9 Estimation of total factor productivity

Robert Breunig and Marn-Heong Wong

Australian National University

Abstract

The micro reality is one of firms being differentiated by productivity differences and industries experiencing constant flux with entry of new firms and exit of failed firms. If firms make decisions on input demand and liquidation based on their productivity (the latter known to them but unobserved by the econometrician), then simultaneity and selection problems arise that bias the traditional estimators of production function coefficients. We apply for the first time on Australian firm-level data a semiparametric production function estimation technique that endogenises input choices and firm exit decisions. Results obtained for over 20 industries at the two-digit ANZSIC level using the Business Longitudinal Survey dataset support the use of this technique to improve productivity estimates.

9.1 Introduction

This paper reviews some of the econometric issues associated with the use of least-squares techniques on firm-level data. Such techniques are applied in estimating industry-level production functions, often with the goal of measuring total factor productivity. Standard techniques have implicit behavioural assumptions that might not be very realistic or match how firms actually behave.

We introduce a recently developed semiparametric method by Olley and Pakes (1996) that incorporates more realistic assumptions about firm behaviour than the standard techniques. We report the results of our application of this method, which is applied for the first time to Australian data.

9.2 Issues in production function estimation

Two commonly used techniques for estimating firm-level productivity are ordinary least squares (OLS) and fixed effects (FE). However, concerns have been raised that these traditional estimators could yield biased coefficients. If productivity differences across firms are the norm, and firms make decisions on input demand and liquidation based on their productivity, the latter known to them but unobserved by the econometrician, then this gives rise to simultaneity and selection biases.

Simultaneity problem

Firm productivity can be both contemporaneously and serially correlated with inputs. If that is so, an OLS estimation that assumes no correlation between input demands and the unobserved productivity term will give inconsistent estimates of the input coefficients. This simultaneity bias has been identified since Marschak and Andrews (1944). Contemporaneous correlation occurs if more productive firms hire more workers and invest in capital in response to higher current and expected future profitability. The problem is likely to be more acute for inputs such as labour that can be adjusted rapidly to current productivity realisations. If a firm's productivity is correlated over time, then input choices will be based on a serially correlated productivity term. The OLS estimates will be biased upwards in a single input case, but the direction of the inconsistency is indeterminate in a multivariate setting. For example, if labour and capital are positively correlated, but labour is more strongly correlated with the productivity term than capital, then the labour coefficient will tend to be overestimated and the capital coefficient underestimated.

A standard solution is to compute an FE or 'within' estimator that uses deviations from firm-specific means in OLS estimation. This controls for simultaneity provided the firm's productivity is time invariant. However, productivity is unlikely to remain constant over long periods of time, especially during periods of significant policy and structural changes. The constant flux in firm decisions regarding input use and firm entry and exit suggests a more general stochastic process for the unobserved productivity term than that specified by fixed effects. If that is the case, the FE estimator will at best remove the effects of the time-invariant component of the productivity variable, but will still lead to inconsistent estimates.

The issue of selection

The econometrician observes only those firms that stay in business. We can reasonably assume that a firm decides to continue operation if its expected future profits exceed its liquidation value. If a firm's future returns are positively related to the size of its capital stock at any given current productivity level, then firms with greater capital stock are more likely to survive lower productivity realisations. The expectation of productivity conditional on the surviving firms is thus decreasing in capital, leading to a negative bias in the capital coefficient. The selection problem is all the more severe in analyses using a ‘balanced’ panel, which is the traditional way to ‘avoid’ dealing with entry and exit, as it keeps only those firms that operate over the entire sample period.

Olley and Pakes (1996) developed an innovative methodology to solve these two problems, which is increasingly being applied in production function estimation using microdata. In particular, their algorithm might be the first to take explicit account of the selection bias, although the issue was discussed in the empirical literature at least since the work of Wedervang (1965). The estimation method is attractive not so much because it introduces sophisticated statistical techniques, but because it is underpinned by a dynamic model of firm behaviour that incorporates time-varying, firm-specific or idiosyncratic productivity differences, and endogenises firm exits. In terms of implementation, Olley and Pakes used a semiparametric approach, which avoided the need to add more structure to the behavioural framework to obtain specific functions for the shutdown and input demand decisions.

9.3 The algorithm of Olley and Pakes

The centrepiece of the Olley and Pakes (1996) methodology (the OP method) is the expression of the unobserved productivity term in terms of observables (specifically, investment demand), as derived from their behavioural framework. This allows for correlation between a firm’s productivity and input choices that the OLS technique disregards. Furthermore, changes in productivity over time can be captured by tracking the observable variables. This makes the OP method a more flexible formulation compared with FE estimation, which is based on the assumption of a time-invariant firm-specific effect.

Behavioural framework

Olley and Pakes extracted features from the models of Ericson and Pakes (1995) and Hopenhayn and Rogerson (1993) to formulate their behavioural model.

At any time *t*, a firm seeks to maximise its expected discounted value of net cash flows. It has to decide, first, whether to continue or cease business. If it stays, it chooses variable factors (labour) and a level of investment.

