

USING FIRM OPTIMIZATION TO EVALUATE AND ESTIMATE RETURNS TO SCALE

Yuriy Gorodnichenko
University of California, Berkeley

First draft: April 2005
This version: August 2007

Abstract

At the firm level, revenue and costs are well measured but prices and quantities are not. This paper shows that because of these data limitations estimates of returns to scale at the firm level are for the revenue function, not production function. Given this observation, the paper argues that, under weak assumptions, micro-level estimates of returns to scale are often inconsistent with profit maximization or imply implausibly large profits. The puzzle arises because popular estimators ignore heterogeneity and endogeneity in factor/product prices, assume perfect elasticity of factor supply curves or neglect the restrictions imposed by profit maximization (cost minimization) so that estimators are inconsistent or poorly identified. The paper argues that simple structural estimators can address these problems. Specifically, the paper proposes a full-information estimator that models the cost and the revenue functions simultaneously and accounts for unobserved heterogeneity in productivity and factor prices symmetrically. The strength of the proposed estimator is illustrated by Monte Carlo simulations and an empirical application. Finally, the paper discusses a number of implications of estimating revenue functions rather than production functions and demonstrates that the profit share in revenue is a robust non-parametric economic diagnostic for estimates of returns to scale.

Keywords: production function, identification, returns to scale, covariance structures.

JEL classification: C23, C33, D24

I am grateful to Susanto Basu, Robert Chirinko, Olivier Coibion, Erwin Diewert, Juan Carlos Hallak, Lutz Kilian, Patrick Kline, Edward Knotek, Yuriy Mokrousov, Alice Nakamura, Serena Ng, Amil Petrin, John van Reenen, Matthew Shapiro, Jan Svejnar and conference and seminar participants in Michigan, Midwest Econometrics Group and the NBER Summer Institute for helpful comments. Correspondence: ygorodni@umich.edu or Yuriy Gorodnichenko, Department of Economics, University of Michigan, 611 Tappan St, Ann Arbor, MI 48109.

Copyright by Yuriy Gorodnichenko. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including notice, is given to the source.

1 INTRODUCTION

Production functions estimated on establishment level data provide essential insights into micro- and macroeconomic phenomena. Estimated returns to scale and measured productivity differences are important inputs to economic theories and policy analyses. However, micro level datasets have significant limitations and recognizing these limitations can dramatically change the economic interpretation of estimated parameters and various measures of relative performance.

Specifically, at the establishment level revenue and costs are well measured but prices and quantities are not.¹ Economic theory suggests that firms with at least some degree of monopoly power should charge different prices if there is heterogeneity in productivity across firms. Since most firms face a downward sloping demand curve, it is not surprising that there is, as I argue below, overwhelming evidence of firms charging different prices even for highly homogenous products. Since the purpose of using micro-level data is often to unearth differences in productivity across firms, it follows that one should be careful in distinguishing quantities and revenues as the latter is a product of quantities and prices varying across firms. In most studies, however, firm revenues are typically deflated by industry price indices to get a measure of quantity. I argue that this measure of quantity, when used as the dependent variable in production function regressions, is effectively the firm's revenue rather than (physical) output precisely because firms have different productivity levels and, consequently, charge different prices. Thus, the estimated returns to scale are returns in *revenue*, not production. This has important implications for the estimation and interpretation of returns to scale as well as productivity measurement. Despite these data limitations, micro level estimates can still be highly informative for positive and normative economics and one should address two outstanding questions. First, does economic theory make predictions about plausible magnitudes of returns to scale in the revenue function? Second, how could one estimate consistently returns to scale with the data actually available?

To answer the first question, I show that under weak assumptions the profit share is intimately related to the elasticity of the total cost with respect to inputs, returns to scale in production, and the markup. With the standard assumption of perfectly elastic factor supply

¹ I use firm and establishment interchangeably.

curves, I demonstrate that returns to scale in the revenue function are equal to one minus the profit share. Given that the share of economic profits is small, there is a tight restriction on the estimates of returns to scale in the revenue function. Importantly, since I make only a few assumptions about the nature of production functions, costs, and market structure, the profit share can serve as a simple litmus test for whether the estimates of returns to scale make economic sense. As I show below, many of the available estimates fail to pass the simple test of this diagnostic.

