THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LV September 2007 No. 3

ROBUSTNESS OF PRODUCTIVITY ESTIMATES*

JOHANNES VAN BIESEBROECK†

Researchers interested in estimating productivity can choose from an array of methodologies, each with its strengths and weaknesses. We compare the robustness of five widely used techniques, two nonparametric and three parametric: in order, (a) index numbers, (b) data envelopment analysis (DEA), (c) stochastic frontiers, (d) instrumental variables (GMM) and (e) semiparametric estimation. Using simulated samples of firms, we analyze the sensitivity of alternative methods to the way randomness is introduced in the data generating process. Three experiments are considered, introducing randomness via factor price heterogeneity, measurement error and differences in production technology respectively. When measurement error is small, index numbers are excellent for estimating productivity growth and are among the best for estimating productivity levels. DEA excels when technology is heterogeneous and returns to scale are not constant. When measurement or optimization errors are nonnegligible, parametric approaches are preferred. Ranked by the persistence of the productivity differentials between firms (in decreasing order), one should prefer the stochastic frontiers, GMM, or semiparametric estimation methods. The practical relevance of each experiment for applied researchers is discussed explicitly.

I. INTRODUCTION

ANY SAMPLE OF FIRMS, EVEN WITHIN A SINGLE INDUSTRY, will typically display a tremendous amount of heterogeneity. This has been known at least since the work of Baily, Hulten and Campbell [1992] and Bartelsman and Dhrymes [1998], who document the distribution and evolution of productivity using data from the U.S. Census of Manufacturers. Similar results from a number of developing countries can be found in Roberts and Tybout [1996].[1] Despite this high degree of heterogeneity, most studies that estimate productivity or

*I would like to thank the editor, an anonymous referee, and Dan Ackerberg, Stephen Bond, Garth Frazer, Mel Fuss, Robert Gagné, Robert McMillan, Marc Melitz, Ariel Pakes, Peter Reiss, Chad Syverson, Frank Wolak, and participants at CIRANO, the NBER Summer Institute, SITE, and the Econometric Society World Congress for comments, All remaining errors are my own. Financial support from SSHRC and the Connaught Fund is gratefully acknowledged.

†Author's affiliation: Department of Economics, University of Toronto; and NBER; 140 St. George Street, Suite 707, Toronto, Ontario M5S 3G6, Canada.
e-mail: johannes.vanbiesebroeck@utoronto.ca

[1] Bartelsman and Doms [2000] summarize the results from the literature drawing on findings from many different countries.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

production functions rely on relatively simple models that ignore most of the relevant heterogeneity. For instance, nonparametric models tend to be deterministic and parametric methods tend to assume a homogeneous production technology with the same input trade-off for all firms.[2]

Using a Monte Carlo study, this paper investigates the sensitivity of five popular estimation methods for productivity to different sources of heterogeneity among firms. The first contribution of the paper, in Sections II and III, is to present the different estimators in a consistent framework and highlight their differences and similarities. The first two methods, index numbers and data envelopment analysis (DEA), are very flexible in the specification of technology, but do not allow for unobservables. Three parametric methods that differ in the way they deal with the problem of simultaneity of inputs and unobserved productivity are evaluated as well. Stochastic frontiers make explicit distributional assumptions on the unobserved productivity; the GMM system estimator (GMM-SYS) relies on instrumental variables; and the semiparametric estimator inverts the investment function nonparametrically to obtain an observable expression for productivity.

We simulate samples of hypothetical firms, introducing heterogeneity in three different ways in the firms' first order condition for labor. The sources of variation considered are (i) factor price heterogeneity, which in the model can also be interpreted as optimization errors or high frequency productivity shocks, (ii) measurement error, and (iii) differences in production technology. Estimated productivity levels and growth rates, as well as input elasticities, are compared with the true values.[3]

Fundamentally, the objective of productivity measurement is to identify output differences that cannot be explained by input differences. To perform this task, one needs to observe inputs and outputs accurately and control for the input substitution that the production technology allows. Methodologies differ by the mix of statistical techniques and economic assumptions they employ to control for input substitution and for unobservable differences. The three different experiments considered in the paper map into different situations that applied researchers often encounter. For example, depending on the source of the data and the type of observations the sample covers, a researcher will have different priors on the importance of measurement error or heterogeneity in technology. The relative performance of the different methods should be informative in choosing

[2] It should be noted that recent innovations generalize some approaches considerably, but they are not widely used. A growing literature explores stochastic nonparametric frontiers; Feenstra and Reinsdorf [2003] illustrate how one can justify standard errors on an index number for TFP growth.

[3] Productivity differences are restricted to be Hicks-neutral, but different processes for their evolution are considered.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

the most appropriate estimator for the situation. Even though the simulated experiments represent only a fraction of the situations one might face, we can draw some lessons as to when we expect each methodology to be particularly suitable. We summarize these here:

Unless one expects the data to be subject to a lot of measurement error, the index number approach produces consistently accurate productivity growth estimates. When firms are likely to employ different technologies, productivity level estimates are among the most robust as well. DEA is the preferred estimator for productivity levels if technology is likely to vary across firms and scale economies are not constant. Examples would include situations where one has to pool firms from very different industries, at different stages in their life cycle or operating in countries with different levels of development.

Given the well-known simultaneity problem between inputs and unobserved productivity, estimating a production function by least squares (OLS) is generally not advisable. Its disadvantages are less acute for productivity growth than for productivity levels. When one has good reason to believe that productivity differences are constant over time, that output is measured accurately, and that observations share the same technology, stochastic frontier estimation produces accurate productivity level estimates. With a lot of measurement error or some technological heterogeneity, the GMM-SYS estimator provides the most robust productivity level and growth estimates of the parametric methods. Even in the absence of these forms of heterogeneity, it provides reliable results if at least part of the productivity difference is constant over time.[4] When firms are subject to idiosyncratic productivity shocks that are not entirely transitory, the semiparametric proxy estimator will exploit the firm's knowledge about these shocks.

In the next section, we give some background on productivity measurement, and in Section III the different methodologies are introduced. An attempt is made to present the basic idea and distinctive features for each estimator concisely and within a consistent framework. References to the literature are included for each estimator. Section IV briefly introduces the model of a representative firm used to generate the data. The setup of the experiments, along with some general findings, are described in Section V, while a more detailed discussion of results, organized by estimator, is given in Section VI. Lessons to take away from these exercises are summarized at the end.

[4] It should be noted that, as it includes firm fixed effects, the GMM-SYS estimator is vulnerable to the problem of downward biased input elasticities if measurement error becomes severe. See Griliches and Hausman [1986] for a discussion. In the simulation results (below), this was rarely an issue, but a comparative analysis with Colombian data in Van Biesebroeck [2003b] suggests the problem can be large.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

II. ESTIMATING PRODUCTIVITY

Economists studying the relative performance of firms have tended to avoid accounting-based measures, such as observed costs or profits.[5] Costs are difficult to decompose into fixed and variable components (which play very different roles in our models) and they exclude opportunity costs. Profits are influenced by many factors other than efficient production: the taxation regime, financial structure, or price setting power, among others. Furthermore, depreciation and valuation assumptions for assets, which are rarely economically motivated, have a large impact on profits and the measured cost of capital.

In the absence of price setting power at the firm level, information on deflated sales or value added and input levels can be used in conjunction with a production function to calculate a firm's productivity.[6] This approach to quantify relative performance has been very popular in the literature and we will evaluate different ways it has been put into practice. It should be noted that when firms have market power, but information on prices is not available (the typical case), one needs a model of demand to isolate quantity changes from price changes.[7]

One firm is more productive than another if it can produce the same output with fewer inputs or if it produces more output from the same inputs. Similarly, a firm has experienced positive productivity growth if outputs have increased more than inputs or inputs have decreased more than outputs. To compare two production plans where one uses more of a first input and the other more of a second input, it becomes necessary to specify a transformation function that links inputs to output. Input substitution possibilities are determined by the technology and each productivity measure is only defined with respect to that specific production technology.

We will discuss three different approaches to control for input substitution embodied in the technology using the following simple model of production:

$$(1) \qquad Q_{it} = A_{it} F_{(it)}(X_{it}).$$

A_{it} is the unobservable productivity term, which differs across firms and time periods. The deterministic part of the model—the technology—is

[5] The reluctance to use accounting data has become much more widespread in the last 20 years. Chapter 16 in the second volume of the *Handbook of Industrial Organization* by Richard Schmalensee lists a number of problems with accounting data.

[6] One can also rely on the dual representation of technology, the cost function, and estimate with cost shares and input prices.

[7] For example, to measure the performance of firms that produce differentiated products, Katayama, Lu and Tybout [2005] use functional form and equilibrium assumptions to recover marginal cost estimates and evaluate production efficiency in such a situation. Other studies have used the approach in Klette and Griliches [1996], relying on the CES demand system to work out an adjustment for the productivity estimates.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

represented by the production function $F(\cdot)$, which takes a vector of inputs X_{it} as arguments. This limits the analysis in a number of ways: only the single output case is considered, because most studies use value added or sales as the output measure; all productivity differences are Hicks-neutral, as the most widely used Cobb-Douglas production function cannot identify factor-bias in technological change; and the productivity measure used is output-based, i.e., it measures how much extra output a firm produces relative to another firm, conditional on its (extra) input use.[8]

Rearranging the production function as

$$\ln\left(\frac{A_{it}}{A_{j\tau}}\right)_k = \ln\frac{Q_{it}}{Q_{j\tau}} - \ln\frac{F_k(X_{it})}{F_k(X_{j\tau})} \tag{2}$$

underscores the point that productivity is intrinsically a relative concept. If the technology is allowed to vary across observations one has to be explicit as to which technology underlies the comparison, hence the k subscript ($k \in \{it, j\tau\}$). Multilateral productivity level comparisons can be achieved if average productivity across all plants in the industry is used in the denominator. In practice, $\ln A_{it} - \overline{\ln A_i}$ is most often used, averaging the logarithms, and for comparability we follow this practice.[9] The calculation of the last term in (2)—the ratio of input aggregators—distinguishes the different methods. It is here that different literatures employ different assumptions; we distinguish three approaches:

(i). Index numbers assume that firms make input choices optimally, with some restrictions on the nature of the technology differences. An exact Taylor expansion of the production function can be calculated using the information on the slope of the $F(.)$ function contained in the factor price ratios.

(ii). Nonparametric frontiers, such as DEA, construct for each observation a linear combination of all other observations (normalized by output) for explicit comparison. Weights are chosen to maximize productivity for the unit considered. The same 'technology' is used in the numerator and denominator of the input aggregator ratio in (2), but for each unit, a separate comparison is performed, i.e., technology is still firm-specific.

(iii). Parametric methods explicitly assume a functional form for the technology, the $F(.)$ function, to be estimated econometrically. Random errors can be introduced straightforwardly in the model, multiplying the right-hand side of (1).

[8] An input-based measure would calculate what the minimum input requirement for one firm is to produce the same output as another firm. Under constant returns to scale, both measures coincide.

[9] In a regression framework, this amounts to including industry-year fixed effects in a regression with unnormalized log-productivity as dependent variable.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

The differences between the approaches can be illustrated graphically. To measure productivity differences, the input-output combinations are decomposed into shifts along a production frontier and shifts of the frontier itself. In the top panel (a) of Figure 1, two production plans (two time periods for one firm or two different firms at one point in time), P_0 and P_1, are compared in input space. Inputs are normalized by output and the technology is represented by the unit isoquant. Part of the difference, from P_0 to A, is a shift along the frontier, exploiting the input substitution the technology allows. The remainder of the difference, from A to P_1, is an

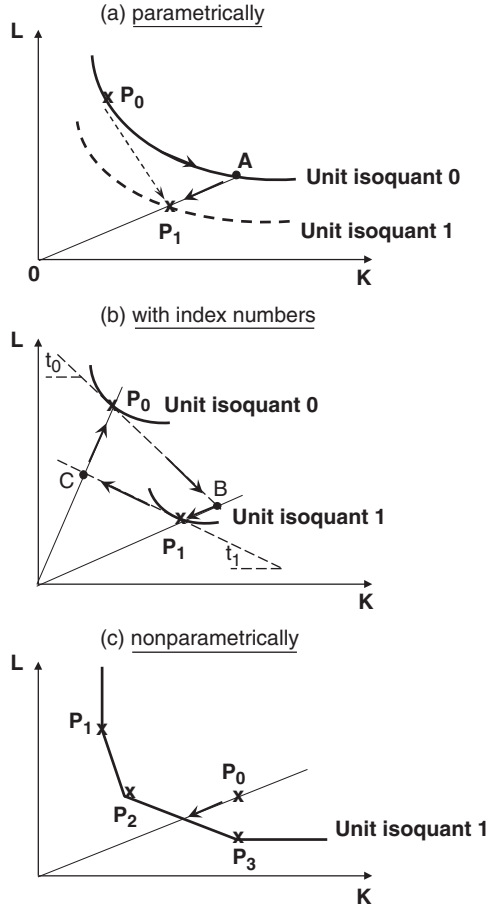


Figure 1
Distinguishing a shift along the frontier from a shift of the frontier (a) parametrically (b) with index numbers (c) nonparametrically

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

actual shift of the frontier, which is counted as productivity growth (from period 0 to 1) or as a productivity difference (between firms 0 and 1).[10]

The first approach (i) relies on index number theory and optimizing behavior. If the first order conditions for input choices hold, the factor price ratio will equal the slope of the input isoquant at the optimum, which determines local input substitution possibilities. For example, production plan P_0 in the middle panel (b) of Figure 1 uses more labor than P_1. Using information contained in the relative factor price ratio for either production plan (t_0 or t_1), we can control for input differences without having to estimate the shape of the isoquant. Instead, we use the average labor share in costs for both production plans to control for the input substitution possibilities. As discussed in Section III(i), this productivity measure corresponds to the geometric mean of the measures using either production technology (B to P_1 using the technology of plan 0 or C to P_0 using technology 1).

