vector of observed state variables – such as capital – all in logarithmic form. The sequence $\{v_{it}: t=1,\ldots, T\}$ is unobserved productivity, and $\{e_{it}: t=1, 2,\ldots, T\}$ is a sequence of shocks that, as we will see, are assumed to be conditional mean independent of current and past inputs.

A key implication of the theory underlying OP and LP is that for some function $g(\cdot,\cdot)$,

$$v_{it} = g(\mathbf{x}_{it}, \mathbf{m}_{it}),\ t = 1, \ldots, T, \tag{2.2}$$

where $\mathbf{m}_{it}$ is a $1\times M$ vector of proxy variables (investment in OP, intermediate inputs in LP). Initially, we assume that $g(\cdot,\cdot)$ is time invariant.

Under the assumption

$$E(e_{it}\,|\,\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}) = 0,\ t = 1, 2, \ldots, T, \tag{2.3}$$

we have the following regression function:

$$\begin{aligned} E(y_{it}\,|\,\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}) &= \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + g(\mathbf{x}_{it}, \mathbf{m}_{it}) \\ &\equiv \mathbf{w}_{it}\beta + h(\mathbf{x}_{it}, \mathbf{m}_{it}),\ t = 1, \ldots, T, \end{aligned} \tag{2.4}$$

where $h(\mathbf{x}_{it}, \mathbf{m}_{it}) \equiv \alpha + \mathbf{x}_{it}\gamma + g(\mathbf{x}_{it}, \mathbf{m}_{it})$. Since $g(\cdot,\cdot)$ is allowed to be a general function – in particular, linearity in $\mathbf{x}$ is a special case – γ (and the intercept, α) is clearly not identified from Eq. (2.4). Nevertheless – at least at first sight – Eq. (2.4) appears to identify β. However, this need not be true, especially if we believe the economics that leads to Eq. (2.2). Particularly problematical is when $\mathbf{m}_{it}$ contains intermediate inputs, as in LP. As shown by ACF, if labor inputs are chosen at the same time as intermediate inputs, there is a fundamental identification problem in Eq. (2.4): $\mathbf{w}_{it}$ is a deterministic function of $(\mathbf{x}_{it}, \mathbf{m}_{it})$, which means β is nonparametrically unidentified. To make matters worse, ACF show that $\mathbf{w}_{it}$ actually drops out of Eq. (2.4) when the production function is Cobb–Douglas.

As in OP and LP, I assume that estimation of γ is also important. In fact, the point of this note is to study the problem of estimating β and γ together. In order to identify γ along with β, I follow OP and LP and strengthen Eq. (2.3) to

$$\begin{aligned} &E\left(e_{it}\,|\,\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}, \mathbf{w}_{i,t-1}, \mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots, \mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}\right) \\ &\quad = 0,\ t = 1, 2, \ldots, T. \end{aligned} \tag{2.5}$$

Assumption (2.5) can be weakened somewhat – in particular, identification could hold with just current values and one lag in the conditioning set – but assuming conditional mean independence given outcomes at t and $t-1$, without also assuming Eq. (2.5), is ad hoc. Assumption (2.5) does allow for serial dependence in the idiosyncratic shocks $\{e_{it}: t=1, 2,\ldots, T\}$ because neither past values of y_{it} nor e_{it} appear in the conditioning set.

Finally, we use an assumption restricting the dynamics in the productivity process, $\{v_{it}: t=1, 2,\ldots\}$. LP state the assumption as

$$E\left(v_{it}\,|\,v_{i,t-1}, \ldots, v_{i1}\right) = E\left(v_{it}\,|\,v_{i,t-1}\right),\ t = 2, 3, \ldots, T, \tag{2.6}$$

along with an assumption that $\mathbf{x}_{it}$ is uncorrelated with the innovation

$$a_{it} \equiv v_{it} - E\left(v_{it}\,|\,v_{i,t-1}\right). \tag{2.7}$$

These assumptions are not quite enough. In the second stage of the LP procedure, the conditional expectation used to identify γ depends on $(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})$. Thus, consistency requires that a_{it} is additionally uncorrelated with $(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})$. A sufficient condition that meshes well with Eq. (2.5) is

$$\begin{aligned} &E\left(v_{it}\,|\,\mathbf{x}_{it}, \mathbf{w}_{i,t-1}\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots, \mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}\right) \\ &\quad = E\left(v_{it}\,|\,v_{i,t-1}\right) \equiv f\left[g\left(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}\right)\right], \end{aligned} \tag{2.8}$$

where the latter equivalence holds for some $f(\cdot)$ because $v_{i,t-1} = g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})$. An important point is that the variable inputs in $\mathbf{w}_{it}$ are allowed to be correlated with the innovations a_{it}, but Eq. (2.8) means that $\mathbf{x}_{it}$, past outcomes on $(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it})$, and all functions of these are uncorrelated with a_{it}.