To answer the second question, it is important to remember that estimation of production (revenue) function parameters is inseparable from simultaneity problem because the volumes of inputs and output are optimizing choices of firms. This fundamental problem is particularly acute in single-equation approaches because in these approaches the researcher does not model the choice of inputs while, as Marschak and Andrews (1944) warn, one cannot treat inputs as independent variables. Modeling the choice of inputs should be an integral part of the production (revenue) function estimation. I show that that one can turn the simultaneity weakness of single-equation estimators into the strength of simple structural estimators as optimizing choices of inputs and output help in identification and estimation of the deep parameters governing the behavior of firms.

Specifically, I extend the full information maximum likelihood estimator of Marschak and Andrews (1944) and Schmidt (1988) to dynamic production (revenue) function models with serially correlated measurement errors and factor prices correlated with productivity. This estimator, which I call the covariance estimator, deals simultaneously with the production and cost sides and with unobserved technology and factor prices. The key idea of the estimator is to use the covariance structures for the firms' observed optimizing choices (inputs, output) to identify and estimate parameters of interest using the restrictions imposed by the economic model on the response of observed variables to unobservables such as productivity and factor prices. In this estimator, the researcher not only focuses on the production (revenue) function relationship but also utilizes information from firm's optimizing choices of inputs. The covariance estimator is easy to implement and interpret since the estimator can have an instrumental variable interpretation.

The covariance estimator adds to a large set of alternative estimators of production (revenue) functions and, therefore, it is important to contrast this estimator with popular rival estimators. I use both simulations and economic arguments to highlight the differences. First, I

show that the covariance estimator outperforms popular alternatives in Monte Carlo experiments in a wide range of setups including empirically important cases such as correlated productivity and factor prices, serially correlated measurement errors, upward sloping factor supply curves, etc. Note that previous estimators cannot handle some or all of these cases. Furthermore, my simulations suggest that popular estimators often yield productivity measures that are poorly correlated with true productivity and, thus, the researcher or policymaker can reach strikingly different (and likely incorrect) conclusions about the relative productivity of firms and the magnitudes of productivity differences. On the other hand, the covariance estimator measures productivity well. The covariance estimator also yields economically more reasonable estimates than those achieved by alternative estimators when confronted with real data.

Second, I use the profit share to show that under the standard assumption of perfectly elastic factor supply curves many estimates of returns to scale are inconsistent with profit maximization or imply implausible large profits. Specifically, returns to scale (RTS) in the revenue function cannot exceed unity otherwise the profit share in revenue is negative. Estimates of RTS frequently exceed unity not only in simple least squares cases (e.g., Griliches and Ringstad 1971, Tybout and Westbrook 1996, Bartelsman and Dhrymes 1998), but also after correcting for the endogeneity of inputs—i.e., the transmission bias (e.g., Pavcnik 2002, Levinsohn and Petrin 2003). In other words, these estimates suggest that firms systematically violate the profit maximization principle. At the other extreme, studies often find low returns to scale that imply a much larger profit share in revenue than is observed in the data. For example, 0.8 returns in the revenue function entails that the share of pure economic profits in revenue is 20% (or 50% in value added if the share of materials is 0.6). In most data, the profit share is 3% or less (Rotemberg and Woodford 1995, Basu and Fernald 1997).

These results raise legitimate concerns about the validity of the applied economic model and statistical estimator. I show that one can reconcile increasing returns to scale in the revenue function and a small profit share by relaxing the assumption of perfectly elastic factor supply curves. Likewise, one can explain large decreasing returns to scale and a small profit share. Unfortunately, available estimators either do not estimate the elasticity of the cost or depend critically on the assumption that factor supply curves are perfectly elastic. Thus, the researcher using these alternative assumptions about factor supply can be unable to check if the estimates make economic sense when he or she uses single-equation approaches.