Its value function is given by:

$$V_t(\omega_t, a_t, k_t) = \max \{\Phi, \sup \pi_t(\omega_t, a_t, k_t) - c(i_t) + \beta E[V_{t+1}(\omega_{t+1}, a_{t+1}, k_{t+1})|J_t] \} \qquad (1)$$

where Φ is the sell-off value of its capital, $\pi_t(\omega_t, a_t, k_t)$ the current profits that are a function of the firm's state variables, in this case, the triple of ω_t, its unobserved productivity, a_t, its age, and k_t, its capital stock. $c(i_t)$ is the cost of investment, β is the discount rate and J_t is the information at time t.

ω_t is assumed to evolve as a first-order Markov process , which means that current period productivity depends on the previous period's productivity. Muendler (2004) introduced an even more appealing model where ω_t is the result of both stochastic factors and the firm's explicit decision to try to increase its productivity by enhancing managerial quality and other productivity-enhancing investments.

A firm's profit is affected also by the market structure and factor prices. However, it is assumed here that factor prices are common across all firms. It is also assumed that all firms in an industry face the same market structure in any given period. Thus, these conditions do not require additional notation, but can be depicted by indexing the various functions by time.

The accumulation equations of capital and age are as follows:

$$k_{t+1} = (1-\delta)k_t + i_t \qquad (2)$$

$$a_{t+1} = a_t + 1 \qquad (3)$$

where δ is the capital depreciation rate.

The solution to the firm's optimisation problem (shown in Ericson and Pakes (1995)) generates an exit rule:

$$\chi_t = 1 \text{ stay, if } \omega_t \geq \omega^*_t(k_t, a_t) \qquad (4)$$

$$\quad 0 \text{ exit, otherwise}$$

and an investment demand function:

$$i_t = i_t(\omega_t, a_t, k_t) \qquad (5)$$

Empirical application

The estimation method starts by specifying a production function in Cobb-Douglas form for a given industry, with firms distinguished by Hicks-neutral efficiency differences:

$$y_{it} = \beta_0 + \beta_a a_{it} + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \eta_{it} \tag{6}$$

where y_{it} is the log of output (value added) from firm i at time t, a_{it} is its age, k_{it} is the log of its capital stock, l_{it} is the log of labour input, and ω_{it} is its productivity. η_{it} is a mean zero error that accounts for measurement error of unanticipated productivity shocks that do not affect the choice of inputs. The firm subscripts are omitted in subsequent equations for ease of presentation.

We assume that labour is the only variable input, whose demand is affected by the current value of ω_t, and capital and age are fixed factors dependent only on the distribution of ω_t conditional on information at time t –1 and past values of ω. These assumptions were also employed by Olley and Pakes (1996).

The procedure is implemented in three stages.

Stage 1

In the first stage, a consistent estimate of the coefficient on the variable input is obtained. Investment is used as a proxy to control for the correlation between the unobserved productivity term and the variable input. The optimal investment level in each period is a function of the state variables ω, a and k (equation 5). Provided $i_t > 0$, Pakes (1994) showed that this equation is strictly increasing in ω (for every (a, k)). This then allows (5) to be inverted to express ω as a function of observables:

$$\omega_t = h_t(i_t, a_t, k_t) \tag{7}$$

Substituting (7) into (6) yields:

$$y_i = \beta_l l_i + \lambda_t(i_t, a_t, k_t) + \eta_{it} \tag{8}$$

$$\text{where } \lambda_t(i_t, a_t, k_t) = \beta_0 + \beta_a a_t + \beta_k k_t + h_t(i_t, a_t, k_t) \tag{9}$$

Equation 8 is a partially linear model estimated with semiparametric regression techniques (see, for example, Engel, Granger, Rice and Weiss 1986; Robinson 1988). Olley and Pakes used a series estimator for $\lambda_t(.)$ as in Newey (1994), which they implemented as a fourth-order polynomial (with a full set of interactions) in the triple (i_t, a_t, k_t). Andrews (1991) showed that a partially linear model using series approximation of the nonlinear portion yields consistent and asymptotically normal estimates of coefficients of the linear part of the model.

Although the labour coefficient is identified, the coefficients of capital and age are not, as equation 8 does not allow us to separate the contribution of capital and age to output from their impact on the investment decision. To see this, note simply that $\frac{\partial y}{\partial k} = \beta_k + h_k$, so that the coefficient on capital can no longer be interpreted in the usual manner.

Stage 2

The second step of the algorithm involves the estimation of survival probabilities to correct for the selection problem. These probabilities, together with the estimates of β_l and $\lambda_t(.)$ in stage 1, will enable the identification of β_a and β_k.

Consider the value of output in the next period:

$$y_{t+1} = \beta_0 + \beta_a a_{t+1} + \beta_k k_{t+1} + \beta_l l_{t+1} + E[\omega_{t+1}|\omega_t, \chi_{t+1}=1] + \xi_{t+1} + \eta_{t+1} \tag{10}$$

where $\omega_{t+1} = E[\omega_{t+1}|\omega_t, \chi_{t+1}=1] + \xi_{t+1}$, the first term being the expected value of next period's productivity conditional on the current period productivity and firm survival, and the second term is the innovation in productivity.