Plugging $v_{it} = f[g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})] + a_{it}$ into Eq. (2.1) gives

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + f\left[g\left(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}\right)\right] + a_{it} + e_{it}. \tag{2.9}$$

Now, we can specify the two equations that identify (β,γ):

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + g(\mathbf{x}_{it}, \mathbf{m}_{it}) + e_{it},\ t = 1, \ldots, T \tag{2.10}$$

and

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + f\left[g\left(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}\right)\right] + u_{it},\ t = 2, \ldots, T, \tag{2.11}$$

where $u_{it} \equiv a_{it} + e_{it}$. Importantly, the available orthogonality conditions differ across these two equations. In Eq. (2.10), the orthogonality condition on the error is given by Eq. (2.5). The orthogonality conditions for Eq. (2.11) are

$$E\left(u_{it}\,|\,\mathbf{x}_{it}, \mathbf{w}_{i,t-1}\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots, \mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}\right) = 0,\ t = 2, \ldots, T. \tag{2.12}$$

In other words, in Eqs. (2.10) and (2.11) we can use the contemporaneous state (capital) variables, $\mathbf{x}_{it}$, any lagged inputs, and functions of these, as instrumental variables. In Eq. (2.10) we can further add the elements of $\mathbf{m}_{it}$ (investment or intermediate inputs).

In the ACF setting, where Eq. (2.10) does not identify β, Eq. (2.11) would still generally identify β and γ provided we have the orthogonality conditions in Eq. (2.12). Effectively, $\mathbf{x}_{it}$, $\mathbf{x}_{i,t-1}$, and $\mathbf{m}_{i,t-1}$ act as their own instruments and $\mathbf{w}_{i,t-1}$ acts as an instrument for $\mathbf{w}_{it}$. Then, Eq. (2.11) can be estimated by an instrumental variable version of Robinson's (1988) estimator to allow f and g to be completely unspecified.

3. Estimating equations in the parametric case

To estimate β and γ, we must deal with the presence of unknown functions, $g(\cdot,\cdot)$ and $f(\cdot)$ in Eqs. (2.10) and (2.11). An approach that LP find works as well is to use third-degree polynomials. So, if x_{it} and m_{it} are both scalars, $g(x,m)$ is linear in terms of the form $x^p m^q$, where p and q are nonnegative integers with $p+q \le 3$. More generally, $g(\mathbf{x},\mathbf{m})$ contains all polynomials of order three or less. In any case, assume that we can write

$$g(\mathbf{x}_{it},\mathbf{m}_{it}) = \lambda_0 + \mathbf{c}(\mathbf{x}_{it},\mathbf{m}_{it})\lambda \tag{3.1}$$

for a $1\times Q$ vector of functions $\mathbf{c}(\mathbf{x}_{it}, \mathbf{m}_{it})$. I assume that $\mathbf{c}(\mathbf{x}_{it}, \mathbf{m}_{it})$ contains at least $\mathbf{x}_{it}$ and $\mathbf{m}_{it}$ separately, since a linear version of $g(\mathbf{x}_{it}, \mathbf{m}_{it})$ should always be an allowed special case. Further, assume that $f(\cdot)$ can be approximated by a polynomial in v:

$$f(v) = \rho_0 + \rho_1 v + \ldots + \rho_G v^G. \tag{3.2}$$

When we plug these choices into Eqs. (2.10) and (2.11), it is evident that neither the original intercept α nor the intercepts λ_0 and ρ_0 are identified.

Given the functions in Eqs. (3.1) and (3.2), we now have

$$y_{it} = \alpha_0 + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + \mathbf{c}_{it}\lambda + e_{it},\ t = 1,\ldots,T \tag{3.3}$$

and

$$y_{it} = \eta_0 + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + \rho_1\left(\mathbf{c}_{i,t-1}\lambda\right) + \ldots + \rho_G\left(\mathbf{c}_{i,t-1}\lambda\right)^G + u_{it},\quad t = 2,\ldots,T, \tag{3.4}$$

where α_0 and η_0 are the new intercepts and we use the notation $\mathbf{c}_{it} \equiv \mathbf{c}(\mathbf{x}_{it}, \mathbf{m}_{it})$. Given Eqs. (2.5) and (2.12), we can easily specify instrumental variables (IVs) for each of these two equations. The most straightforward choice of IVs for Eq. (3.3) is simply

$$\mathbf{z}_{it1} \equiv (1, \mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{c}_{it}^o), \tag{3.5}$$

where $\mathbf{c}_{it}^o$ is $\mathbf{c}_{it}$ but without $\mathbf{x}_{it}$. The choice in Eq. (3.5) corresponds to the regression analysis in OP and LP for estimating β in a first stage. Of course, under Eq. (2.5), any nonlinear function of $(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{c}_{it}^o)$ is also a valid IV, as are all lags and all functions of these lags. Adding a lag could be useful for generating overidentifying restrictions to test the model assumptions, particularly Eq. (2.2).