Furthermore and most importantly, some of the popular estimators ignore that firms optimize given their technology and factor/product prices. Sweeping the latter variation under the “*ceteris paribus*” rug can greatly distort estimates of returns to scale, measures of productivity and resource reallocation, calibration of economic models, etc. I emphasize that consistent estimation requires modeling not only unobserved technology but also unobserved factor prices and, possibly, other structural shocks. It is equally important to model both the revenue and cost sides of optimizing firms. Finally, optimization imposes restrictions on how firms react to changes in technology and prices and, thus, makes certain moment conditions redundant. I show that this reduction in the number of informative moments can be so acute that certain estimators become not identified. Likewise, tight theoretical restrictions on contemporaneous and dynamic responses of observed choices of firms to structural shocks such as productivity suggest that estimators based on inverting factor demands to construct proxy variables for unobserved productivity can be underidentified. In fact, the problem can be so acute that these types of estimators can be forced to make internally inconsistent assumptions to “achieve” identification, which means that these estimators cannot yield consistent estimates even in theory. I demonstrate that puzzling estimates of returns to scale can be an artifact of these misspecifications while simple structural estimators such as the proposed covariance estimator can address these problems.

In summary, this paper makes three important contributions to measuring productivity and estimating returns to scale with the micro level data. First, I show that because of data limitations many of the available estimates of returns to scale are for the revenue function, not production function. I argue that this alternative concept, the returns to scale in revenue, is economically meaningful and can be robustly estimated with the data that are actually observable at the establishment level. Second, I show how one can fruitfully merge economic theory and statistical methods into simple structural estimators to obtain consistent estimates of returns to scale. Furthermore, I prove that the profit share (i.e., the share of economic profits in total revenue) can work as a robust, non-parametric diagnostic for returns to scale in the revenue function. Third, I demonstrate that many of the popular estimators of returns to scale are inconsistent with simple economic arguments and tend to perform poorly in a wide range of settings.

In the following section of the paper I elaborate on the ideas and claims I make in this introduction. In the next section, I present theoretical results, discuss the sources of identification

in production (revenue) functions, and examine the variables used in the production (revenue) function regressions. In Section 3, I present the covariance estimator and discuss identification and estimation issues. In Section 4, I derive the theoretical predictions about the performance of OLS, instrumental variables and inversion estimators. Monte Carlo experiments in Section 5 illustrate the performance of alternative estimators. In Section 6, I use a well-known Chilean firm-level data to compare RTS estimates from the covariance estimator and popular alternatives. I present conclusions in Section 7.

2 SETUP

In this section, I derive the relationship between the markup, returns to scale in production, the elasticity of the cost and the profit share. I demonstrate that the profit share can serve as a robust non-parametric diagnostic for economic tests of the estimates of production (revenue) functions.

2.1 ECONOMIC MODEL OF PRODUCER BEHAVIOR

Consider a firm that minimizes cost in expectation or non-stochastically. I assume that the cost of inputs is separable in inputs and factor prices, i.e., cross-partial derivatives of the cost with respect to factor prices and inputs are equal to zero. Hence, the cost can be written as

$C(L, w) = \sum_{j=1}^n C_j(L_j, w_j)$ where L and w are vectors of inputs and factor prices, L_j is the j^{th} input, and w_j is its price. The elasticity of the cost C_j with respect to input j is $\phi_j = \frac{\partial C_j(L_j, w_j)}{\partial L_j} \cdot \frac{L_j}{C_j(L_j, w_j)}$. The share of input j in total cost is $\omega_j = C_j(L_j, w_j) / C(L, w)$.

Returns to scale in production γ is defined as $\gamma = \sum_{j=1}^n (\partial Q / \partial L_j) L_j / Q(L)$, where $Q(L)$ is the production function. Analogously, RTS in the revenue function η is defined as $\eta = \sum_{j=1}^n (\partial Y / \partial L_j) L_j / Y(L)$, where Y is total revenue. I define the markup μ as the ratio of the output price to the marginal cost. The share of economic profits in revenue (henceforth, profit share) is $s_\pi = (Y - C) / Y$. Note that I make no assumptions about the production function or the structure of product and factor markets.