The second approach (ii) is entirely nonparametric. It constructs a piecewise linear isoquant to maximize the productivity estimate for P_0, without allowing any other observation to lie below the isoquant. The relevant section of the isoquant, connecting P_2 and P_3 in the bottom panel (c) of Figure 1, implicitly defines relative weights for labor and capital. Weights are chosen to minimize the distance from P_0 to the isoquant. For each observation, different weights are calculated using linear programming techniques, as discussed in Section III(ii). The productivity of P_0 relative to the frontier indicates how much its inputs can be reduced for the plan to equal a linear combination of the two comparison plans (P_2 and P_3).[11]

The parametric approach (iii) to measure the shift relies on functional form assumptions about technology (i.e., on the input trade-off) to estimate the exact shape of the unit isoquant. Most studies use a simple functional form[12] and virtually all applications restrict parameters to be equal across firms.[13] One reason for the restrictiveness in the deterministic portion of the model is that unobserved productivity differences make input choices

[10] Throughout, we assume firms are only producing technologically efficient production plans; i.e., while we allow allocative inefficiency (except with the index number approach), we rule out technical inefficiencies. To some extent, this is merely definitional as the feasible output for each firm is shifted in or out by its productivity. In contrast, the stochastic frontier literature has traditionally interpreted all firms with below first-best productivity as 'inefficient.'

[11] Inputs are rescaled in proportion, along a ray through the origin. The output-based definition of productivity rescales output, but this cannot be represented in input space.

[12] Even though a large literature investigates the impact of functional form assumptions in the parametric framework (see Berndt and Khaled [1979] for an early example and Gagné and Ouellette [1998] for a Monte Carlo investigation) the simple Cobb-Douglas production function remains extremely popular.

[13] A notable exception is the random coefficients model in Mairesse and Griliches [1990], but they find very volatile and unstable technology parameters, which seem unrelated to any observable plant differences. A more recent paper by Klette [1999] achieved more stable estimates with random coefficients, but only by introducing variations in price-cost margins, which in turn makes deflated sales a problematic measure of output.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

endogenous.[14] Estimating flexible functional forms is likely to exacerbate this simultaneity problem. It also explains the popularity of nonparametric approaches which are able to calculate productivity without estimating the exact input trade-off. We implement several estimators that control for the simultaneity, discussed in Sections III(iii) to III(vi).

The different methodologies are introduced briefly in the following section. Appendix B describes two more estimation methods that explicitly purge the estimates of random errors in the production function. Estimators from different literatures are presented in a unified framework with references to the literature.[15]

III. ESTIMATION METHODS

III (i). *Index Numbers (IN)*

Index numbers provide a theoretically motivated aggregation method for inputs and outputs, while remaining fairly agnostic as to the exact shape of the production technology. For example, Solow [1957] used the following formula for total factor productivity (TFP) growth in his original growth accounting exercise:

$$\begin{aligned}\ln A_{it}^{IN} - \ln A_{it-1}^{IN} \;=\; & \ln \frac{Q_{it}}{Q_{it-1}} - \left(\frac{s_{it}^{L} + s_{it-1}^{L}}{2}\right) \ln \frac{L_{it}}{L_{it-1}} \\ & - \left(1 - \frac{s_{it}^{L} + s_{it-1}^{L}}{2}\right) \ln \frac{K_{it}}{K_{it-1}}. \end{aligned} \tag{3}$$

All variables on the right-hand side of (3) are observable: output, inputs, and the fraction of the wage bill in output (s^L).

While Solow rationalized the TFP formula as a discrete approximation to the continuous divisia index, Caves, Christensen and Diewert [1982a] show that this Törnqvist index has a much broader validity. It exactly equals the geometric mean of the two Malmquist productivity indices—using either production plan's technology—if the technology is characterized by a translog (output) distance function.[16] Some heterogeneity in production

[14] This has been a well-known problem since the early work of Marschak and Andrews [1944] and Zellner, Kmenta and Drèze [1966].

[15] For a more detailed exposition, the reader is referred to the working paper version, Van Biesebroeck [2004].

[16] The production technology can be equivalently represented by the input requirements function $x_1 = g^i(q, \tilde{x})$or the output distance function $d^i(q, x) \equiv \min_\delta \{\delta : g^i(q/\delta, \tilde{x}) \leqslant x_1\}$. The latter measures the amount of output deflation (or inflation) needed for any production plan to lay on the transformation function of technology i; by definition $d^i(q^i, x^i) = 1$. The Malmquist firm i output-based productivity index comparing production plan j to i is the ratio of two output distance functions: $M^i(x^j, x^i, q^j, q^i) \equiv d^i(q^j, x^j)/d^i(q^i, x^i)$. The higher M^i, the more productive plan j is relative to plan i, using the input trade-off embodied in technology i. Equivalently, the firm j based Malmquist index is $M^j(x^j, x^i, q^j, q^i) \equiv d^j(q^j, x^j)/d^j(q^i, x^i)$. If the distance function is translog, the Törnqvist index comparing plans j and i (t and $t-1$ in equation (3)) exactly equals $\sqrt{M^i M^j}$, see Theorem 3 in Caves *et al.* [1982a].

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

technology across observations is allowed; only the coefficients on the second order terms have to be equal for the two units compared. The relative weights on the input factor ratios (labor versus capital shares in output) exploit the information on the slope of the input trade-off contained in the factor prices.

For multilateral productivity level comparisons, Caves, Christensen and Diewert [1982b] propose an index where each firm is compared to a hypothetical firm, with average log output ($\overline{\ln Q}$), labor share ($\overline{s^L}$), etc. The productivity level of firm i at time t is

$$\ln A_{it}^{IN} - \overline{\ln A_t}^{IN} = (\ln Q_{it} - \overline{\ln Q_t}) - \tilde{s}_{it}(\ln L_{it} - \overline{\ln L_t}) - (1 - \tilde{s}_{it})(\ln K_{it} - \overline{\ln K_t}) \quad (4)$$

with $\tilde{s}_{it} = \frac{s_{it}^L + \bar{s}_t^L}{2}$. This yields bilateral comparisons that are transitive, and still allows for technology that is firm-specific.

For the theory to apply, i.e., to enable the last term in equation (2) to be calculated from observables without having to estimate the production function, a number of assumptions have to be satisfied. The most important ones are perfect competition in output and input markets, optimizing behavior by firms, and the absence of measurement error. While it is not strictly necessary to assume constant returns to scale, one would need outside information on scale economies to control for them in the productivity estimates. To absorb the scale effect, an additional factor has to be added to (subtracted from) equation (3) for each input if returns to scale are decreasing (increasing): see equation (8) in Van Biesebroeck [2004]. Estimating scale economies parametrically or obtaining information on the cost of capital would be sufficient, but we will follow common practice and limit attention to the constant returns to scale case.

The main advantages of the index number approach are the straightforward computations (no estimation is required), the ability to handle multiple outputs and many inputs, and the flexible and heterogeneous production technology it allows. Technical change can be factor-biased and the only separability assumption is between outputs and inputs. The main disadvantages are the deterministic nature of the model and the necessary assumptions on firm behavior and market structure.[17] It is impossible to deal with measurement errors or outliers, except for some *ad hoc* data trimming.

III (ii). *Data Envelopment Analysis (DEA)*

Data envelopment analysis, also called non-parametric frontier estimation, dates back to Farrell [1957]. It was operationalized by Charnes, Cooper and

[17] Adjustments exist for regulated firms, non-competitive output markets and temporary equilibrium, but they either involve estimating some structural parameters or are more data intensive.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

Rhodes [1978] and an overview of the method with applications can be found in Seiford and Thrall [1990]. No particular production function is assumed. Instead, efficiency is defined as the ratio of a linear combination of outputs over a linear combination of inputs. Weights on inputs (u_l, u_k) and output (v_q) are chosen directly to maximize efficiency (productivity) for the unit under consideration, indicated by θ.[18] Observations that are not dominated are labeled 100% efficient. Domination occurs when another firm, or a linear combination of other firms, produces more of all outputs with the same input aggregate, using the same weights to aggregate inputs.

A linear programming problem is solved separately for each observation. For unit (firm-year) 1, in the single-output case, the problem amounts to

$$
\begin{aligned}
&\max_{v_q,u_l,u_k} \ \theta_1 = \frac{v_q Q_1 + v^*}{u_l L_1 + u_k K_1} \\
(5) \qquad &\text{subject to} \ \ \frac{v_q Q_i + v^*}{u_l L_i + u_k K_i} \leqslant 1 \qquad\qquad i = 1 \ \ldots \ N \\
&\qquad\qquad v_q, u_l + u_k > 0, \ u_l, u_k \geqslant 0 \\
&\qquad\qquad v^* \geqslant 0 \qquad\qquad (v^* = 0 \text{ for constant returns to scale})
\end{aligned}
$$

In addition to sign restrictions on the weights, the only constraints are that the efficiency of all firms i cannot exceed 100% when the same weights are applied to them. For the problem to be well-defined, a normalization is needed; doubling all input and output weights in (5) would leave productivity estimates unchanged. The usual choice is to put $u_l L_1 + u_k K_1$ equal to one. v^* is a complementary slack variable to allow for variable returns to scale. Under constant returns to scale ($v^* = 0$) the production frontier is a ray through the origin in output-aggregate input space; under variable returns to scale, it is the piecewise linear envelope of all production plans.[19]

The efficiency measure θ_i can be interpreted as the productivity difference between unit i and the most productive unit: $\theta_i = A_i / A_{max}$. Estimates of productivity levels and growth rates comparable to the ones obtained with other methodologies can be defined as

$$
(6) \qquad \ln A_{it}^{DEA} - \overline{\ln A}_t^{DEA} = \ln \theta_{it} - \frac{1}{N_t} \sum_{j=1}^{N_t} \ln \theta_{jt}
$$

$$
(7) \qquad \ln A_{it}^{DEA} - \ln A_{it-1}^{DEA} = \ln \theta_{it} - \ln \theta_{it-1}.
$$

The θ coefficients are estimated on the full sample that includes all firm-years as separate observations. The transformations do not change the ranking of firms, only the absolute productivity levels and growth rates.

[18] Multiple outputs would be aggregated linearly.

[19] In practice, most estimation algorithms solve the dual problem, where θ_1 is chosen directly. The other choice variables in the dual problem are weights on all firms to form linear combinations of output and inputs and construct an explicit comparison production plan.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

The main advantage of DEA is the absence of functional form or behavioral assumptions. The underlying technology is left entirely unspecified and allowed to vary across firms. The linear aggregation is natural in an activities analysis framework. Each firm is considered a separate process that can be combined with others to replicate the production plan of the unit under investigation. On the other hand, the flexibility in weighting has drawbacks: Each firm with the highest ratio for any output-input combination is 100% efficient. Under variable returns to scale, each firm with the lowest input or highest output level in absolute terms is also fully efficient. The most widely used implementations are not stochastic, making estimates sensitive to outliers. Because each observation is compared to all others, measurement error for a single firm can affect all productivity estimates.