Note that expected productivity ω_{t+1} is conditional not only on ω_t but also on survival. A firm's decision to stay or shut down in the next period depends on whether its productivity at t + 1 is above some threshold value ω^*_{t+1}. This value, which is a function of the firm's age and capital stock, is endogenously determined in equilibrium and is known to the firm but not the econometrician. Thus, there is a need to control for the impact of the unobservable ω^*_{t+1} on selection in the estimation procedure.

Define $g(\omega^*_{t+1}, \omega_t) = \beta_0 + E[\omega_{t+1}|\omega_t, \chi_{t+1}=1]$ (11)

$$= \beta_0 + \int_{\omega^*_{t+1}} \omega_{t+1} \frac{F(d\omega_{t+1} | \omega_t)}{\int_{\omega^*_{t+1}} F(d\omega_{t+1} | \omega_t)}$$

g(.) is a function, up to an additive constant, of two indices of firm-specific state variables ω_t and ω^*_{t+1}. Information on ω^*_{t+1} can be obtained by evaluating the probability that a firm continues to produce in time t+1:

$$\Pr\{\chi_{t+1}=1\} = \Pr\{\omega_{t+1} \geq \omega^*_{t+1}(k_{t+1}, a_{t+1}) | \ \omega^*_{t+1}(k_{t+1}, a_{t+1}), \omega_t\} \tag{12}$$

$= \varphi_t\{\omega^*_{t+1}(k_{t+1}, a_{t+1}), \omega_t\}$

$= \varphi_t(i_t, a_t, k_t)$

$\equiv P_t$

The third line follows from the investment rule and the accumulation equations for capital and age. The survival probabilities can then be estimated by running a probit regression on investment, age and capital. Instead of specifying a linear index function, we allow the index function to be of unspecified non-parametric form. This is estimated by a fourth-order polynomial in investment, age, and capital.

From the second line, φ_t can be inverted to express ω^*_{t+1} as a function of P_t and ω_t, provided the density of ω_{t+1} conditional on ω_t is positive in the region about ω^*_{t+1} (for every ω_t). Furthermore, by conditioning on the nonlinear term in equation (9) from stage 1, ω_t can be expressed as $h_t = \lambda_t - \beta_a a_t - \beta_k k_t$. Thus, g(.) can be rewritten as $g(P_t, h_t)$.

Stage 3

In the third and final stage, equation (10) is rewritten as follows:

$$y_{t+1} - \hat{\beta}_l l_{t+1} = \beta_a a_{t+1} + \beta_k k_{t+1} + g(\hat{P}_t, \hat{h}_t) + \xi_{t+1} + \eta_{t+1} \qquad (13)$$

Since k_{t+1} is uncorrelated with both ξ_{t+1} and η_{t+1}, the coefficient on capital can be consistently estimated. Since the equation is nonlinear in β_k, the nonlinear least squares technique can be used, with the unknown function g(.) approximated by a fourth-order polynomial expansion in $(\hat{P}_t, \hat{h}_t)$ of the form $\hat{g}(P_t, \hat{h}_t) = \sum_{j=0}^{4-m}\sum_{m=0}^{4}\beta_{mj}\hat{h}_t^m \hat{P}_t^j$. (The non-linearity arises through *h*, which takes the form $\lambda_t - \beta_a a_t - \beta_k k_t$ — see above).

Review of results using OP estimation

Olley and Pakes and a few related papers compared the production function estimates obtained from OP with estimators such as OLS and FE, which is what we intend to do in our empirical analysis. Hence, we briefly review their findings for later comparison. Generally, estimations carried out on different datasets found that the labour coefficients using OP were lower and the capital coefficients higher, compared with the traditional estimators, thus supporting the theory of a simultaneity and a selection bias.

Olley and Pakes found that the move from a balanced panel to the full sample almost doubled the capital coefficients and lowered the labour coefficient by about 20 per cent, in both the OLS and FE estimations. The labour coefficient in stage 1 of OP was 15 per cent lower than the OLS value, while the capital coefficient from the series estimator in stage 3 was 12.5 per cent higher than the OLS (full panel) estimate.

Pavcnik (2002), using Chilean manufacturing data, reported that the OP coefficients for the variable inputs of unskilled and skilled labour and materials were lower than the OLS estimates for all but one of the eight industries she considered, based on the unbalanced panel. Five (or 63 per cent) of the industries reported higher point estimates of capital coefficients based on the OP estimation compared with OLS, by 45–300 per cent. The FE coefficients were often much lower than the OLS or OP estimates, especially for capital.

Estimations by Levinsohn and Petrin (1999) using an intermediate input (electricity consumption) proxy (the LP estimator) yielded higher capital coefficients compared with the OLS estimator in all eight industries in the Chilean manufacturing sector. The increase was large (ranging from 35 to 110 per cent) for all except two industries. Levinsohn and Petrin also found that the coefficient on blue collar labour fell in every industry, while results for white collar labour were mixed.

Levinsohn and Petrin (2003) observed that 'the fixed effect estimator is in the most pronounced disagreement with the other estimators', which is taken to imply that the productivity shock seems to vary within firm over time.

More on empirical implementation

We add year dummies to the basic specification of the OP model to control for year-to-year changes in the data. We also introduce dummies to account for observations with zero investment. The theoretical model of Olley and Pakes requires that investment be strictly positive, which permits the inversion of the investment function on which the estimation of the unobserved productivity term is based. In their empirical implementation, Olley and Pakes dropped all observations with zero investment.