To simplify exposition, assume that firms freely adjust factors of production to avoid unnecessary complications arising from dynamic optimization. This assumption implies that firms solve a *static* profit maximization problem in every period and inputs and output are

chosen simultaneously. One can interpret this assumption as describing a large cross-section of firms or the long run when firms can adjust all inputs. In this general setup, the following result can be proven:

Proposition 1.

Suppose a firm minimizes cost, all inputs are variable, and its cost is separable in inputs. Then $\gamma/\mu = (1 - s_\pi)\phi$, where μ is the markup, $\phi = \sum_{j=1}^n \phi_j \omega_j$ is the elasticity of the cost with respect to inputs, ϕ_j is the elasticity of the j^{th} factor cost, ω_j is the share of input j in total cost, γ is returns to scale in production, and s_π is the profit share in revenue. Furthermore, if the firm maximizes profit, then $\eta = \gamma/\mu$, where η is returns to scale in the revenue function.

Proof: see Appendix B.

One can draw several conclusions from Proposition 1. First, consider the case where factor supplies are perfectly elastic (i.e., $\phi_j = 1$ for all j). Since the profit share s_π is close to zero (Rotemberg and Woodford 1995, Basu and Fernald 1997), by Proposition 1 the returns in the revenue function η , which is equal to γ/μ , should be approximately unity. Furthermore, industries with large RTS in production γ should have a large markup μ such that $\mu \approx \gamma$. Hence, finding constant RTS in revenue is likely to indicate that there are increasing RTS in production since the markup is often greater than 1.05-1.1 (e.g., Bresnahan 1988). Proposition 1 also shows that low RTS in the revenue function imply a large profit share. For instance, $\eta=0.8$ implies $s_\pi=20\%$.² Similarly, finding $\eta>1$ is not consistent with profit maximization since $\eta>1$ implies a negative profit share. More generally, if the profit share implied by an estimate of η is far from the profit share observed in the data, then one has a signal that either the statistical or economic model is incorrect. This point is first raised by Basu and Fernald (1997) in the context of estimating aggregate production functions. Because Proposition 1 makes weak assumptions about producer behavior, the profit share serves as a robust non-parametric economic diagnostic for statistical estimates of η .

² Griliches and Hausman (1986) attribute low returns to large, (possibly) serially correlated measurement errors, which are hard to handle in the instrumental variables (IV) framework as there are few good instruments to cope with measurement errors. It is not clear, however, why measurement errors are so pervasive in some industries and not others.

Second, consider the case where factor supplies are not perfectly elastic (i.e., $\phi_j \neq 1$ for some j). In this case, there is no tight link between the profit share s_π and RTS in the revenue function η . Increasing RTS in the revenue function η and a small positive profit share s_π can be reconciled by a steep cost (i.e., large ϕ). For example, monopsony power, rent sharing or a shift premium can result in an upward-sloping labor supply schedule. Likewise, decreasing RTS in revenue or production functions can be consistent with a small profit share if ϕ is less than unity, i.e., the marginal unit cost of inputs is (locally) declining. Table 1 summarizes the relationship between s_π , η and ϕ .

Note that in the case with $\phi \neq 1$ the cost-based Solow residual does not measure technology (or revenue-generating ability) correctly because cost shares are not equal to the elasticities of output with respect to corresponding inputs. Specifically, the cost-based Solow residual depends on factor ratios and thus can be procyclical and serially correlated.

For the case with some inputs being fixed, Proposition 1 needs a slight modification:

Corollary 1

Suppose that the assumptions of Proposition 1 hold. Also suppose that the first k inputs are variable and the other $n-k$ inputs are fixed. Then, $\gamma^ / \mu = (1 - s_\pi^*) \phi^* \omega^*$, where $\phi^* = \sum_{j=1}^k \phi_j \omega_j$ is the elasticity of the cost with respect to variable inputs, ω^* is the cost share of variable inputs in total cost, γ^* is returns to scale in production with respect to variable inputs, and s_π^* is the profit share in revenue. Furthermore, if the firm maximizes profit, then returns to scale in revenue with respect to variable inputs is $\eta^* = \gamma^* / \mu$.*

Proof: see Appendix B.