III (iii). *Parametric Estimation: Least Squares (OLS)*

The parametric methods assume the same input trade-off and returns to scale for all firms. The functional form assumption on the production function concentrates all heterogeneity in the productivity term, but the explicitly stochastic framework is likely to make estimates less susceptible to measurement error. We follow most of the literature by estimating a Cobb-Douglas production function in logarithms,

$$q_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \omega_{it} + \varepsilon_{it}. \tag{8}$$

ω_{it} represents a productivity difference known to the firm, but unobservable to the econometrician; ε_{it} captures other sources of i.i.d. error.

Consistent estimation of the input parameters faces an endogeneity problem, first discussed by Marschak and Andrews [1944]. Firms choose inputs knowing their own level of productivity and a least squares regression of output on inputs will give inconsistent estimates of the production function coefficients. We use the OLS estimates simply as a benchmark and implement three estimators that explicitly address the endogeneity problem. The stochastic frontiers in Section III(iv) make explicit distributional assumptions on the unobserved productivity term; the GMM-SYS estimator in Section III(v) relies on instrumental variables; the semiparametric estimator in Section III(vi) inverts the investment function nonparametrically to obtain an expression for productivity.

Because the input aggregator is the same for all time periods and firms, calculating productivity levels and growth rates is straightforward:

$$\ln A_{it}^{OLS} - \overline{\ln A_t}^{OLS} = (q_{it} - \overline{q_t}) - \hat{\alpha}_l^{OLS}(l_{it} - \overline{l_t}) - \hat{\alpha}_k^{OLS}(k_{it} - \overline{k_t}) \tag{9}$$

$$\begin{aligned} \ln A_{it}^{OLS} - \ln A_{it-1}^{OLS} =& (q_{it} - q_{it-1}) - \hat{\alpha}_l^{OLS}(l_{it} - l_{it-1}) \\ & - \hat{\alpha}_k^{OLS}(k_{it} - k_{it-1}). \end{aligned} \tag{10}$$

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

To obtain a clean estimate of ω_{it} one should subtract an estimate for the difference in error terms from the right-hand sides: $(\varepsilon_{it} - \overline{\varepsilon_t})$ in (9) and $(\varepsilon_{it} - \varepsilon_{it-1})$ in (10). Generally this is not possible and is ignored because $E(\varepsilon_{it}) = 0$. For two alternative estimators introduced in Appendix B (SF2 and OP2), different formulas are used to purge the random noise from the productivity estimates.

III (iv). *Parametric Estimation: Stochastic Frontiers (SF)*

The stochastic frontier literature uses assumptions on the distribution of the unobserved productivity component to separate it from the random error. The method is credited to Aigner, Lovell and Schmidt [1977] and Meeusen and van den Broeck [1977] who model productivity as a stochastic draw from the negative of, respectively, an exponential or half-normal distribution. Estimation is usually with maximum likelihood. In the production function (8), the term ω_{it} is weakly negative and is interpreted as the inefficiency of firm i at time t relative to the best-practice production frontier. An alternative interpretation is that the firm-specific production function lies ω_{it} below the most productive observation in the sample.

Initially developed to measure productivity in a cross section of firms, the model was generalized for panel data in a number of ways. Battese and Coelli [1992] provide the most straightforward, but also the most restrictive generalization, modeling the inefficiency term as

$$(11) \qquad \omega_{it} = -e^{-\eta(t-T)}\omega_i \; with \; \omega_i \sim N^+(\gamma, \sigma^2).$$

The relative productivity of each firm (ω_i) is a time-invariant draw from a truncated normal distribution. Inefficiency increases (decreases) deterministically over time if η is positive (negative) at the same rate for all firms. If firms are only observed once, making strong assumptions is the only possibility to separate the productivity component from the random error. Panel data contain more information on each firm and allow identification under weaker assumptions. More flexible generalizations are discussed in Van Biesebroeck [2004].[20]

While it is customary in this literature to calculate technical inefficiency as $TE_{it} = E(e^{\omega_{it}}|\omega_{it} + \varepsilon_{it})$, for comparability with the other methods, we calculate the expected value of log-productivity. This boils down to the same formulas as before, (9) and (10), but using different input coefficient estimates.[21]

[20] The estimator by Cornwell, Schmidt and Sickles [1990] (see Appendix B) is one example. It uses a flexible parameterization of the firm effects included in the production function using three coefficients per firm: a fixed effect and linear and quadratic firm-specific time trends.

[21] The best estimate of $E(\omega_{it}|\hat{\omega}_{it} + \hat{\varepsilon}_{it})$ is $\ln A_{it}^{SF} = \hat{\omega}_{it} + \hat{\varepsilon}_{it}$ if ω_{it} is independent of ε_{it}.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

An advantage of stochastic frontiers is that the deterministic part of the production function can be generalized easily to allow more sophisticated specifications, e.g., to incorporate factor-bias in technological change or flexible functional forms. The model described above and the alternative stochastic frontier implementation in the Appendix trade off flexibility in the characterization of productivity with estimation precision. The assumption that the relative position of firms never changes, in equation (11), only requires the estimation of three coefficients, but it will be less appropriate for longer samples. The alternative estimator is much less restrictive, but it uses $3 \times N$ degrees of freedom and it is the only estimator where consistency relies on asymptotics in the time dimension. With either approach, one might be uncomfortable with identification coming solely from distributional assumptions, which are especially restrictive in the first model.

III (v). *Parametric Estimation: Instrumental Variables (GMM)*

The general approach to estimate error component models of Blundell and Bond [1998] was first applied to production functions in Blundell and Bond [2000]. The productivity term is modeled as a firm fixed effect (ω_i) with an additive autoregressive component ($\omega_{it} = \rho\omega_{it-1} + \eta_{it}$). An additional term ε_{it} captures measurement error and transitory productivity shocks. Quasi-differencing the production function gives the estimating equation in its dynamic representation,

$$(12) \qquad q_{it} = \rho q_{it-1} + \alpha_l(l_{it} - \rho l_{it-1}) + \alpha_k(k_{it} - \rho k_{it-1}) + \omega'_i + \underbrace{(\eta_{it} + \varepsilon_{it} - \rho\varepsilon_{it-1})}_{\varepsilon_{it}}.$$

There is still a need for moment conditions to provide instruments, because the inputs will be correlated with the composite error ε_{it} through η_{it}, the annual productivity innovation, and the transitory productivity shocks ε_{it} and ε_{it-1}.

Estimating equation (12) in first-differenced form takes care of the firm fixed effects.[22] Three and more periods lagged inputs and output will be uncorrelated with $\Delta\varepsilon_{it}$ under standard exogeneity assumptions on the initial conditions. Three lags are necessary as the first differenced equation will contain errors as far back as ε_{it-2}. This error directly affects twice lagged output and firms will have based their input choices two periods ago on this shock. Additional lags can be used to increase precision.[23]

[22] Time dummies can be included in both the level and first differenced production function to absorb productivity growth that is common to all firms.

[23] Van Biesebroeck [2004] lists the exact form of the necessary assumptions and moment conditions.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

Blundell and Bond [1998] illustrate theoretically and with a practical application that these instruments can be weak. If one is willing to make the additional assumption that input changes are uncorrelated with the firm fixed effects and that the same holds for the initial output change, twice lagged first differences of the variables are valid instruments for the production function in levels. As before, the presence of the lagged error ε_{it-1} in equation (12) necessitates an additional lag on the instruments. In this case, including more distant lags does not provide additional information. The production function is estimated as a system combining the first differenced and level equations, each with the appropriate set of instruments. Productivity is again calculated using equations (9) and (10).

The GMM system estimator is flexible in generating instruments and one can test the overidentification restrictions if many instruments are used for the first differenced equation. This estimator allows for an autoregressive component to productivity, in addition to a fixed and an idiosyncratic component. Relative to the simple fixed effects estimator, it also uses the information contained in the levels, which is likely to help with measurement error (see Griliches and Mairesse [1998]). The major disadvantage is the need for a long panel: at least four time periods are required if we allow for transitory productivity shocks. If instruments are weak, the method risks underestimating the input elasticities, just like the fixed effects estimator.

III (vi). *Semiparametric Estimation (OP)*

The last method we consider was introduced by Olley and Pakes [1996] to estimate productivity effects of restructuring in the U.S. telecommunications equipment industry. Two assumptions are important. First, productivity—a state variable in the firm's dynamic problem—is assumed to follow a Markov process unaffected by the firm's control variables. Second, investment—one of the control variables of the firm—becomes part of the capital stock with a one period lag.

In a first step, coefficients on the variable inputs in the production function and the joint effect of all state variables on output are estimated. In our case, the former is just labor and the latter are capital and productivity. Under relatively mild conditions, investment is shown to be a monotonically increasing function of productivity. Inverting the investment equation nonparametrically provides an observable expression for productivity. This expression is used to substitute the unobserved productivity term (ω_{it}) from of the production function, hence allowing identification of the variable input elasticities. If market structure or factor prices change over time, the inversion can even be performed separately by year.[24]

[24] The methodology is more general than this exposition makes it appear. The basic idea is to use another observable decision by the firm to provide separate information on the unobserved

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

In a second step, the coefficients of the observable state variables (only capital in our case) are identified, exploiting the orthogonality of the quasi-fixed capital stock and the current innovation in productivity. A nonparametric term is included in the production function to absorb the impact of productivity, to the extent it was known to the firm when it chose investment in the last period. The second term included in equation (14) captures $E[\omega_{it}|\omega_{it-1}, t-1]$ and uses the results of the first stage.[25]

The estimating equations for the two steps are

$$q_{it} = \alpha_0 + \alpha_l l_{it} + \phi_t(i_{it}, k_{it}) + \varepsilon^1_{it} \tag{13}$$

$$q_{it} - \hat{\alpha}_l l_{it} = \alpha_k k_{it} + \psi(\hat{\phi}_{it-1} - \alpha_k k_{it-1}) + \varepsilon^2_{it}. \tag{14}$$

The functions ϕ_t and ψ are approximated nonparametrically by a fourth order polynomial or a kernel density. Productivity is calculated as in Olley and Pakes [1996] from (9) and (10) and these estimates will be labeled OP.[26]

A major advantage of this approach is the flexible characterization of productivity, only assuming that it evolves according to a Markov process. A potential weakness is the use of nonparametric approximations. The functions that are inverted are complicated mappings from states to actions, since they have to hold for all firms regardless of their size or competitive position. The prevalence of zero investment observations in many data sets weakens the mapping, prompting Levinsohn and Petrin [2003] to explore an alternative proxy for productivity (material input). Ackerberg, Caves and Frazer [2005] illustrate that the implicit assumptions required to identify the variable input coefficients are relatively restrictive. The latter paper generalizes the approach, estimating the elasticities on both the variable and quasi-fixed inputs in the second step.

IV. DATA GENERATION

The different methodologies are compared using simulated data, constructed from a representative firm model. Details on the model, calibration of the parameters, and evaluation of the different experiments are in Appendix A. We highlight a number of features here.

productivity term. Alternatively, Levinsohn and Petrin [2003] invert the material input demand and Van Biesebroeck [2003a] inverts the first order condition on the choice of the number of shifts to operate a plant.

[25] An intermediate step to control for endogenous exit is omitted here, as no other method that we consider explicitly deals with selection.

[26] For an alternative way to calculate productivity, purging the random error ε_{it} from the productivity estimates, see the alternative estimator (OP2) in Appendix B.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

The model we use is the simplest possible that allows for input substitution in the production function and optimal input choices over time. The firm chooses labor input and investment over time to maximize the net present value of profits, subject to a production function, an exogenously evolving productivity level, the initial capital stock, and a capital accumulation equation. The firm observes all variables at time t, including its own productivity level. The first order condition for labor,

$$\ln L_{it} = \frac{1}{1-\alpha_l}(\ln(\alpha_l) + \ln A_{it} - \ln W_t) + \frac{\alpha_k}{1-\alpha_l}\ln K_{it}, \tag{15}$$

provides an expression for optimal labor input as a function of predetermined variables, including capital, and parameters. Under constant returns to scale, rational expectations, and assumptions on the evolution of exogenous variables, forward substituting the Euler equation for investment gives an analytic expression for investment. Under variable returns to scale, it is impossible to solve the investment function analytically and we use univariate bifurcation methods to calculate optimal investment numerically as a function of the capital stock and productivity level.[27]

Simulating the sample starts with drawing values for the initial values of capital and productivity for all firms. In addition, for each year, we draw the industry wage rate and firm-specific productivity shocks. Adding a set of parameters, we can generate all (endogenous) variables from the production function, the first order condition for labor and the investment equation. The results we report for all experiments are the averages over 50 simulated samples, each consisting of 200 firms that are observed over 10 years. We calibrated the parameters in the model to approximate a number of summary statistics on an actual sample of plants taken from the textile industry in Colombia between 1981 and 1991.