Other authors have noted that while this is a theoretical requirement, in practice, zero investment is often observed, and that the methodology seems to work even when the theory is violated somewhat. Dropping firm/year combinations with zero investment would lead us to drop over half of our observations. So our approach is to retain all the observations with zero investment but to introduce dummy variables (investment interacted with state inputs) to account for these observations, as in

Blalock and Gertler (2003). If we estimate the model dropping these observations with zero investment, we get similar coefficient estimates to those presented below.

Boostrapped standard errors (using 200 replications) are computed and reported for the age and capital coefficient estimates. This is because the series estimator used for g(.) in equation 13 has no known limiting properties, although Olley and Pakes, who proved asymptotic results for the kernel estimator of g(.), had suggested that the series estimator should have the same properties as the kernel estimator, since the parameter estimates yielded by the two were not significantly different.

Levinsohn and Petrin (1999) implemented a specification test to examine whether there were further grounds to justify the use of the OP/LP way of estimation compared with OLS and FE. Since the OP/LP approach assumes that the residuals follow a first-order Markov process, the assumption nests both the OLS and the FE specifications. Thus, a Wald test can be conducted to test the hypothesis that residuals in each period are uncorrelated with those in the period before (OLS), or the hypothesis that there is perfect correlation (FE). We shall do likewise.

9.4 Data description

Overview of the Business Longitudinal Survey

We apply the OP production function estimation algorithm to data from the Business Longitudinal Survey (BLS), which is Australia’s only business longitudinal micro-dataset that tracks firm entry and exit. The BLS panel, compiled by the Australian Bureau of Statistics (ABS), contains data from four waves of the survey covering 1994-95 to 1997-98. Businesses were chosen from the ABS Business Register based on the stratified random sampling method, where the stratification was by both industry and employment size classification. The first wave of 9000 live responses were further stratified into two categories in 1995-96: firms identified as innovators, exporters, or those with high employment or sales growth, which numbered about 3400, continued to be surveyed; of the remaining 5600 live respondents, about 2200 were selected for inclusion in the survey. In addition, a random sample of new firms, or births, was selected for the 1995-96 survey. In subsequent years, all firms surveyed in the previous year were traced, with exits recorded, and births were included.

The full sample of the BLS can be accessed only remotely, through the ABS running program codes prepared by researchers and then conveying the results back to them. The publicly available file is known as the CURF (Confidentialised Unit Record File) and excludes information on large businesses, namely, firms with more than 200 employees, or those with large measures other than employment, such as sales. The version of CURF we are using is the November 2001 release that had corrected data errors identified by various earlier users.

The BLS covered only non-agricultural market sectors and excluded industries with heavy government involvement, such as health, education and communications services. While each business in the survey is coded to the four-digit ANZSIC level, only its two-digit industry codes are released in the CURF. The regressions are run on industries classified at the two-digit level. Several industries have been excluded from our estimation. These are the mining, transport and storage, and finance and insurance industries. In the former two, firms in the CURF are coded by their one-digit industry codes because of the predominance of larger firms. In the latter, the problem lies with the measurement of firm output, which was highlighted in Rogers (1998). The income from sales of goods and services used to calculate output is not a good measure for this industry, which derives its income also from interest earned on financial assets. Interest income is available in the BLS under the question 'other income', but this 'other income' includes nine other types of income.

The 'full sample', or unbalanced panel, is constructed by retaining firms that eventually exit until the year prior to their exit, and introducing new entrants as they appear. One important issue is the classification of 'truly' new entrants and exiting firms. Will and Wilson (2001) discovered anomalies in the data on births and deaths, and derived criteria for identifying 'true' births and deaths. We investigated this issue further and decided to modify their 'true' birth rule but adopt their rule for removing 'illegitimate' deaths. In short, true births are identified as firms coded as entrants that are aged less than 4 years, with total employment of less than 30 or not more than median industry sales at survey entry. True deaths are defined as firms that exit the survey and record no change or a fall in employment, and a rise in capital stock of no more than 5 per cent, in the year prior to exit.[1]

Variable definitions

Variable definitions used in the production function regressions are given in box 9.1.

1 Readers interested in obtaining a more detailed write-up on the correction for true births and deaths can email Marn-Heong Wong (wmhoz@yahoo.com.sg).

Box 9.1 **Variable definitions**

y: value-added: sales plus change in inventories less purchases of intermediate inputs and other operating expenses

k: capital stock: book value of total non-current assets plus leasing stock. Leasing capital is obtained by dividing leasing expenses by (0.05 + 0.0803), where 0.05 = 1/20 is the average years of depreciation, and 0.0803 is the average 10-year treasury bond rate form July 1994 to June 1998.

l: full-time equivalent persons: the number of full-time employees plus 0.426 * the number of part-time employees, averaged over two years

i: investment: the sum of capital expenditure on plant, machinery, equipment, land, dwellings, other buildings and structures, and intangible assets

a: age of firm: calculated as the midpoint of the range of responses — that is, less than 2 years = 1, 2–4 years = 3, … over 30 years = 35

Dy: year dummies.