Corollary 1 suggests that the argument about the profit share should be applied to variable inputs only. The corollary explains that the profit share can be temporarily large since γ^* can be significantly less than unity or temporarily small since the short term elasticity of the variable factor supplies can be low (i.e., ϕ^* large). Since there is no optimization with respect to fixed inputs, cross-sectional variation in the fixed inputs is sufficient to identify the RTS with respect to fixed inputs and, hence, RTS with respect to all inputs.

In summary, Proposition 1 justifies using the profit share as an economic check to verify that statistical estimates of returns to scale make economic sense.³ Put differently, since the parameter ϕ can be interpreted as RTS in the cost, RTS in the revenue function η is always less than RTS in the cost but the difference is small. Furthermore, since the profit share is typically small, a consistent estimate of RTS in the revenue function can inform the researcher about the properties of the cost, specifically ϕ . Likewise, one can infer η from ϕ . In addition, using $\eta = \gamma / \mu$, one can get an estimate of RTS in production γ (markup μ) provided a measure of markup μ (RTS in production γ) is available.

2.2 FIRST ORDER APPROXIMATION

To make further progress in the analysis of estimated RTS, I make a few assumptions about production, demand, and cost.⁴ Specifically, the inverse demand function is $P_{it} = D(G_{it}, Q_{it}) = D_1(G_{it}) \cdot D_2(Q_{it})$, where i and t index firms and time, P_{it} is the price of the good, Q_{it} is the quantity of the good, G_{it} is a separable demand shifter (e.g., quality of a good, macroeconomic conditions). If σ is the elasticity of demand, then the markup is $\mu = \sigma / (\sigma - 1)$. The production function is $Q_{it} = F(A_{it}^\mu, Z_{it}) = F_1(A_{it}^\mu) \cdot F_2(Z_{it})$ where A_{it} is Hicks-neutral firm-specific productivity (the power of A_{it} is a normalization to simplify notation), and Z_{it} is a composite input. Inputs are measured in physical units. The cost of employing Z_{it} is $C(W_{it}, Z_{it}) = C_1(W_{it}) \cdot C_2(Z_{it})$ where W_{it} is the separable base price of the input. To be consistent with previous notation, γ is local returns to scale in production and ϕ is the elasticity of the cost with respect to the input Z_{it} . The case of $\phi = 1$ corresponds to supply of Z_{it} being perfectly elastic. Hence, profits are $\pi_{it} \equiv Y_{it} - C(W_{it}, Z_{it})$, where $Y_{it} = P_{it}Q_{it}$ is the revenue function. The profit function is (locally) concave in the input if and only if $\gamma / \mu - \phi = -s_\pi \phi < 0$.

After log-linearizing the first order conditions, suppressing uninteresting constants, and partialing out industry-wide shocks, one obtains the following expressions for profit-maximizing input and revenue:

$$z_{it} = \frac{1}{\eta - \phi} w_{it} - \frac{1}{\eta - \phi} (a_{it} + g_{it}), \quad (1)$$

³ Of course, some caution is warranted since measured profits can deviate from economic profits.

⁴ This model of producer behavior is similar to the model analyzed by Marschak and Andrews (1944) and Klette and Griliches (1996).

$$y_{it} = \frac{\eta}{\eta - \phi} w_{it} - \frac{\phi}{\eta - \phi} (a_{it} + g_{it}), \quad (2)$$

where small letters denote log deviation of the respective variables from steady state (or industry averages), and $\eta = \gamma/\mu$ is the RTS in the revenue function. Note that demand shocks G_{it} and technology shocks A_{it} are isomorphic and, thus, are not identified separately.⁵ Henceforth, I treat G_{it} as if it were a shock to technology and consider only A_{it} . It will be convenient in further analysis to write (1)-(2) in matrix form:

$$\mathbf{X}_{it} \equiv \begin{bmatrix} z_{it} \\ y_{it} \end{bmatrix} = \begin{bmatrix} \frac{1}{\eta - \phi} & \frac{-1}{\eta - \phi} \\ \frac{\eta}{\eta - \phi} & \frac{-\phi}{\eta - \phi} \end{bmatrix} \begin{bmatrix} w_{it} \\ a_{it} \end{bmatrix} \equiv \Lambda \mathbf{F}_{it}. \quad (3)$$