Instruments for Eq. (3.4) would include $(\mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1})$ and, especially if $G>1$, nonlinear functions of $\mathbf{c}_{i,t-1}$ (probably low-order polynomials). Lags more than one period back are valid, too, but adding more lags can be costly in terms of lost initial time periods. So, write

$$\mathbf{z}_{it2} = \left(1, \mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1}, \mathbf{q}_{i,t-1}\right), \tag{3.6}$$

where $\mathbf{q}_{i,t-1}$ is a set of nonlinear functions of $\mathbf{c}_{i,t-1}$, probably consisting of low-order polynomials.

We can easily verify that we have enough moment conditions to identify the $2+J+K+Q+G$ parameters in Eq. (3.3). In fact, we can identify the parameters β and γ off Eq. (3.4). As remarked earlier, $(\mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1})$ would act as their own instruments, and then we would include enough nonlinear functions in $\mathbf{q}_{i,t-1}$ to identify $\rho_1,\ldots,\rho_G$.

A key difference between Eqs. (3.5) and (3.6) is that $\mathbf{z}_{it2}$ does not contain contemporaneous values of $\mathbf{w}_{it}$ and $\mathbf{m}_{it}$. One possibility is to choose, for each i and t, a matrix of instruments, with two rows, as

$$\mathbf{Z}_{it} \equiv \begin{pmatrix} (\mathbf{w}_{it}, \mathbf{c}_{it}, \mathbf{z}_{it2}) & 0 \\ 0 & \mathbf{z}_{it2} \end{pmatrix},\ t = 2,\ldots,T. \tag{3.7}$$

This choice makes it clear that all instruments available for Eq. (3.5) are also valid for Eq. (3.4), and we have some additional moment restrictions in Eq. (3.4).

GMM estimation of all parameters in Eqs. (3.3) and (3.4) is now straightforward. For each $t>1$, define a 2×1 residual function as

$$\mathbf{r}_{it}(\theta) = \begin{pmatrix} \mathbf{r}_{it1}(\theta) \\ \mathbf{r}_{it2}(\theta) \end{pmatrix} = \begin{pmatrix} y_{it} - \alpha_0 - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - \mathbf{c}_{it}\lambda \\ y_{it} - \eta_0 - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - \rho_1(\mathbf{c}_{i,t-1}\lambda) - \ldots - \rho_G(\mathbf{c}_{i,t-1}\lambda)^G \end{pmatrix}, \tag{3.8}$$

so that

$$E[\mathbf{Z}_{it}'\mathbf{r}_{it}(\theta)] = 0,\ t = 2,\ldots,T. \tag{3.9}$$

Then, these $T-1$ conditions can be stacked for each i, and standard GMM estimation can be used; see, for example, Wooldridge (1996, 2002, Chapter 14).

Interestingly, in one leading case – namely, that productivity follows a random walk with drift – the moment conditions are linear in the parameters. Using $G=1$ and $\rho_1=1$, the residual functions become $r_{it1}(\theta) = y_{it} - \alpha_0 - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - \mathbf{c}_{it}\lambda$ and $r_{it2}(\theta) = y_{it} - \eta_0 - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - \mathbf{c}_{i,t-1}\lambda$, which results in a particularly straightforward GMM estimation problem. In fact, we can write the system as $\mathbf{y}_{it} = \mathbf{X}_{it}\theta + \mathbf{r}_{it}$, where $\mathbf{y}_{it}$ is the 2×1 vector with y_{it} in both elements,

$$\mathbf{X}_{it} = \begin{pmatrix} 1 & 0 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{it} \\ 0 & 1 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{i,t-1} \end{pmatrix}, \tag{3.10}$$

and $\theta = (\alpha_0, \eta_0, \beta', \gamma', \lambda')'$. We can choose $\mathbf{Z}_{it}$ as in Eq. (3.7). Identification does not require including $\mathbf{q}_{i,t-1}$ in $\mathbf{z}_{it2}$, but we might include $\mathbf{q}_{i,t-1}$ among the instruments and test the several overidentifying restrictions.

Acknowledgement

I am grateful to Dan Ackerberg for providing helpful comments on earlier drafts.

References

Ackerberg, D.A., Caves, K., and Frazer, G., 2006. Structural identification of production functions, mimeo, UCLA Department of Economics.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70, 317–341.

Olley, S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1298.

Petrin, A., Poi, B.P., Levinsohn, J., 2004. Production function estimation in Stata using inputs to control for observables. Stata Journal 4, 113–123.

Robinson, P.M., 1988. Root-n consistent semiparametric regression. Econometrica 55, 931–954.

Wooldridge, J.M., 1996. Estimating systems of equations with different instruments for different equations. Journal of Econometrics 74, 387–405.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.