Dik: indicator dummy for observations with zero investment interacted with capital stock

Dia: indicator dummy for observations with zero investment interacted with age

9.5 Analysis of estimation results

Table 9.1 shows estimation results for 23 industries across the three estimation techniques of OLS, FE, and OP, the former two on both a balanced and unbalanced panel.

First, we examine the OLS coefficients between the balanced and unbalanced panel (columns 1 and 3). If restoring observations to form an unbalanced panel alleviates the simultaneity and selection problems, we would expect the labour coefficient to fall and the capital coefficient to rise when estimating by OLS on a full sample. This direction of change occurs in slightly half of the industries. More industries (65 per cent) register a higher capital coefficient in the full sample, compared with the number of industries (48 per cent) that have a lower labour coefficient. Where the labour coefficient is lower, the reduction is usually below 10 per cent. Where the capital coefficient is higher, the increase is usually within the range of 2–38 per cent. These percentage changes are nowhere near as dramatic as the changes to the coefficients in Olley and Pakes, which is unsurprising since they increased the sample size by 189 per cent by moving from a balanced to unbalanced panel, compared with our much more modest average rise of 32 per cent.

We now compare the OLS (full sample) and OP coefficients (columns 3 and 5). Since OLS regression, even on a full sample, does not control for firm-specific differences in productivity, we would expect the OLS labour coefficient to remain biased upwards because of the endogeneity of input choices, which is corrected in OP. This is strongly supported by 87 per cent of the industries having lower labour coefficients in the OP estimates. The drop in point estimates ranges from 0.5 to 13 per cent. All the labour coefficients are significant at the 1 per cent level.

The direction of change of the capital coefficient from OLS (full sample) to OP is predominantly negative, with the reduction by between 1 and 80 per cent. This implies a positive bias in the OLS coefficient. Only five of the industries have higher point estimates of capital for OP, with the increase in the range of 2 and 40 per cent. Although these results differ from that obtained in Olley and Pakes and several others, they are not perplexing within the current framework, as there can be several biases working in different directions, and at varying magnitudes, on the capital coefficient at the same time. If selection for survival is important, there will be a negative bias in the OLS estimate, as emphasised in OP. However, the OLS capital coefficient can be biased upwards if capital usage is correlated with the productivity shock contemporaneously or serially. Levinsohn and Petrin (2003) also pointed out that if capital positively covaries with labour, but is uncorrelated with the productivity shock, or if this correlation is much weaker than that between the variable inputs and productivity, then the OLS estimate on capital is likely to be biased downwards. Our findings would indicate that there is strong correlation between capital and productivity, and that the simultaneity bias dominates the selection bias in most cases. Seventy-eight per cent of the capital coefficients in CURF are significant at the 1 per cent level. The capital coefficients are insignificant for three industries:

- petroleum, coal, chemical and associated products
- motion picture, radio and television services
- sport and recreation.

Relative to the OLS and OP estimates, both the labour and capital coefficients from running FE, even on a full sample, are much lower. On average, they are about half the value of the OLS and OP coefficients. This is in line with studies that find that FE estimates usually disagree markedly with other estimators, and is further evidence that the assumption of a time-invariant, firm FE is quite poor.

The age coefficients are always small in value, and all of them are insignificant. Thus, they will not be further discussed. Dropping them and re-estimating does not affect the substantive results presented here.

Wald tests on the alternative specifications strongly reject the respective assumptions of OLS and FE that the residuals are uncorrelated and perfectly correlated, which would validate the OP premise that the residuals are correlated, but in a time-varying manner.

9.6 Summary conclusion

This paper applies a method in production function estimation that corrects for the twin biases of simultaneity and selection usually associated with traditional estimators such as OLS and fixed effects panel regression (FE). Estimation using the semiparametric technique suggested by Olley and Pakes is carried out on Australia's BLS, covering 23 industries at the two-digit ANZSIC level.

We find that labour parameter estimates mainly become lower, as we move from running OLS on a balanced panel, to OLS on the full sample, and then to OP. This supports the hypothesis that there is a positive bias in the OLS estimate, because it does not control for the simultaneity between firms' labour input choices and productivity, which is corrected under the model assumptions and econometric techniques employed here.

As for the capital coefficients, the majority of OLS estimates using the full sample are higher than the values yielded by a balanced panel. However, the capital estimates from OP are predominantly less than those obtained from OLS on the full sample. This suggests that any downward bias on the OLS capital coefficient, exerted by less productive firms' selection to stay in business based on their higher levels of capital stock, is countered by a larger positive bias from the simultaneity between firms' capital usage and productivity. This is perhaps not surprising, given the fairly modest exit rates in the sample (around 12 per cent on average over three years) due to the sample period being one of steady economic expansion.

As for FE estimation, the labour and capital coefficients obtained are usually much lower than the OLS and OP estimates, which indicate that firm-specific productivity shocks seem to vary over time.

Lastly, Wald tests comparing the OLS, FE and OP specifications overwhelmingly reject the hypothesis that the residuals are uncorrelated, or perfectly correlated. This justifies the choice of OP in production function estimation.