Equations (1) and (2) indicate that output and input demand are increasing in productivity a_{it} and decreasing in the factor price w_{it} . Equations (1)-(2) are a first-order log-linear approximation to the optimal behavior of firms. This approximation is exact if demand and factor supply are isoelastic and production function is Cobb-Douglas. Since variation in technology a_{it} across firms is not controversial (e.g., Bartelsman and Doms 2000), in the next section I focus on factor price w_{it} as a source of variation in (1)-(2).

2.3 ON SOURCES OF VARIATION

In the model (1)-(2), I use variation in the factor price w_{it} to address two stylized facts. First, inputs and output are not collinear in the data. Second, there is enormous variation in input mixes. For example, the interquantile ($Q_3 - Q_1$) range of $\log(\text{capital/labor})$ and $\log(\text{materials/labor})$ for Chilean and U.S. manufacturing firms is typically above 100% even at four-digit SIC industries. Note that in any model that assumes Hicks-neutral technology such variation in input mixes can happen only if firms face different input prices or technology or firms cannot satisfy profit maximizing (cost minimizing) conditions (e.g., because of managerial errors).

This paper does not seek to explain why firms face different input prices. Possible reasons include unionization, regulation, location, composition of capital, and subjective beliefs of the management about factor prices. Search and information costs result in equilibrium price dispersion even if firms are identical ex ante (e.g., Stigler 1961, Salop and Stiglitz 1982, Burdett and Judd 1983, Stahl 1989).

There is substantial direct evidence on the dispersion of prices even for precisely defined products (Stigler 1961, Pratt, Wise and Zeckhauser 1979, Dahlby and West 1986, Abbott 1992,

⁵ Under stronger assumptions it is possible to separate demand and technology shocks. For example, Katayama, Lu and Tybout (2003) assume Bertrand pricing and constant marginal cost to identify demand and technology shocks.

Sorensen 2000). Using firm-level U.S. Census data, Abbott (1992) reports that the mean coefficient of variation for output prices at 7-digit product codes is at least 55% (see also Roberts and Supina 1996). Even prices of homogenous inputs such as cement have significant dispersion at local markets (Abbott 1992, Adams 1997, Lach 2002, Yoskowitz 2002). For 70% of firms, other firms are the main customers (Fabiani et al 2004) and, thus, such price dispersion is an important source of variation in input mixes.

Likewise, there is voluminous evidence that similar workers are paid different wages (e.g., Mortensen 2003 and references cited therein). Abowd, Creedy and Kramarz (2002) find that approximately 40-50% of wage dispersion in France and the state of Washington in U.S. is determined by firm effects. Price dispersion in capital/financial markets is less documented yet it exists (see Hortaçsu and Syverson (2004) for an example of dispersion of fees charged by mutual index funds). Multiplicity of interest rates also suggests that different firms face different prices of capital even within the same industry and location. Furthermore, firms may have different *shadow* prices of inputs (because of adjustment costs, for example) even when they face the same posted market prices for inputs.