Using the simulated data, we estimate the production function parameters (α_l and α_k), productivity levels ($\ln A_{it} - \overline{\ln A_t}$) and growth rates ($\ln A_{it} - \ln A_{it-1}$) using each methodology. The estimates for each of the three statistics and each of the six methods are compared with the true values of the data generating process. For the production function parameters, we report the bias between estimates and true values.[28] For the productivity level we report the Spearman rank correlation between estimates and true productivity and for productivity growth we report the simple correlation

[27] Bond and Söderbom [2005] use an alternative approach, numerically calculating the value function first and subsequently deriving optimal input choices.

[28] For the two methods that allow for heterogeneity in technology—index numbers and DEA—we report the average bias.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

statistic. Correlations are calculated by year and then averaged over the ten year sample period.

V. RESULTS FOR THREE EXPERIMENTS

V (i). *Heterogeneity in Factor Prices*

The procedure outlined in the previous section can be used to simulate a sample of firms, but without additional heterogeneity the model cannot be estimated. The production function amounts to a transformation of the first order condition for labor, which is used to calculate labor input. There is no unobserved heterogeneity across firms that makes it a stochastic relationship, see also Bond and Söderbom [2005]. For each of the three experiments we perform, a different source of heterogeneity is introduced; each is illustrated using the first order condition for labor in logarithms—equation (15).

Heterogeneity in factor prices enters (15) linearly, adding an additive error (η_{it}^{w}) to $\ln W_t$:

$$\ln L_{it} = \frac{1}{1-\alpha_l}(\ln(\alpha_l) + \ln A_{it} - \ln W_t - \eta_{it}^{w}) + \frac{\alpha_k}{1-\alpha_l}\ln K_{it}. \tag{15a}$$

If the error is assumed to be independent over time and normally distributed, the derivation of the investment equation remains unchanged. Interpreted this way, the error will also appear in the wage bill. An alternative interpretation for the same random variable is as optimization error in the choice of labor. While firms on average choose labor input according to the first order condition, some random mistakes are made.[29] A third way to rationalize the error term is as a shock to productivity after labor input is chosen, but before production takes place or investment is chosen.[30] In this case, the productivity term used in the Euler equation for investment and in the production function differs from the productivity term in the first order condition for labor by the same error. The results we present are consistent with the first two interpretations, which only differ for the index numbers.

Table I contains the results using three different assumptions on the evolution of unobserved productivity. In the top panel, productivity is the sum of two processes: a firm fixed effect and decaying productivity shocks

[29] Heterogeneity in the cost of capital would enter the model in the same way as optimization error in the choice of investment. Results in Van Biesebroeck [2004] indicate that such variation would be of lesser importance, as the endogeneity problem in estimating the production function is most severe for variable inputs. Heterogeneous adjustment costs for capital would enter similarly as multiplicative optimization errors.

[30] As discussed in Ackerberg *et al.* [2005], this allows for identification of the production function with the Olley-Pakes methodology, at least if the timing of input choices is exactly as discussed above.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

TABLE I
WAGE HETEROGENEITY—OPTIMIZATION ERROR

	correlation with true productivity		coefficient bias for	
	level	growth	labor	capital
Productivity is the sum of an AR(1) process and a fixed effect	(1a)	(2a)	(3a)	(4a)
IN (wage heterogeneity)	0.969	0.942	0.003	−0.003
IN (optimization error)	0.705	0.665	0.027	−0.027
DEA	0.771	0.568	−0.086	−0.010
OLS	0.792	0.963	0.022	0.083
SF	0.969	0.975	0.027	−0.011
GMM	0.975	0.996	−0.002	0.048
OP	0.973	0.990	0.009	−0.024
Productivity consists of a firm-fixed effect and entirely transitory shocks	(1b)	(2b)	(3b)	(4b)
IN (wage heterogeneity)	0.994	0.920	0.003	−0.003
IN (optimization error)	0.916	0.615	0.027	−0.027
DEA	0.812	0.332	−0.042	−0.133
OLS	0.594	0.592	0.124	0.149
SF	0.993	0.929	0.040	0.004
GMM	0.996	0.987	−0.007	0.016
OP	0.737	0.702	0.088	0.167
Productivity follows an AR(1) process	(1c)	(2c)	(3c)	(4c)
IN (wage heterogeneity)	0.885	0.912	0.029	−0.029
IN (optimization error)	0.863	0.904	0.026	−0.026
DEA	0.878	0.904	−0.119	−0.052
OLS	0.942	0.981	0.082	−0.072
SF	0.709	0.880	0.193	−0.228
GMM	0.889	0.994	0.030	0.210
OP	0.984	0.995	0.043	−0.058

Notes: Columns (1) contain the Spearman rank-correlation between estimated productivity levels and the true values of the data generating process. Columns (2) contain the correlation between estimated and true productivity growth. Columns (3) and (4) contain the bias in the estimated labor and capital coefficients. For IN and DEA, the average of the firmspecific biases are reported. IN is the only method where it matters whether the errors in the data generating process are modeled as wage heterogeneity or optimization errors. Results for both cases are reported.

following an AR(1) process with $\rho = 0.3$. Both productivity terms are mean-zero normally distributed random variables with a standard deviation of 0.5. In the next two panels, both processes are included separately; in these cases, the productivity terms have a standard deviation of one.

When both the constant and autoregressive components are included, all methods obtain high correlations between true and estimated productivity levels, in column (1a), and generally even higher correlations for growth rates, in column (2a). Especially the parametric methods provide accurate estimates with correlation statistics of approximately 0.97 for the levels and 0.99 for growth rates. The improvement over OLS is much greater for

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

productivity levels than for growth.[31] Especially the GMM and OP estimators are able to reduce the upward bias in the labor coefficient estimate, in column (3a). The bias for the capital coefficient remains nonnegligible, in column (4a), and is smallest for the SF estimator. The index numbers are almost as reliable as the parametric methods, but only if the errors in (15a) are interpreted as wage heterogeneity, i.e., if the errors also enter the wage bill. Finally, DEA results marginally improve over the IN results for productivity levels, in the optimization error case, but for productivity growth they are least accurate of all.

The different assumptions on the productivity process will influence a firm's investment policy and affect how serious the endogeneity problem is. The lower the persistency of the shocks across periods, the lower the weight of current productivity in the investment equation. In the limit, if ρ falls to zero, the firm ignores productivity shocks, only investing to make up for depreciation. This is the situation in the middle panel, and methods including fixed effects, SF and GMM achieve almost perfect correlations for productivity level estimates. Finally, results without permanent productivity differences, but with persistent productivity shocks, are in the bottom panel. The OP estimator that exploits the firm's knowledge about these shocks, as captured by its investment decisions, estimates productivity levels and growth rates almost perfectly. A more detailed discussion of the results for each individual estimator follows in Section VI.

Results in Table I can be used to guide applied researchers in a choice of productivity measurement methodology, at least if the data is measured accurately, and if one is fairly confident that all firms in the sample share the same production technology. Measurement errors are incorporated in Section V(ii) (Table II) and heterogeneity in production parameters in Section V(iii) (Table III). Situations where we expect this to be the case are, for example, large census data sets in developed countries where the analysis can be performed on narrowly defined industries. In developing countries, where data is likely to be more noisy, the current results could still apply reasonably well to agricultural applications or even for manufacturing firms where output and inputs are more straightforward to measure.

These results will be especially relevant if we expect wage (or relative price) heterogeneity to be important, e.g., when labor markets are geographically or functionally segregated or when firms face radically different costs of capital, for example because smaller firms have no access to formal credit markets. If optimization errors are expected to be particularly important, these results will again be relevant. Examples would be data sets with many new firms or with

[31] It is a strong assumption to introduce wage heterogeneity as purely independent across firms and independent of productivity levels. In the likely case that productivity is positively correlated with wages, the improvement of the more sophisticated parametric methods over OLS would be less pronounced.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

TABLE II
MEASUREMENT ERROR

Measurement error in output	correlation with true productivity		coefficient bias for	
	level (1a)	growth (2a)	labor (3a)	capital (4a)
IN	0.307	0.131	0.014	−0.014
DEA	0.090	0.086	0.255	0.088
OLS	0.027	0.079	0.404	−0.404
SF	0.024	0.077	0.419	−0.426
GMM	0.212	0.077	0.380	−0.554
OP	−0.011	0.078	0.417	−0.389
Measurement error in output with (some) wage heterogeneity	correlation with true productivity		coefficient bias for	
	level (1b)	growth (2b)	labor (3b)	capital (4b)
IN	0.302	0.129	0.013	−0.013
DEA	0.233	0.114	−0.068	0.208
OLS	0.176	0.125	0.080	0.010
SF	0.178	0.122	0.099	−0.024
GMM	0.323	0.142	−0.057	0.094
OP	0.292	0.128	0.042	−0.048
Measurement error in inputs	(1c)	(2c)	(3c)	(4c)
IN	0.390	0.159	0.012	−0.012
DEA	0.446	0.170	−0.223	0.046
OLS	0.417	0.177	0.019	−0.031
SF	0.725	0.252	−0.166	−0.044
GMM	0.776	0.455	−0.368	−0.392
OP	0.734	0.252	−0.119	−0.213

Note: Statistics are the same as in Table 1. The wage heterogeneity in the middle panel can alternatively be interpreted as optimization errors (not for IN). Productivity is the sum of an AR(1) process and a fixed effect.

inexperienced owners, or firms facing rapid technological change or other changes in their environment. High frequency productivity shocks, occurring after one input is fixed but before investment for next period's capital takes place, are a third situation where the current results are instructive.

V (ii). *Measurement Error*

Measurement error can be added to inputs or outputs to break the exact relationship between the first order condition and the production function. In this case, all variables are calculated as described in Appendix A, but the variables the econometrician observes differ from these by some random error. Mean-zero, independent and standard normally distributed error

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

TABLE III
HETEROGENEITY IN PRODUCTION TECHNOLOGY

Enforcing constant returns to scale (α_{Li} & $1-\alpha_{Li}$)				
	correlation with true productivity		coefficient bias for	
	level (1a)	growth (2a)	labor (3a)	capital (4a)
IN	0.918	0.995	0.003	−0.003
DEA	0.822	0.632	−0.071	−0.008
OLS	0.236	0.734	0.378	−0.381
SF	0.213	0.737	0.378	−0.376
GMM	0.490	0.884	0.273	−0.243
OP	0.352	0.764	0.357	−0.370
Enforcing constant returns to scale with (some) wage heterogeneity				
	correlation with true productivity		coefficient bias for	
	level (1b)	growth (2b)	labor (3b)	capital (4b)
IN	0.878	0.982	0.002	−0.002
DEA	0.744	0.667	0.005	−0.024
OLS	0.536	0.812	0.116	−0.025
SF	0.721	0.838	0.105	−0.094
GMM	0.800	0.889	0.060	−0.080
OP	0.777	0.843	0.091	−0.142
Allowing for variable returns to scale with (some) wage heterogeneity (α_{Li} & α_{Ki})				
	(1c)	(2c)	(3c)	(4c)
IN	0.753	0.954	0.003	0.005
IN*	0.393	0.986	0.003	−0.001
DEA	0.754	0.640	0.055	0.010
DEA*	0.457	0.553	0.191	−0.162
OLS*	0.181	0.586	0.346	−0.311
SF*	0.254	0.613	0.329	−0.302
GMM*	0.384	0.813	0.104	−0.131
OP*	0.200	0.605	0.335	−0.294

Note: Statistics are the same as in Table 1. Productivity is the sum of an AR(1) process and a fixed effect.
*Indicates results when (some) wage heterogeneity (which can alternatively be interpreted as optimization errors—not for IN) is added in the data generating process.

draws (ε_{it}^{L}, ε_{it}^{K}, ε_{it}^{Q}, ε_{it}^{I}, ε_{it}^{W}) are added to the logarithmic values of labor, capital, output, investment and the wage bill.