Table 9.1 **Results from production function estimation**[a]

ANZSIC/industry	Balanced panel (1) OLS	(2) Within	Full sample (3) OLS	(4) Within	(5) OP
Manufacturing					
21 Food, beverage and tobacco					
Labour	0.818	0.586	0.816	0.647	0.780
	(0.032)**	(0.071)**	(0.032)**	(0.069)**	(0.032)**
Capital	0.312	0.111	0.318	0.156	0.298
	(0.023)**	(0.035)**	(0.022)**	(0.035)**	(0.046)**
Age	0.001	0.012	0.002	0.011	–0.010
	(0.002)	(0.020)	(0.002)	(0.021)	(0.019)
N	536	536	647	629	454
22 Textile, clothing, footwear and leather					
Labour	0.779	0.304	0.701	0.393	0.638
	(0.034)**	(0.082)**	(0.034)**	(0.085)**	(0.035)**
Capital	0.281	0.208	0.318	0.216	0.300
	(0.025)**	(0.040)**	(0.024)**	(0.036)**	(0.086)**
Age	0.005	0.049	0.009	0.047	0.017
	(0.002)*	(0.021)*	(0.003)**	(0.023)*	(0.018)
N	416	416	496	481	349
23 Wood and paper product					
Labour	0.923	0.661	1.021	0.565	0.953
	(0.056)**	(0.098)**	(0.053)**	(0.099)**	(0.068)**
Capital	0.257	0.073	0.183	0.093	0.166
	(0.040)**	(0.043)+	(0.035)**	(0.042)*	(0.096)+
Age	0.003	0.003	0.004	–0.024	0.003
	(0.004)	(0.026)	(0.004)	(0.026)	(0.014)
N	252	252	348	338	238
24 Printing, publishing and recorded media					
Labour	0.799	0.340	0.811	0.302	0.660
	(0.046)**	(0.085)**	(0.043)**	(0.076)**	(0.045)**
Capital	0.278	0.151	0.241	0.156	0.229
	(0.033)**	(0.037)**	(0.030)**	(0.035)**	(0.066)**
Age	0.006	0.020	0.009	0.025	–0.005
	(0.003)+	(0.024)	(0.003)**	(0.022)	(0.020)
N	384	384	467	457	330

(Continued on next page)

Table 9.1 (continued)

	Balanced panel		Full sample		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
ANZSIC/industry	*OLS*	*Within*	*OLS*	*Within*	*OP*
25 Petroleum, coal, chemical and associated product					
Labour	0.801	0.597	0.871	0.513	0.863
	(0.032)**	(0.063)**	(0.032)**	(0.061)**	(0.033)**
Capital	0.342	0.109	0.269	0.089	0.052
	(0.020)**	(0.024)**	(0.020)**	(0.025)**	(0.043)
Age	–0.002	0.015	–0.001	0.025	0.018
	(0.002)	(0.015)	(0.002)	(0.016)	(0.012)
N	644	644	737	718	528
26 Non-metallic mineral product manufacturing					
Labour	0.955	0.384	0.919	0.305	0.890
	(0.070)**	(0.100)**	(0.063)**	(0.104)**	(0.061)**
Capital	0.162	0.075	0.198	0.069	0.242
	(0.047)**	(0.052)	(0.040)**	(0.055)	(0.084)**
Age	0.008	0.000	0.009	0.023	0.000
	(0.005)+	(0.030)	(0.005)*	(0.031)	(0.013)
N	224	224	298	285	203
27 Metal product					
Labour	0.916	0.502	0.964	0.520	0.929
	(0.028)**	(0.052)**	(0.027)**	(0.053)**	(0.031)**
Capital	0.237	0.109	0.216	0.128	0.238
	(0.020)**	(0.025)**	(0.019)**	(0.026)**	(0.060)**
Age	–0.002	–0.004	–0.004	–0.004	–0.006
	(0.002)	(0.015)	(0.002)+	(0.016)	(0.011)
N	728	728	825	810	591
28 Machinery and equipment					
Labour	0.878	0.481	0.851	0.495	0.839
	(0.022)**	(0.063)**	(0.022)**	(0.058)**	(0.024)**
Capital	0.220	0.121	0.229	0.124	0.222
	(0.016)**	(0.022)**	(0.016)**	(0.021)**	(0.045)**
Age	0.001	0.000	0.002	0.003	–0.004
	(0.002)	(0.015)	(0.002)	(0.016)	(0.013)
N	1312	1312	1475	1452	1061

(Continued on next page)

Table 9.1 (continued)