There are alternative explanations for variation in input mixes. Early studies (e.g., Marschak and Andrews 1944, Hoch 1961, Zellner, Kmenta and Dreze 1966) assumed that managerial errors determine the variation in input ratios. In another interpretation (e.g., Stigler 1976, McElroy 1987), managerial errors reflect constraints known to the management but unobserved by the econometrician.⁶

Although the managerial errors theory may be right, it can hardly explain immense variation in input mixes. (Recall that the interquartile range of log input ratios is generally above 100%.) In addition, all measures of dispersion for input ratios increase with aggregation. It is hard to reconcile these facts with managerial errors theory because there is no reason to expect

⁶ Another explanation of variability in input mixes is variation in parameters of the production function. A typical approach to estimate models with parameter heterogeneity (e.g., Mairesse and Griliches 1990, Biorn, Lindquist and Skjerpen 2002) is to use the random coefficients estimator (Swamy 1970) that assumes zero covariance between random coefficients and regressors. This assumption is, however, clearly violated in the context of production functions if management knows the parameters of its production function. Consider the model in (1)-(2) with no measurement errors, $\mu = 1$ and random firm-specific RTS parameter γ_i such that $\gamma_i \sim iid(\bar{\gamma}, \sigma_\gamma^2)$ and $cov(w_{it}, \gamma_i) = cov(a_{it}, \gamma_i) = 0$. The estimated model is $y_{it} = \gamma_i z_{it} + u_{it}$. It is not hard to find $cov(x_{it}, \gamma_i) \approx -[\gamma^2(\gamma - 1)(\gamma - 2) + \sigma_\gamma^2]/(\gamma - 1)^2 < 0$. Because $cov(x_{it}, \gamma_i) \neq 0$, the estimator is not consistent and results should be interpreted very carefully.

that managerial errors become more important with aggregation. In contrast, variation in prices for labor, capital and materials can approximately explain the volatility in input mixes.⁷

Differences in interpretation, however, do not generally imply differences in estimates of RTS. For example, suppose that factor prices are the same across firms and consider a Cobb-Douglas production function with labor L_{it} and capital K_{it} inputs and managerial errors $\zeta_{it}^K, \zeta_{it}^L$ in the first order conditions so that $\beta_K Y_{it}/K_{it} = R_t \exp(\zeta_{it}^K)$ and $\beta_L Y_{it}/L_{it} = W_t \exp(\zeta_{it}^L)$, where β_K and β_L are elasticities of the revenue function with respect to capital and labor, Y_{it} is revenue, R_t is the cost of capital and W_t is wages. After taking logs and ignoring uninteresting constants, one has $y_{it} = k_{it} + \zeta_{it}^K$ and $y_{it} = l_{it} + \zeta_{it}^L$. If one assumes firm-specific factor prices, the corresponding first order conditions lead to $y_{it} = k_{it} + r_{it}$ and $y_{it} = l_{it} + w_{it}$. Thus, the models are observationally equivalent and give identical estimates of parameters in the revenue function. As a result, I treat factor prices as generic shocks to input ratios.

2.4 WHAT DO “PRODUCTION FUNCTION” REGRESSIONS ESTIMATE?

Firm-level data sets (e.g., Longitudinal Business Database at the U.S. Census Bureau) rarely contain information about prices paid/charged by firms or quantities consumed/produced by firms. In the vast majority of cases, the econometrician observes only inputs and revenue of the firm and, hence, a typical production function regression is

$$y_{it} - \bar{p}_t = bz_{it} + u_{it}, \quad (4)$$

where \bar{p}_t is the industry price index, b is estimated RTS, u_{it} is the error term, and the dependent variable is the firm revenue deflated by industry price index.⁸ In the standard framework of monopolistic competition (Dixit and Stiglitz 1977), the demand function is

$$p_{it} = \bar{p}_t - \frac{1}{\sigma} q_{it} + g_{it} + const \text{ and}$$

⁷ For example, Abowd, Creedy and Kramarz (2002) report that the standard deviation of log real hourly wages is 53%. If one takes the coefficient of variation as a proxy for the standard deviation of log deviations from the mean, then the standard deviation of material prices is 55% (Abbott 1992) at the 7-digit level. At the 4-digit level, the standard deviation is likely to be several times larger. Hence, variation in the ratio of prices for labor and materials, which is equal to $\log(\text{labor/materials})$, can be as large as 100%. Likewise, the standard deviation of log fees in mutual funds is about 50% (Hortaçsu and Syverson 2004), which, however, can be an upper bound. Hence, variation in the log wage-to-capital price, which is equal to the log labor-to-capital ratio, can also be as volatile as 100%. Of course, the observed variation can be endogenous, yet this calculation is suggestive. Note that this simple calculation ignores possible variation in shadow prices which can be considerably larger than the variation in posted prices because shadow prices can differ across firms due to adjustment costs, complementarity of inputs, composition of inputs (especially vintages of capital), etc.