For the results in the first panel of Table II, only output is measured with error. While this is sufficient to make the model estimable, the identical first order condition for all firms leads to parametric estimates of the labor coefficient equaling unity and zero for the capital coefficient.[32] A more

[32] Solving the first order condition (15) for $\ln A_t$ and substituting in the production function (A.1) immediately makes clear why this is the case.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

interesting exercise is to include measurement error in output, together with wage heterogeneity with only half the standard deviation used in the previous section. These results are in the second panel of Table II. The reduction in correlations for all methods, relative to Table I, is substantial.[33] Finally, we add measurement error to the factor inputs, investment and the wage bill, but not to output. The errors are added after all endogenous variables are generated. For example, the labor input that the econometrician observes ($\tilde{L}_{it}$) contains measurement error (η^L_{it}),

$$\ln \tilde{L}_{it} = \underbrace{\frac{1}{1-\alpha_l}\left(\ln(\alpha_l) + \ln A_{it} - \ln W_t\right) + \frac{\alpha_k}{1-\alpha_l}\ln K_{it}}_{\ln L_{it}} + \eta^L_{it}, \tag{15b}$$

but the labor input (L_{it}) substituted in the production function to calculate output does not contain the error. The effect on the various methods differs markedly; results are in the bottom panel of Table II. It is here that the advantage of the stochastic nature of the parametric methods becomes apparent.

Results in Table II apply when measurement errors are important, which begs the question of how one is to know. In some cases variables will fail obvious consistency checks, e.g., the wage bill might exceed value added for an implausibly large number of firms, output levels might display large spikes which are not matched by input increases and disappear after a single year, etc. Another situation in which these results are relevant is when input measures are relatively crude, e.g., they lump together workers of varying quality, do not adjust for intensity of input use such as capacity utilization or temporary workers, or ignore cross-firm variation in hours worked. Similarly, the output measure may combine different products and the composition changes over time or between firms. If such data imperfections are exogenous or at least uncorrelated with productivity, the effect will be similar to measurement error. When variables are constructed from firm surveys that use recall information in the absence of accounting data, as is often the case in developing countries, measurement errors are of obvious concern.

At this point, it is also useful to mention the results for two additional estimators, which are elaborated on in Appendix B. All parametric methods distinguish implicitly between two sources of variation in the production function (8), ω_{it} and ε_{it}, because only the first term leads to the endogeneity problem. When productivity estimates are calculated according to equations (9) and (10), the random errors (ε_{it}) are implicitly lumped into the productivity estimate. To the extent that these capture unobserved (to the firm) productivity shocks or shocks that are realized after inputs are chosen, that is still

[33] The correlations for productivity growth tend to be affected much more, as i.i.d. measurement errors for two time periods enter these calculations.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

satisfactory. If they represent measurement error, it reduces the accuracy of the productivity estimates. For the stochastic frontier and semiparametric estimators, it is possible to calculate a pure estimate of productivity level ($\omega_{it} - \overline{\omega_t}$) or growth ($\omega_{it} - \omega_{it-1}$), purged from the random errors.

The correlations in the third line of Table B.1 indicate that these methods perform much better than any of the estimators in Table II if output is measured with error, at least if one is interested in productivity levels. Especially the alternative Olley-Pakes estimator (OP2) achieves a high correlation (0.783), almost as high as for the experiments in Table I. The alternative stochastic frontier method (SF2) also outperforms the other methods in Table II and becomes the preferred estimator if all productivity differences are fixed over time (results not reported).

These two alternative methods are more appropriate to estimate productivity levels than growth rates—note the consistently low correlations in columns (2a) and (2b). The SF2 method estimates a firm-specific deterministic productivity growth rate, which performs very poorly in all experiments. The OP2 estimator is the preferred estimator for productivity growth if only output is measured with error and is second best if inputs are noisy. Still, the correlations are sufficiently below the results with the usual OP estimator in all the other exercises reported in Tables I and III to make one wary of using OP2 for productivity growth calculations.

V (iii). *Heterogeneity in Technology*

Finally, heterogeneity can also be introduced through the technology parameters. Normally distributed errors ($\eta_i^l$and/or η_i^k) are added to the input elasticity parameters in the production function. These errors are assumed to differ by firm, but to be constant over time. The first order condition for labor (15c) and the Euler equation for investment will also contain the firm-specific production parameters:

$$(15c) \qquad \ln L_{it} = \frac{1}{1-\alpha_l-\eta_i^l}\left(\ln(\alpha_l+\eta_i^l) + \ln A_{it} - \ln W_t\right) + \frac{\alpha_k+\eta_i^k}{1-\alpha_l-\eta_i^l}\ln K_{it}.$$

This change makes the parametric approaches misspecified, but not the two nonparametric approaches, IN and DEA, as they make only limited functional form assumptions on the deterministic part of the production model.

For the results in the top panel of Table III, the constant returns to scale assumption in production is maintained in the data generating process. A normally distributed mean-zero term with standard deviation of 0.1 is added to the labor coefficient ($\alpha_{li} = 0.6 + \eta_i^l$), while the capital coefficient is still $1 - \alpha_{li}$. In the simulated samples, this amounts to a labor coefficient varying

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

between 0.35 and 0.84, with an interquartile range of 0.55–0.65. Introducing these errors makes the production function estimable, but they do not break the link between the first order condition and production function.[34] In the middle panel of Table III, some wage heterogeneity is added as well, as in the previous experiments. Finally, both input coefficients are varied independently, results in the bottom panel, which relaxes the constant returns to scale assumption in the data generating process. As a result, it becomes impossible to derive an explicit investment equation and we have to rely on equation (A.5) to compute optimal investment numerically.[35]

Without wage heterogeneity, the parametric approaches are very unreliable, evident from the large biases on the input elasticity estimates and the low correlations for productivity levels. In contrast, the IN and DEA estimators are able to deal very well with the technological heterogeneity. Even when returns to scale vary, productivity level correlations for IN and DEA are still around 0.75. With wage heterogeneity, the parametric estimators perform not quite as badly under constant returns to scale, although correlations are never as high as for the index numbers. Under variable returns to scale, their adjustments for the endogeneity problem are ineffective, resulting in upward biases in the labor coefficients almost as large as with OLS. The correlation with the true productivity levels is on average only 0.28 for the three parametric methods.

Given the stark differences between estimators in these experiments, it would be particularly useful to know when the results in Table III apply. One situation would be if one's data set pools firms from different industries or different countries and the sample size is not large enough to split the sample. If observations are aggregate entities—e.g. industries—or are mostly larger firms in developed economies, the results under the constant returns to scale assumption are likely to be most appropriate. If the sample contains firms from developing economies, firms from a new industry, or many entrants, the variable returns results are likely to be more relevant.

Even if all firms come from a single industry, there might be good reason to suspect that they do not all use the same production technology. For example, plants producing steel have the choice to operate as an integrated mill using the basic oxygen furnace to melt iron ore or operate as a so-called mini-mill using electric furnaces to process scrap metal. Automobile assembly plants can use highly automated mass manufacturing

[34] The model is only estimable because the intercept varies with the labor parameter.

[35] Ideally, one would like to investigate the impact of the functional form of the production function more thoroughly, as was done with U.S. data in Berndt and Khaled [1979]. Previous Monte Carlo studies, for example Gagné and Ouellette [1998] and sources cited there, use exogenously generated inputs to study robustness to functional form. In the current framework, it is only possible to explicitly derive consistent input choices for the simplest of technologies (Cobb-Douglas) and only if no scale economies are present.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

methods to maximize scale economies or alternatively rely on lean, more flexible, manufacturing techniques. Larger textile firms in developing countries that produce for the export market are often much more automated than domestically-oriented, smaller competitors. In each case, one might be hesitant to split the sample *a priori*, but when diverse firms are pooled, heterogeneity in technology is likely to be important.

A final point to note is that these experiments underscore that identifying the parameters of a Cobb-Douglas production function is not obvious. For example, introducing measurement error in output or heterogeneity in technology in the data generating process was not sufficient to break the link between the first order condition and the production function. In these experiments, we simply added some wage heterogeneity or optimization errors when necessary. An alternative solution would have been to make both inputs subject to adjustment costs or allow productivity to change after labor input is chosen, but before production takes place. Structural identification of the parametric estimators is studied in greater detail in Bond and Söderbom [2005] and Ackerberg *et al.* [2005]—see also footnote 16.[36]

VI. STRENGTHS AND WEAKNESSES OF EACH METHOD

VI (i). *Index Numbers (IN)*

This method tends to produce very robust results across the different experiments. Correlations are consistently among the highest of the different methods, especially for productivity growth estimates. If one has no *a priori* information about the importance and the degree of persistence of productivity shocks, the index numbers provide a compromise—performing intermediate in the extreme situations of no or full persistence (compare the bottom two panels of Table I).

When heterogeneity is added linearly in the first order condition, estimates are especially accurate if it enters as wage heterogeneity, i.e., when the errors also carry over to the wage bill. This is not surprising as the method exploits firms' optimization decision on the input trade-off, as reflected by the share of labor in total costs. If the heterogeneity is merely optimization error, i.e., does not appear anywhere else, correlations are notably lower.

[36] One of the first papers to discuss explicitly the difficulty identifying parameters in a Cobb-Douglas production function when inputs are chosen optimally is Zellner *et al.* [1966]. Bond and Söderbom [2005 demonstrate that it is impossible to identify the parameters from cross section variation if all inputs are perfectly flexible and input prices are common to all firms. They introduce adjustment costs (also to labor) to achieve identification. Ackerberg *et al.* [2005] investigate the (implicit) identification strategy for the semiparametric estimator. They argue that the first stage suffers from a collinearity problem. Their alternative estimator purges the random errors (ε_{it}) from the production function in the first stage and jointly estimates all coefficients in the second stage using GMM.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

The advantage of the index numbers over other methods is most pronounced when technology is heterogeneous. Correlations are especially high when returns to scale are constant, when other errors are absent, and for the productivity growth estimates. We implemented the Törnqvist index enforcing constant returns to scale, measuring the capital share in costs as one minus the labor share—as is most commonly done in applications.[37] Still, even in the case of variable returns to scale, the IN results provide the highest correlation for productivity growth.

As expected, the method is vulnerable to measurement error, but so are the alternatives. The index numbers only perform notably worse than the parametric alternatives when inputs, including the wage bill, are measured with error. When only output is measured inaccurately, the productivity level and growth correlations for IN are only slightly lower than for GMM, the best-performing method in this case.

Bias in input weights is almost entirely absent throughout the three tables. Even with optimization error, where correlations for IN are relatively low, the biases are still moderate (± 0.027). To a large extent, this is by construction as the data generating process also features constant returns to scale production; even when returns vary across firms, they are constant on average. Moreover, when firms optimize input factors subject to a Cobb-Douglas production function, the wage bill will always be equalized to the labor input coefficient. Only for this particular production function do we obtain the input parameter estimates with the IN method at all.

VI (ii). *Data Envelopment Analysis (DEA)*

Unlike the results for other methods, the DEA correlations for productivity growth are rarely higher than the correlations for productivity levels. It underscores the comparative advantage of this method—calculating productivity levels. Productivity growth comparisons are largely *ad hoc*: production plans for different time periods are compared to the same frontier, which is estimated by pooling all observations. For comparability with other methods, the efficiency numbers in percentage terms are rescaled to log differences, which further influences the simple correlations for productivity growth, but not the rank correlations for productivity levels.

For DEA, it does not matter whether the linear errors in the first order condition represent wage heterogeneity or optimization errors. In the latter

[37] Generalizations exist that control for variable returns to scale, see Van Biesebroeck [2004], but they either require an outside estimate for the returns to scale or a user cost of capital. Results in the latter paper suggest that, in some cases, productivity estimates are more accurate when constant returns are enforced, even if they are increasing or decreasing in the data generating process.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

case, DEA outperforms IN in the benchmark case (top panel of Table I), but it still lags the parametric methods. The DEA method handles decaying productivity shocks, according to the AR(1) process, slightly better than firm fixed effects. The *ad hoc* weighting of inputs leads to sizeable biases on the input elasticities in most cases, without any particular pattern. Moreover, the variation in weights across plants tends to be large; much larger than for the index numbers.

Other patterns are as expected. As it is a deterministic method, it is vulnerable to measurement error. In contrast to the IN results, inaccurately measured outputs pose more of a problem than measurement errors in inputs. The correlation for productivity levels in the middle panel of Table II stands at only 30% of the comparable benchmark statistic in Table I; correlations are almost twice as high in the bottom panel. The main advantage of DEA is to estimate productivity levels when the production technology is heterogeneous (Table III). Especially in the variable returns to scale case, it is the most robust method of all. The linear aggregation of inputs is surprisingly effective in matching the varying weights on the two inputs. Absent wage heterogeneity/optimization errors (top panel) the average bias for the labor and capital coefficients is only 0.055 and 0.010, respectively. Its advantage over the runner-up (IN) in estimating productivity levels becomes more pronounced if wage heterogeneity compounds the heterogeneity in technology (middle panel).