	Balanced panel		Full sample		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
ANZSIC/industry	*OLS*	*Within*	*OLS*	*Within*	*OP*
29 Other					
Labour	0.924	0.834	0.945	0.714	0.831
	(0.033)**	(0.071)**	(0.032)**	(0.072)**	(0.036)**
Capital	0.176	0.085	0.159	0.104	0.173
	(0.023)**	(0.034)*	(0.021)**	(0.031)**	(0.073)*
Age	0.007	0.001	0.01	0.003	0.001
	(0.003)**	(0.022)	(0.003)**	(0.023)	(0.013)
N	500	500	601	586	420
Construction					
41 General construction					
Labour	0.935	0.529	0.884	0.505	0.790
	(0.060)**	(0.148)**	(0.048)**	(0.132)**	(0.054)**
Capital	0.208	0.188	0.215	0.221	0.301
	(0.038)**	(0.059)**	(0.031)**	(0.053)**	(0.085)**
Age	0.001	0.042	0.001	0.021	0.019
	(0.005)	(0.058)	(0.005)	(0.054)	(0.034)
N	236	236	324	305	214
42 Construction trade services					
Labour	0.907	0.457	0.911	0.438	0.869
	(0.033)**	(0.078)**	(0.034)**	(0.081)**	(0.037)**
Capital	0.263	0.094	0.247	0.092	0.233
	(0.021)**	(0.033)**	(0.021)**	(0.032)**	(0.052)**
Age	–0.002	–0.014	–0.003	–0.036	–0.003
	(0.003)	(0.027)	(0.003)	(0.028)	(0.024)
N	452	452	592	579	410
Wholesale trade					
45 Basic material wholesaling					
Labour	0.768	0.763	0.811	0.789	0.756
	(0.043)**	(0.107)**	(0.048)**	(0.111)**	(0.052)**
Capital	0.211	0.081	0.257	0.003	0.237
	(0.031)**	(0.030)**	(0.033)**	(0.033)	(0.061)**
Age	0.011	0.008	0.013	0.009	–0.011
	(0.003)**	(0.020)	(0.003)**	(0.023)	(0.015)
N	496	496	584	574	413

(Continued on next page)

Table 9.1 (continued)

ANZSIC/industry	Balanced panel (1) OLS	Balanced panel (2) Within	Full sample (3) OLS	Full sample (4) Within	Full sample (5) OP
46 Machinery and motor vehicle wholesaling					
Labour	0.996	0.595	1.017	0.558	0.988
	(0.028)**	(0.068)**	(0.027)**	(0.066)**	(0.031)**
Capital	0.143	0.085	0.164	0.076	0.167
	(0.020)**	(0.024)**	(0.020)**	(0.023)**	(0.034)**
Age	0.002	0.010	0.004	0.027	0.003
	(0.002)	(0.015)	(0.002)+	(0.016)+	(0.016)
N	848	848	1054	1039	742
47 Personal and household good wholesaling					
Labour	0.886	0.584	0.858	0.519	0.760
	(0.032)**	(0.064)**	(0.033)**	(0.063)**	(0.036)**
Capital	0.181	0.104	0.235	0.088	0.199
	(0.022)**	(0.023)**	(0.022)**	(0.022)**	(0.058)**
Age	–0.002	–0.008	0.001	0.009	0.026
	(0.002)	(0.016)	(0.002)	(0.016)	(0.017)
N	644	644	794	775	554
Retail trade					
51 Food retailing					
Labour	0.657	0.063	0.685	0.086	0.615
	(0.036)**	(0.093)	(0.032)**	(0.079)	(0.035)**
Capital	0.378	0.298	0.353	0.287	0.334
	(0.029)**	(0.036)**	(0.025)**	(0.033)**	(0.059)**
Age	0.018	0.003	0.018	0.021	0.007
	(0.004)**	(0.027)	(0.003)**	(0.026)	(0.013)
N	344	344	469	438	314
52 Personal and household good retailing					
Labour	0.805	0.212	0.790	0.237	0.793
	(0.032)**	(0.069)**	(0.037)**	(0.088)**	(0.041)**
Capital	0.278	0.145	0.313	0.177	0.257
	(0.022)**	(0.024)**	(0.024)**	(0.031)**	(0.081)**
Age	0.001	0.032	0.007	0.049	–0.006
	(0.003)	(0.020)	(0.003)*	(0.026)+	(0.017)
N	480	480	625	598	428

(Continued on next page)

Table 9.1 (continued)

	Balanced panel		*Full sample*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
ANZSIC/industry	*OLS*	*Within*	*OLS*	*Within*	*OP*
53 Motor vehicle retailing and services					
Labour	0.977	0.388	0.975	0.434	0.942
	(0.023)**	(0.074)**	(0.023)**	(0.071)**	(0.028)**
Capital	0.159	0.065	0.179	0.060	0.117
	(0.018)**	(0.023)**	(0.018)**	(0.021)**	(0.043)**
Age	0.001	–0.029	0.002	–0.02	–0.013
	(0.002)	(0.018)	(0.002)	(0.019)	(0.015)
N	532	532	612	602	437
57 Accommodation, cafés and restaurants					
Labour	0.954	0.586	0.910	0.477	0.799
	(0.035)**	(0.071)**	(0.033)**	(0.071)**	(0.037)**
Capital	0.273	0.109	0.289	0.105	0.270
	(0.022)**	(0.025)**	(0.020)**	(0.027)**	(0.072)**
Age	0.000	0.021	0.000	0.030	0.003
	(0.002)	(0.021)	(0.002)	(0.022)	(0.025)
N	472	472	651	617	437
Property and business services					
77 Property services					
Labour	0.838	0.460	0.865	0.437	0.867
	(0.042)**	(0.103)**	(0.033)**	(0.083)**	(0.035)**
Capital	0.281	0.137	0.306	0.161	0.226
	(0.025)**	(0.052)**	(0.022)**	(0.042)**	(0.080)**
Age	–0.004	0.028	–0.002	0.019	–0.053
	(0.004)	(0.044)	(0.004)	(0.038)	(0.040)
N	376	376	593	562	388
78 Business services					
Labour	0.948	0.471	0.936	0.507	0.858
	(0.022)**	(0.048)**	(0.020)**	(0.044)**	(0.022)**
Capital	0.179	0.067	0.186	0.078	0.156
	(0.014)**	(0.017)**	(0.012)**	(0.016)**	(0.028)**
Age	0.003	–0.007	0.004	–0.008	–0.010
	(0.002)	(0.018)	(0.002)	(0.017)	(0.022)
N	1252	1252	1585	1535	1093

(Continued on next page)

Table 9.1 (continued)

ANZSIC/industry	Balanced panel (1) OLS	(2) Within	Full sample (3) OLS	(4) Within	(5) OP
Cultural and recreational services					
91 Motion picture, radio and television services					
Labour	0.429	0.271	0.447	0.259	0.592
	(0.077)**	(0.119)*	(0.074)**	(0.113)*	(0.087)**
Capital	0.504	0.179	0.515	0.192	0.096
	(0.047)**	(0.083)*	(0.042)**	(0.069)**	(0.104)
Age	–0.002	0.007	0.003	0.023	–0.028
	(0.008)	(0.069)	(0.008)	(0.065)	(0.048)
N	128	128	170	162	114
93 Sport and recreation					
Labour	0.223	0.530	0.564	0.424	0.564
	(0.162)	(0.230)*	(0.096)**	(0.147)**	(0.136)**
Capital	0.499	0.178	0.368	0.142	0.315
	(0.085)**	(0.122)	(0.054)**	(0.066)*	(0.195)
Age	–0.024	–0.164	–0.018	–0.149	–0.012
	(0.012)*	(0.084)+	(0.007)*	(0.055)**	(0.036)
N	52	52	121	110	69
Personal and other services					
95 Personal services					
Labour	0.898	0.370	0.717	0.378	0.645
	(0.041)**	(0.110)**	(0.046)**	(0.070)**	(0.054)**
Capital	0.270	0.175	0.373	0.183	0.341
	(0.028)**	(0.046)**	(0.031)**	(0.043)**	(0.068)**
Age	–0.003	0.025	0.008	0.047	0.002
	(0.003)	(0.030)	(0.004)+	(0.032)	(0.022)
N	280	280	383	367	262

[a] Standard errors in parentheses (bootstrapped s.e. reported for capital and age coefficients in column 5). + significant at 10 per cent. * significant at 5 per cent. ** significant at 1 per cent.

References

ABS (Australian Bureau of Statistics) 2000, *Business Longitudinal Survey, Confidentialised Unit Record File*, cat. no. 8141.0.30.001, Canberra.

Andrews, D.W.K. 1991, 'Asymptotic normality of series estimators for nonparametric and semiparametric regression models', *Econometrica*, 59(2), pp. 307–45.

Blalock, G. and Gertler, P. 2003, *Learning from Exporting: Evidence from Indonesia*, National Bureau of Economic Research Working Paper.

Engle, R.F., Granger, J.R. and Weiss, A.1986, 'Semiparametric estimates of the relation between weather and electricity sales', *Journal of the American Statistical Association,* 81(394), pp. 310–20.

Ericson, R. and Pakes, A. 1995, ‘Markov-perfect industry dynamics: a framework for empirical work’, *Review of Economic Studies*, 62(1), pp. 53–82.

Hopenhayn, H. and Rogerson, R., 1993, ‘Job turnover and policy evaluation: a general equilibrium analysis’, *Journal of Political Economy*, 101(5), pp. 915–38.

Levinsohn, J. and Petrin, A. 1999, *When Industries Become More Productive, Do Firms? Investigating Productivity Dynamics*, National Bureau of Economic Research Working Paper.

—— 2003, ‘Estimating production functions using inputs to control for unobservables’, *Review of Economic Studies*, 70, pp. 317–41.

Marschak, J. and Andrews, W.H. 1944, ‘Random simultaneous equations and the theory of production’, *Econometrica*, 12(3/4), pp. 143–205.

Muendler, Marc-Andreas 2004, ‘Estimating production functions when productivity change is endogenous’, University of California Working Paper, San Diego.

Newey, W.K., 1994, ‘The asymptotic variance of semiparametric estimators’, *Econometrica*, 62(6), pp. 1349–82.

Olley, G.S. and Pakes, A. 1996, ‘The dynamics of productivity in the telecommunications equipment industry’, *Econometrica*, 64(6), pp. 1263–97.

Pakes, A. 1994, ‘Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interactions’ in Sims, C (ed.), *Advances in Econometrics*, Cambridge University Press, Cambridge, Massachusetts.

Pavcnik, N. 2002, ‘Trade liberalization, exit, and productivity improvement: evidence from Chilean plants’, *Review of Economic Studies*, 69(1), pp. 245–76.

Robinson, P.M. 1988, ‘Root-N-consistent semiparametric regression’, *Econometrica*, 56(4), pp. 931–54.

Rogers, M. 1998. *Productivity in Australian Enterprises: Evidence from the ABS Growth and Performance Survey*, Melbourne Institute Working Paper, University of Melbourne.

Wedervang, F. 1965, *Development of a Population of Industrial Firms*, Scandinavian University Books, Oslo, Norway.

Will, L. and Wilson, H. 2001, *Tricks and Traps of the Business Longitudinal Survey*, Productivity Commission Staff Working Paper, Melbourne.