VI (iii). *Parametric Estimation: Least Squares (OLS)*

Ignoring the simultaneity of unobserved productivity and input choices—as with OLS—leads to upwardly biased parameter estimates in the production function. While we expected a larger bias for the more variable input, consistent with the discussion in Olley and Pakes [1996], the built-in correlation between the initial capital stock and productivity (see footnote A.5) leads to a larger bias for the capital coefficient in the benchmark results. For longer panels, the relative bias eventually turned in the expected direction. Adding any sort of measurement error has a similar effect. Moreover, actual data sets tend to be unbalanced panels with a greater exit probability for less productive firms. If firms with a larger capital stock overcome greater negative productivity shocks, sample selection will impart a further downward bias on the capital coefficient.

In the benchmark case, the three parametric methods that explicitly control for the simultaneity have much smaller biases than the OLS estimates. While small in absolute value, some biases are still significantly different from zero, as coefficients are estimated very precisely. When measurement error or production heterogeneity is added, the advantage of the more sophisticated estimators over OLS in estimating the input elasticities becomes less clearcut.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

It is worthwhile to note that the most accurate parametric method in one situation can be among the least reliable in another situation. This is especially true for the productivity level estimates. In Table I, the OP estimator achieves the highest correlation when there are no fixed effects (bottom panel), while the SF correlation is the lowest. Even without changing the source of heterogeneity, if productivity is constant over time instead of evolving according to an AR(1) process, the relative performance of these two estimators is reversed. At the same time, the OLS correlations are consistently among the lowest. There is clearly a payoff to the more sophisticated estimators, but the magnitude of the benefit will depend on selecting the right method for the situation.

Relative to IN or DEA, the parametric methods are superior if there is measurement error, even though the production function parameters are estimated with a downward bias. As expected, this is especially pronounced with measurement in inputs, and affects the capital coefficient estimates most. For the OLS results, the effects of the endogeneity problem and measurement error in the explanatory variables work in opposite directions in Table II. Coincidentally, for the parameter values and standard deviations used, OLS estimates happen to be virtually unbiased.[38]

VI (iv). *Parametric Estimation: Stochastic Frontiers (SF)*

The highest correlations for SF are obtained in the middle panel of Table I, where productivity is a firm fixed effect with only transitory productivity shocks. The correlation with true productivity level drops by 2.5% if an AR(1) process is added to the data generation process. If only the AR(1) process is present, the correlation is a full 28.5% lower. While the correlation for productivity level is an amazing 0.993 with only fixed effects present, the GMM results which also contain plant dummies are slightly better in the first two panels and their advantage grows if no fixed effects enter the data generating process.

Differences are smaller for productivity growth, but in the same direction: the correlation is 10.7% higher when fixed effects are added to the AR(1) process. The SF estimator we implemented is very restrictive on the evolution of productivity—assumed to be the same for all firms and constant over time. Results with the more flexible alternative estimator, see

[38] Additional results in Van Biesebroeck [2004] indicate that the performance of parametric methods improves initially if measurement error is gradually increased. While the correlations for the nonparametric methods decrease monotonically with increasing standard deviation on the measurement error, the correlations for parametric methods tend to increase first and only decline for very large errors. This stems from the difficulty identifying a Cobb-Douglas production function. Adding some randomness helps (initially) in breaking the link between the production function and the first order condition used to calculate labor input.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

SF2 in Appendix B, are not very robust either; at least in the current experiments without any productivity trend in the data.

Given that the stochastic frontier estimators make explicit assumptions on the underlying distribution of productivity, they are likely to be most sensitive to the exact modeling of productivity in the data generating process. For example, they generally assume that the productivity terms are left-skewed, as they are truncated from above at zero. Generating errors with such skewness built-in improved correlations slightly, see Van Biesebroeck [2004]. It makes evaluating this estimator sensitive to the maintained assumptions.

Overall, the results for SF tend to be worse than for either the OP or GMM estimators, which each have their particular advantages. Nowhere in Tables II or III are the SF correlations higher than both the OP and GMM results. However, the difference are often small. The results for SF are weakest when there are no fixed effects in productivity and with measurement error in output. In these cases, the correlations are lower and the bias on the parameter estimates larger than for the OLS results. In the latter case, the SF2 method does provide good productivity level estimates.

VI (v). *Parametric Estimation: Instrumental Variables (GMM)*

The inclusion of firm dummies makes this estimator, just as the stochastic frontiers, more apt at dealing with firm fixed effects in the productivity process than with persistent productivity shocks. The built-in AR(1) process still makes the GMM estimator perform rather well in most experiments, but this could change if actual productivity dynamics took a different form. When all productivity growth is random and transitory, in the second panel of Table I, it produces the highest correlations for both productivity levels and growth rates, and parameter biases are almost as low as for IN.

The GMM estimator achieves the highest correlations in several of the panels in Tables I–III, including in the benchmark results. Its advantage is often higher for productivity growth rates than for the levels. In particular, GMM results tend to be least sensitive to measurement error in both inputs or output. It is the only parametric method to outperform the nonparametric but deterministic approaches when output is measured with error (middle panel of Table II). Still, results with measurement error tend to be worse than for the other types of heterogeneity considered in Tables I or III. Correlation statistics are less than one third of the benchmark results and only slightly higher than for the deterministic DEA method that simply ignores all errors. Finally, the stochastic frontier and semiparametric estimators have alternatives that are particularly accurate when measurement error primarily affects output.

In the benchmark case (top panel of Table I) and with measurement error for inputs, the highest correlations for productivity levels and growth

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

rates are achieved with the GMM estimator, but it also has the highest bias for the input parameter estimates. This discrepancy is particularly apparent in the bottom panel of Table II. Its correlation for productivity growth is almost twice as high as for any other method, but both input coefficients are estimated with a very large downward bias. While the method allows for general (unobserved) productivity processes, the trade-off is lower precision in the estimates of the deterministic portion of the production process. The coefficient restrictions on the lagged input terms—e.g., the coefficient on lagged capital should equal $\rho\alpha_k$—are always rejected and not enforced.

The inclusion of lagged terms and fixed effects in the model is also an advantage when there is heterogeneity in the production technology; see results in Table III. At least when some wage heterogeneity is added to the first order condition, the GMM correlations approach the IN results. Just as for the other parametric estimators, the average bias on the production function parameter estimates is higher than in the benchmark case. However, in the variable returns to scale case (in the bottom panel), only the GMM estimator is able to improve over the OLS parameter estimates.

VI (vi). *Semiparametric Estimation (OP)*

This estimator has its strong points, but is not as consistent as the index numbers or GMM estimator. In the benchmark case (top panel of Table I), OP results virtually tie with the GMM results. Correlations are marginally lower, but so is the bias on the coefficient estimates. Correlations for OP are in turn marginally higher than for IN, but coefficient biases are somewhat higher as well.

OP productivity estimates are extremely accurate in the absence of fixed effects. If productivity shocks are idiosyncratic, large, and at least partly persistent, the control function approach in OP contains useful information about the arrival of shocks. In particular, when productivity evolves according to a pure AR(1) process, which satisfies the Markov assumption in the model, the correlations are 0.984 for the levels and 0.995 for the growth rates. The advantage of OP over SF or GMM is likely to be even larger if productivity evolves according to a more general process, e.g., if shocks decay at firm-specific rates or if persistency varies over time. In contrast, if shocks are entirely transitory and firm fixed effects are important (middle panel of Table I), correlations with OP are low and the biases on the parameter estimates are even larger than for OLS.

Among the parametric methods, if one is mainly interested in measuring productivity growth, the GMM and OP methods seem preferable.[39] If the

[39] The low correlation for productivity growth with OP in column (2b) of Table 1 is not very informative. The firm-specific growth in this case is merely the arrival and disappearance of transitory shocks, which averages to zero over the sample.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

principal interest is in cross-firm comparisons of productivity levels, one has to consider the process that governs the evolution of productivity. If productivity differences are likely to be stable and productivity shocks are transitory, GMM and SF are preferable. The same would hold if productivity shocks are mostly common to all firms. If shocks are persistent, OP is the most reliable method.

In the experiments that introduce measurement error and heterogeneity in technology, but no other errors (top panels in Tables II and III), the correlations for OP are clearly lower than for GMM. When some wage heterogeneity is introduced additionally (middle panels) the differences are smaller. OP correlations with measurement error and wage heterogeneity (or optimization error) stand at 91% of the GMM correlations; with heterogeneity in technology they are even 96% of the GMM correlations, on average. If the extra heterogeneity is introduced as high frequency productivity shocks occuring after labor input is chosen—the third interpretation of the shocks in the first experiment—OP correlations are higher than the GMM statistics. In this case, the investment equation provides the most up to date information on the state of productivity of any observable variable.[40]

Finally, results in the bottom two panels of Tables II and III are less encouraging for OP. If measurement errors affect inputs and investment, OP is almost as effective as GMM to estimate productivity levels, but not for growth rates. Results in Van Biesebroeck [2004] indicate that if only investment is observed with error, the effect on OP is smaller than for similarly sized errors to output or inputs, which is reassuring. When technology is heterogeneous and returns to scale are variable, the advantage of OP over OLS in each of the four columns is marginal. In this case, investment decisions are more complicated as they also depend on the capital stock. Moreover, the actual impact of both productivity and capital differs by firm as the input elasticities enter the Euler equation for investment.

VII. CONCLUSIONS

For five widely used techniques to estimate productivity, we have compared the robustness of the results to different ways randomness can enter the model of production. While the different literatures have evolved more or less independently, our discussion of the different estimators highlights the fact that the underlying model is very similar, but that methods differ in the assumptions they make on two unobservables: the nature of the input trade-off embodied in technology and the evolution of productivity. Using

[40] These last results are not reported, as the empirical relevance is unclear, but are available upon request.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

simulated samples of firms, we introduced three sources of variation in the data generating process: factor price heterogeneity (which alternatively can be interpreted as optimization error or productivity shocks occurring after inputs are chosen), measurement error, and differences in production technology.

As mentioned earlier, each method has its own strengths and weaknesses. It is no surprise then, that the different experiments tend to favor different estimators. Even though several of the findings are in line with *a priori* expectations—for instance, nonparametric methods do well when the production function is heterogeneous; stochastic methods do well if there is measurement error—we believe several findings are novel. Moreover, researchers do not always base their choice of methodology for estimating productivity explicitly on the expected performance in the specific situation at hand. Even though the simulated experiments represent only a fraction of the situations one might face, we can draw some lessons concerning when we expect each methodology to be particularly appropriate. We summarize these here:

(i). Index numbers (IN): Unless one expects the data to be subject to a lot of measurement error, this method produces consistently accurate productivity growth estimates. When the units of analysis (firms, industries, . . .) are likely to employ different technologies, productivity level estimates are among the most robust as well.
(ii). Data envelopment analysis (DEA): While never the ideal method for estimating productivity growth, it is the preferred estimator for productivity levels if technology is likely to vary across firms and scale economies are not constant. Examples would include situations where one has to pool firms from very different industries, at different stages in their lifecycle or operating in countries with different levels of development.
(iii). Ordinary least squares (OLS): Given the well-known problem of simultaneity among inputs and unobserved productivity, this method is generally not advisable. A simple fixed effects estimator will do better in most situations. Its disadvantage is less acute for productivity growth than for productivity levels.
(iv). Stochastic frontiers (SF): When one has good reason to believe that productivity differences are constant over time, that output is measured accurately and that observations share the same technology, this method produces accurate productivity level estimates.
(v). GMM-SYS: In the presence of measurement error or heterogeneous production technology, this provides the most robust productivity level and growth estimates of the parametric methods. In the absence of these forms of heterogeneity, it provides the most reliable results overall if at least part of the productivity difference is constant over time.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

(vi). Semiparametric estimator (OP): When firms are subject to idiosyncratic productivity shocks that are not entirely transitory, this estimator will exploit the firm's knowledge about these shocks. In addition, if output is measured with error, the alternative estimator, which purges random noise from the productivity estimates (discussed in Appendix B), provides accurate results, especially for productivity levels.

APPENDIX A
DATA GENERATION

A (i). *A Representative Firm*

The different methodologies are compared using simulated data, constructed from a representative firm model. The model is the simplest possible that allows for input substitution in the production function and optimal input choices over time. The firm chooses labor input and investment over time to maximize the net present value of profits, subject to a production function, initial capital stock and a capital accumulation equation:[41]

$$\begin{aligned} &\max_{L_t, I_t} E_t \sum_{t=0}^{\infty} \beta^t [Q_t - W_t L_t - g(I_t)] \\ &\text{subject to} \quad Q_t = A_t L_t^{\alpha_l} K_t^{\alpha_k} \\ &\qquad\qquad\quad K_{t+1} = (1-\delta)K_t + I_t \\ &\qquad\qquad\quad K_0 = \overline{K}. \end{aligned} \tag{A.1}$$

Q_t is the value of output, W_t the wage rate, and $g(.)$ is a convex function capturing all costs associated with investment, including adjustment costs, the purchase price of capital, and the present value of all user costs. Current investment only becomes productive in the next period. The firm observes all variables at time t, including its own productivity level A_t.

The first order condition for labor immediately gives optimal labor input as a function of predetermined variables and parameters

$$L_t = \left(\frac{\alpha_l A_t}{W_t}\right)^{1/(1-\alpha_l)} K_t^{\alpha_k/(1-\alpha_l)}. \tag{A.2}$$

The Euler equation for investment is

$$\begin{aligned} g'(I_t) =& \beta\alpha_k\alpha_l^{1/(1-\alpha_l)} E_t[A_{t+1}^{1/(1-\alpha_l)} W_{t+1}^{\alpha_l/(1-\alpha_l)} K_{t+1}^{(\alpha_l+\alpha_k-1)/(1-\alpha_l)}] \\ &+ \beta(1-\delta)E_t[g'_{t+1}(I_{t+1})]. \end{aligned} \tag{A.3}$$

Under constant returns to scale, the capital stock drops out of (A.3). Assuming quadratic investment costs, $g(I_t) = \frac{b}{2} I^2$, and forward substituting equation (A.3) gives

[41] The exposition in this section benefited from Chapter 4 in Syverson [2001]. For an alternative approach to simulate the data, using the firm's value function, see Bond and Söderbom [2005].

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

the investment function as a function of current and future exogenous variables,

$$I_t = \frac{\beta\alpha_k\alpha_l^{1/(1-\alpha_l)}}{b} E_t[\sum_{\tau=0}^{\infty} [\beta(1-\delta)]^\tau A_{t+1+\tau}^{1/(1-\alpha_l)} W_{t+1+\tau}^{-\alpha_l/(1-\alpha_l)}]. \tag{A.4}$$

Adding rational expectations and assumptions on the evolution of exogenous variables, we can forward substitute $A_{t+1+\tau}$ and $W_{t+1+\tau}$ to express investment as a function only of current variables and parameters. The wage rate is modeled as a random walk ($\ln W_t \sim i.i.d.\, N(0, \sigma_w^2)$) and log-productivity is an autoregressive process with a firm fixed effect ($\ln A_{it} = a_i + a_{it}$ with $a_{it} = \rho a_{it-1} + \eta_{it}$ and $\eta_{it} \sim$ i.i.d. $N(0, \sigma_a^2)$). This reduces the investment equation to a deterministic function[42]

$$I_t = \frac{\beta\alpha_k\alpha_l^{1/(1-\alpha_l)}}{b} \sum_{\tau=0}^{\infty} [\beta(1-\delta)]^\tau \exp(\frac{a_i + a_{it}\rho^{\tau+1}}{1-\alpha_l}) \prod_{s=0}^{\tau} \exp(\frac{1}{2}(\frac{\sigma_a\rho^s}{1-\alpha_l})^2). \tag{A.5}$$

Under variable returns to scale, it is impossible to solve the investment function analytically as a function of exogenous variables. Because the future capital stock is a function of current investment, equation (A.3) is a nonlinear function of I_t. Univariate bifurcation methods can be used to find the investment that solves for $\phi[I] = 0$, as a function of productivity and capital, in

$$\begin{aligned}
\phi[I(A_t, K_t)] =& bI(A_t, K_t) \\
&- \alpha_1 E_t[\underbrace{A_{t+1}^{1/(1-\alpha_l)}}_{exp(a_i+\rho a_{it}+\varepsilon_{it})} W_{t+1}^{\alpha_l/(1-\alpha_l)} \underbrace{K_{t+1}^{(\alpha_l+\alpha_k-1)/(1-\alpha_l)}}_{(1-\delta)K_t+I(A_t,K_t)}] \\
&- \alpha_2 E_t[bI(\underbrace{A_{t+1}}, \underbrace{K_{t+1}})] \\
=& bI(A_t, K_t) \\
&- \alpha_1 exp(\frac{\sigma_w^2\alpha_l^2 + \sigma_a^2}{2(1-\alpha_l)^2} + \frac{a_i + a_{it}\rho}{1-\alpha_l})[(1-\delta)K_t + I(A_t, K_t)]^{(\alpha_l+\alpha_k-1)/(1-\alpha_l)} \\
&- \alpha_2 b \int_\varepsilon I(exp(a_i + \rho a_{it} + \varepsilon), (1-\delta)K_t + I(A_t, K_t)) f(\varepsilon) d\varepsilon,
\end{aligned}$$

where the coefficients $\alpha_1 = \beta\alpha_k\alpha_l^{1/(1-\alpha_l)}$ and $\alpha_2 = \beta(1-\delta)$ come from equation (A.3) and the logarithm of the wage rate is assumed to follow a random walk with mean zero. We use the Newton-Raphson algorithm to find the root for $\phi[I] = 0$ and the integration is handled by Gaussian quadrature using five points of support.

Simulating the sample starts by drawing values for the annual industry wage rate (W_t) and initial values for capital (K_{i0}) for all firms. In the benchmark model, the wage rate and cost of capital are normalized to one, but this can easily be relaxed. The capital series is initialized to resemble the first year of an actual sample of firms, details are in Van Biesebroeck [2004]. Log-capital is drawn from a normal distribution with mean five and standard deviation 1.5. In addition, for each firm we draw the productivity shocks (η_{it}), the fixed component of productivity (a_i), and initial values for productivity (a_{i0}) from mean-zero normal distributions.[43] The parameters governing the evolution

[42] The derivation uses that $E[\exp(\gamma\varepsilon_t)] = \exp(\frac{1}{2}(\gamma\sigma_a)^2)$ for normally distributed variables.

[43] According to the Euler equation, more productive firms will invest more in the simulated data. We incorporate this past investment behavior by forcing a correlation of 0.3 between initial productivity and the capital stock in the first year. Without such an initialization, the

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

of productivity are discussed in the main text; see Section V(i). Adding a set of parameters $\Gamma = [\alpha_l, \alpha_k, \beta, \delta, b]$, one can generate all (endogenous) variables $\mathbf{y} = [Q, K, L, I, WL]$. We use the following parameter values: 0.6 for the labor coefficient in the production function (α_l), 0.4 for capital (α_k), 0.95 for the discount rate (β), 0.15 for the depreciation rate (δ), and the adjustment cost for capital (b) is calibrated to match the investment share in an actual data set (see below).

Without additional heterogeneity, the model cannot be estimated, as the first order condition for labor is merely a transformation of the production function. There is no unobserved heterogeneity across firms that makes it a stochastic relationship. For each of the three experiments discussed in Section V, we introduce a different source of heterogeneity. Each can be illustrated straightforwardly modifying the deterministic first order condition for labor in logarithms:

$$(15') \qquad \ln L_{it} = \frac{1}{1-\alpha_l}(\ln(\alpha_l) + \ln A_{it} - \ln W_t) + \frac{\alpha_k}{1-\alpha_l}\ln K_{it}.$$

A (ii). *A Reality Check*

To verify that the firm samples simulated with this data generation process resemble an actual data set, we compare a number of summary statistics with the corresponding statistics for an actual sample of plants. The real data was taken from the Colombian census of manufacturing, limited to 256 textile plants over the 1981–91 period; details on the data source are in Roberts [1996]. The data generating process used to compare the simulated and real data includes the first two sources of heterogeneity (heterogeneous wages/optimization error and measurement error). The standard deviations for the error terms are only half as large as in the main analysis in Section V, where each source is considered separately.[44]

The comparison indicates that appropriately choosing the parameters in the representative firm model generates a sample of firms that closely resembles an actual data set. Full results are in the working paper version (see Tables I and II in Van Biesebroeck [2004]) and available upon request for each of the individual sources of heterogeneity as well. There are sufficient degrees of freedom to match the observed statistics much better, for example the size of measurement error could vary by variable, but we are mainly interested in verifying whether production function estimates are similar for the real and simulated data.

A number of findings are worthwhile mentioning specifically. Average levels and standard deviations of all observable variables match reasonably well, even without including any time trends in the data. On the simulated data, the labor coefficient in a Cobb-Douglas production function is estimated much higher than the wage share in output, 0.89 versus 0.62 (0.98 versus 0.62 in Colombia), while the R^2, at 0.83, is similar to that for the Colombian data (0.82). The correlation of the (log) level of output with

standard deviation of capital almost doubles, as some highly productive firms go on an investment binge because they were randomly started off with a low capital stock and vice versa for large but unproductive firms.

[44] To obtain an unbalanced panel, as in most data sets, we augment the algorithm with an exogenous exit rule that drops firms randomly. The probability of exit is modeled as a decreasing function of productivity and the size of the capital stock.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

labor is 28% higher (29% in Colombia) than the correlation with capital. The correlation between capital and labor is one of the lowest of all pairs in both cases: 0.70 for the simulated and 0.69 for the Colombian data. The correspondence is not as tight for the growth rates: the average correlation between all simulated variables is 0.33, while it is only 0.22 for the Colombian variables. In addition, the relative correlations tend to vary quite a bit. One notable pattern is that correlation between the wage bill and labor input in the observed data is close to unity (in levels or growth rates), suggesting a small role for wage heterogeneity. Hence, the additive errors in (15′)—results in Table I and the middle panels of Tables II and III—are more likely to represent optimization error than wage heterogeneity, which matters for the IN results.

The simulated and actual samples differ most notably on three dimensions, which can be moderated by introducing further sources of heterogeneity in the data generating process. First, investment as a share of capital is higher in the simulated data and the standard deviations of investment and capital, but not the investment share, are lower. These differences are reduced if some heterogeneity in the (adjustment) cost of capital is introduced. Second, correlations between output and input growth are higher in the simulated data. This becomes less pronounced if a stochastic component is added to underlying productivity growth (currently there is no trend at all). Both generalizations are explored in Van Biesebroeck [2004]. Third, the standard deviation of the wage share is much lower in the simulated data. While it increases when we introduce heterogeneity in the technology parameters, in the third experiment, the standard deviation in the simulated data is never as high as in the the observed data. This suggests that the technology in the real data is more flexible than the simple Cobb-Douglas production function.

A (iii). *Evaluation*

The results we report for all experiments are the averages over 50 simulated samples, each consisting of 200 firms that are observed over 10 years. Panels are balanced. The limitation on the number of simulations is the DEA method, which can take up to an hour for a sample of 2,000 observations if firms vary dramatically in size.

Using the simulated data, production function parameters (α_l and α_k), productivity levels ($\ln A_{it} - \overline{\ln A_t}$) and growth rates ($\ln A_{it} - \ln A_{it-1}$) are estimated using each methodology. The following table summarizes the acronyms we will use when discussing the results and lists the formulas used to calculate productivity for each method.

Mehtod		(level)–(growth)
IN	Törnqvist index enforcing constant returns to scale	(4)–(3)
DEA	Data Envelopment Analysis: nonparametric frontier	(6)–(7)
OLS	Least squares estimation of output on L and K	(9)–(10)
SF	Stochastic frontier with time-invariant productivity ranking	(9)–(10)
GMM	joint estimation of prod. function in levels and first differences	(9)–(10)
OP	Semiparametric inversion of investment equation	(9)–(10)
Alternative estimators (in Appendix B)		
SF2	Stochastic frontier with three sets of dummies per firm	(B.1)–(B.2)
OP2	Estimation as OP, productivity calculated without error term	(B.3)–(B.4)

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

For each method, the estimates for each of the three statistics are compared with the true values of the data generating process. To evaluate the results, for the production function parameters we report the bias between estimates and true values. For methodologies that estimate a distribution of parameters or for the parametric methods when the underlying technology is heterogeneous, we report the average bias. For the productivity level, we report the Spearman rank correlation between estimates and true productivity. This focuses on the order of firms rather than the absolute size of the estimates, but results are extremely similar using the simple correlation statistic. For productivity growth, we report the simple correlation statistic. Correlations are calculated by year and averaged over the ten year sample period. An alternative evaluation criterion, the mean squared error (MSE) between the actual and estimated values for productivity and parameter estimates, is reported in the working paper version; see Van Biesebroeck [2004]. The ranking of estimators was invariably the same with the MSE as with the correlations or biases.

APPENDIX B
TWO ALTERNATIVE ESTIMATORS

Two alternative estimators that calculate productivity (ω_{it} in equation (8)) purged from the random error (ε_{it}) are worth mentioning. Cornwell *et al.* [1990] introduced a more flexible generalization to panel data of the cross-sectional stochastic frontier, estimating a time-varying, firm-specific effect using three coefficients per firm:

$$\omega_{it} = \alpha_{i0} + \alpha_{i1} t + \alpha_{i2} t^2.$$

Productivity evolves deterministically, but the growth rate changes over time and varies by firm. Productivity level and growth can be calculated as

$$\text{(B.1)} \qquad \ln A_{it}^{SF2} - \overline{\ln A}_t^{SF2} = (\hat{\alpha}_{i0} - \overline{\hat{\alpha}_0}) + (\hat{\alpha}_{i1} - \overline{\hat{\alpha}_1})t + (\hat{\alpha}_{i2} - \overline{\hat{\alpha}_2})t^2$$

$$\text{(B.2)} \qquad \ln A_{it}^{SF2} - \ln A_{it-1}^{SF2} = (\hat{\alpha}_{i1} - \hat{\alpha}_{i2}) + 2\hat{\alpha}_{i2} t.$$

While the estimator uses a lot of degrees of freedom, it is easy to implement and straightforwardly generalizes the popular fixed effects estimator.

When the production function (8) is estimated with the semiparametric estimator of Olley and Pakes [1996], one can calculate a direct estimate of $\hat{\omega}_{it}$, purged from random noise $\hat{\varepsilon}_{it}$, but only for firms with non-missing investment. The semiparametric part of the production function estimated in the first stage (ϕ) combines the effects of productivity and all other state variables. As this is computable as a nonparametric function of observable variables, we can obtain a productivity estimate as follows:[45]

$$\text{(B.3)} \qquad \ln A_{it}^{OP2} - \overline{\ln A}_t^{OP2} = (\hat{\phi}_{it} - \overline{\hat{\phi}_t}) - \hat{\alpha}_k (k_{it} - \overline{k_t})$$

$$\text{(B.4)} \qquad \ln A_{it}^{OP2} - \ln A_{it-1}^{OP2} = (\hat{\phi}_{it} - \hat{\phi}_{it-1}) - \hat{\alpha}_k (k_{it} - k_{it-1}).$$

[45] Yet another alternative estimator, described in Ackerberg *et al.* [2005], also purges the random error from the production function in a first stage, but treats variable and quasi-fixed inputs similarly.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

TABLE B.I
TWO ALTERNATIVE ESTIMATORS

Productivity is the sum of an AR(1) process and a fixed effect				
Stochastic frontier alternative (SF2)				
	correlation with true productivity		coefficient bias for	
	level (1a)	growth (2a)	labor (3a)	capital (4a)
wage heterogeneity	0.898	0.182	0.019	−0.111
errors in output	0.048	0.019	0.384	−0.387
errors in output with W.H.*	0.406	0.033	0.095	−0.157
errors in inputs	0.764	0.067	−0.527	−0.372
technology diff. (CRS)	0.250	0.151	0.373	−0.392
technology diff. with W.H.* (CRS)	0.752	0.151	0.088	−0.195
technology diff. with W.H.* (VRS)	0.302	0.117	0.226	−0.189
Olley & Pakes alternative (OP2)				
	(1b)	(2b)		
wage heterogeneity	0.866	0.385		
errors in output	−0.214	−0.009		
errors in output with W.H.*	0.783	0.309		
errors in inputs	0.761	0.281		
technology diff. (CRS)	0.469	0.119		
technology diff. with W.H.* (CRS)	0.771	0.301		
technology diff. with W.H.* (VRS)	0.126	0.173		

Notes: Statistics are comparable to the results in Tables 1–3. The biases in coefficient estimates for OP2 are the same as for OP in Tables 1–3.
*Wage heterogeneity added to aid identification of production function parameters (as before).

The OP2 productivity estimate only captures the part of productivity known to the firm at the time it chooses investment, not any subsequent innovation in productivity that might still contribute to output. In this sense, its interpretation differs from the other estimators as it explicitly takes out idiosyncratic productivity shocks or shocks that are realized after inputs are chosen.

The results for these two estimators in Table A.1 are comparable to the three experiments considered in Tables I–III.[46] Their advantage shows up in the results with measurement error (lines 2–4 in both panels of Table A.1 are comparable to the results in Table II). The two new estimators outperform their alternative implementations—even all other methods—when measurement errors are large. When only output is measured very imprecisely, they are clearly superior. The correlation with true productivity level for the SF2 estimator is at least twice as high as for SF: 0.048 versus 0.024 without any wage heterogeneity and 0.406 versus 0.178 in the more realistic case. The OP2 estimator only works well when there is at least some wage heterogeneity, in which case it obtains a surprisingly high correlation with true productivity levels of

[46] We limit the comparison to the results where productivity is the sum of an AR(1) process and a fixed effect, i.e., from the nine panels in Tables 1–3, we omit results comparable to the middle and bottom panels of Table 1.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

0.783 versus 0.292 for OP.[47] Interestingly, when only inputs are measured imprecisely, SF2 and OP2 still have advantages. Even though the correlation differences are reduced, they still improve over SF and OP: 0.752 versus 0.725 for SF and 0.771 versus 0.734 for OP. Still, the correlation statistic for the GMM estimator (0.776) is marginally higher.

Their advantage is almost as large for estimating productivity growth in the presence of measurement error, as for productivity levels. However, the uniformly low correlations in columns (2a) and (2b) indicate that these approaches are not preferable if one is mainly interested in estimating productivity growth. The SF2 and OP2 productivity growth correlations for the exercises in Table I and III are that much lower than for SF and OP, that it seems extremely risky to use the alternatives to estimate productivity growth. For productivity levels, correlations for SF and OP also tend to be higher in Table I and for the CRS results in Table III, but the differences with SF2 and OP2 are relatively small. Furthermore, if measurement errors are firm-specific, the error in the growth estimates ($\varepsilon_{it} - \varepsilon_{it-1}$) is likely to be relatively small and randomly distributed in the sample, leading to smaller effects on the productivity growth correlations.

REFERENCES

Ackerberg, D. A.; Caves, K. and Frazer, G., 2005, *Structural Identification of Production Functions* (mimeo: University of California at Los Angeles).

Aigner, D.; Lovell, C. A. K. and Schmidt, P., 1977, 'Formulation and Estimation of Stochastic Frontier Production Function Models,' *Journal of Econometrics*, 6 (1), pp. 21–37.

Baily, M. N.; Hulten, C. and Campbell, D., 1992, 'Productivity Dynamics in Manufacturing Plants,' *Brookings Papers: Microeconomics*, 4 (1), pp. 187–267.

Bartelsman, E. J. and Dhrymes, P. J., 1998, 'Productivity Dynamics: U.S. Manufacturing Plants, 1972–1986,' *Journal of Productivity Analysis*, 9 (1), pp. 5–34.

Bartelsman, E. J. and Doms, M., 2000, 'Understanding Productivity: Lessons from Longitudinal Microdata,' *Journal of Economic Literature*, 38 (3), pp. 569–594.

Battese, G. E. and Coelli, T. J., 1992, 'Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India,' *Journal of Productivity Analysis*, 3 (1–2), pp. 153–169.

Berndt, E. R. and Khaled, M. S., 1979, 'Parametric Productivity Measurement and Choice among Flexible Functional Forms,' *Journal of Political Economy*, 87 (6), pp. 1220–1245.

Blundell, R. W. and Bond, S. R., 1998, 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,' *Journal of Econometrics*, 87 (1), pp. 115–143.

Blundell, R. W. and Bond, S. R., 2000, 'GMM Estimation with Persistent Panel Data: An Application to Production Functions,' *Econometric Reviews*, 19 (3), pp. 321–340.

Bond, S. R. and Söderbom, M., 2005, 'Adjustment Costs and the Identification of Cobb Douglas Production Functions,' IFS working papers W05-04.

Caves, D. W.; Christensen, L. R. and Diewert, E. W., 1982a, 'The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity,' *Econometrica*, 50 (6), pp. 1393–1414.

[47] It should be noted that if the error added to output is a transitory productivity shock not observed to the firm, which is an alternative interpretation for the results in Table 1, SF2 and OP2 would erroneously subtract the shock from the productivity estimate.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

Caves, D. W.; Christensen, L. R. and Diewert, E. W., 1982b, 'Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers,' *Economic Journal*, 92 (365), pp. 73–86.
Charnes, A.; Cooper, W. W. and Rhodes, E., 1978, 'Measuring the Efficiency of Decision Making Units,' *European Journal of Operational Research*, 2 (6), pp. 429–444.
Cornwell, C.; Schmidt, P. and Sickles, R. C., 1990, 'Productivity Frontiers with Cross-sectional and Time Series Variation in Efficiency Levels,' *Journal of Econometrics*, 46 (1–2), pp. 185–200.
Farrell, M. J., 1957, 'The Measurement of Productive Efficiency,' *Journal of the Royal Statistical Society, Series A*, 120 (3), pp. 253–290.
Feenstra, R. C. and Reinsdorf, M. B., 2003, 'Should Exact Index Numbers Have Standard Errors? Theory and Application to Asian Growth,' NBER working paper, no. 10197.
Gagné, R. and Ouellette, P., 1998, 'On the Choice of Functional Forms: Summary of a Monte Carlo Experiment,' *Journal of Business and Economic Statistics*, 16 (1), pp. 118–124.
Griliches, Z. and Hausman, J. A., 1986, 'Errors in Variables in Panel Data,' *Journal of Econometrics*, 31 (1), pp. 93–118.
Griliches, Z. and Mairesse, J., 1998, 'Production Functions: The Search for Identification,' in Strom, S. (ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, (Cambridge University Press, Cambridge). pp. 169–203.
Katayama, H.; Lu, S. and Tybout, J. R., 2005, 'Firm-level Productivity Studies: Illusions and a Solution' (mimeo: Penn State University).
Klette, T. J., 1999, 'Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data,' *Journal of Industrial Economics*, 47 (4), pp. 451–476.
Klette, T. J. and Griliches, Z., 1996, 'The Inconsistency of Common Scale Estimators when Output Prices Are Unobserved and Endogenous,' *Journal of Applied Econometrics*, 11 (4), pp. 343–361.
Levinsohn, J. and Petrin, A., 2003, 'Estimating Production Functions Using Inputs to Control for Unobservables,' *Review of Economic Studies*, 70 (2), pp. 317–342.
Mairesse, J. and Griliches, Z., 1990, 'Heterogeneity in Panel Data: Are There Stable Production Functions?,' in Champsaur, P. *et al.* (eds.), *Essays in Honor of Edmond Malinvaud, Volume 3*, (MIT Press, Cambridge, Massachusetts). pp. 125–147.
Marschak, J. and Andrews, W. H., 1944, 'Random Simultaneous Equations and the Theory of Production,' *Econometrica*, 12 (3–4), pp. 143–205.
Meeusen, W. and van den Broeck, J., 1977, 'Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error,' *International Economic Review*, 18 (2), pp. 435–444.
Olley, G. S. and Pakes, A., 1996, 'The Dynamics of Productivity in the Telecommunications Equipment Industry',' *Econometrica*, 64 (6), pp. 1263–1297.
Roberts, M., 1996, 'Colombia, 1977–85: Producer Turnover, Margins and Trade Exposure,' in Roberts, M. and Tybout, J. R. (eds.), *Industrial Evolution in Developing Countries*, (Oxford University Press for the World Bank, New York). pp. 227–259.
Roberts, M. and Tybout, J. R., 1996, *Industrial Evolution in Developing Countries*, (Oxford University Press for the World Bank, New York).
Seiford, L. L. and Thrall, R. M., 1990, 'Recent Developments in DEA. The Mathematical Programming Approach to Frontier Analysis,' *Journal of Econometrics*, 46 (1–2), pp. 7–38.
Solow, R. M., 1957, 'Technical Change and the Aggregate Production Function,' *Review of Economics and Statistics*, 39 (3), pp. 312–320.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

Syverson, C. W., 2001, *Market Structure and Productivity*, Ph.D. dissertation, University of Maryland.

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

Van Biesebroeck, J., 2003b, 'Revisiting some Productivity Debates', forthcoming, *Journal of Business and Economic Statistics*.

Van Biesebroeck, J., 2004, 'Robustness of Productivity Estimates', NBER working paper, no. 10303.

Zellner, A.; Kmenta, J. and Drèze, J. H., 1966, 'Specification and Estimation of Cobb-Douglas Production Function Models,' *Econometrica*, 34 (4), pp. 784–795.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.