⁸ Foster, Haltiwanger and Syverson (2005) is an important exception. They consider firms producing homogenous goods so that information on revenue and physical output is available.

$$\begin{aligned}
y_{it} - \bar{p}_t &= (p_{it} + q_{it}) - \bar{p}_t = (1 - \frac{1}{\sigma})q_{it} + g_{it} + const \\
&= (1/\mu)(\gamma z_{it} + \mu a_{it}) + g_{it} + const = \eta z_{it} + (a_{it} + g_{it}) + const. \quad (5)
\end{aligned}$$

Clearly, the coefficient b in (4) reflects returns in the revenue function η , not returns to scale in production γ . Furthermore, because firms face different productivity and/or wage realizations, the price of the good varies across firms and, as I discussed in the previous section, dispersion of prices is not trivial even in narrowly defined industries. Therefore, deflating the firm's revenue with an industry price index \bar{p}_t does not generally yield the firm's output. In the limiting case where the share of the firm in industry output converges to zero and shocks to productivity and factor prices are not perfectly correlated across firms, the cross-sectional variation of $(y_{it} - \bar{p}_t)$ converges to the cross-sectional variation in y_{it} , i.e., the dependent variable in typical firm-level production function regressions is effectively the firm's revenue y_{it} , not the firm's output q_{it} .⁹

One has to be careful with the interpretation of the residual in (4). Note that the error term in (5) combines demand shocks G_{it} and technology shocks A_{it} and, hence, one should not attribute large residuals to high technology because a large residual can stem from a large demand shock. Likewise, large variation of u_{it} in (4) should not be interpreted as large variation in technology.

Nonetheless, u_{it} is an extremely interesting object from the economic standpoint. The entry/exit decisions of firms depend on both technology and demand conditions and u_{it} conveniently summarizes this information about profitability or revenue generating ability of firms (see e.g. Foster, Haltiwanger, and Syverson 2005). Specifically, firms with large u_{it} are more likely to survive than firms with low u_{it} . Thus, even when one does not measure technology a_{it} , the measured u_{it} is still a very useful statistic.

2.5 DISCUSSION

This section makes several points. First, because of data limitations, typical production function regressions based on firm level data use revenue as the dependent variable and, hence, estimate RTS in the revenue function and do not yield the Solow residual measuring technical efficiency of firms. However, residual from the revenue function regression contains important economic information about firms' viability. Likewise, an estimate of η can inform us about important

⁹ See Klette and Griliches (1996) for further discussion. Also note that time dummies are often included in (4) so that deflation by \bar{p}_t is irrelevant.

economic parameters such as cost elasticity ϕ and under certain conditions RTS in production γ and markup μ . Second, profit share in revenue s_π should be used as a robust nonparametric diagnostic for the estimates of RTS in the revenue function. Third, increasing or decreasing RTS in the revenue function and a small profit share s_π can be reconciled by ϕ , the elasticity of the cost with respect to inputs. Hence, the parameter ϕ is of central importance. Fourth, there is sizable variation in factor prices across firms.

Unfortunately, available estimators either do not yield an estimate of ϕ or hinge critically on the assumption that $\phi=1$ (see Section 4). To address this problem, I develop a full-information estimator that deals with the production and cost sides simultaneously.

3 COVARIANCE ESTIMATOR

To consistently estimate parameters of production (revenue) and cost functions, I suggest an estimator based on explicit specification and modeling of unobserved shocks (i.e., productivity, demand, wages, etc.) where factor price shocks are treated symmetrically with productivity shocks, instead of just focusing on productivity shocks. The idea of the estimator is to identify and estimate parameters of the model by matching the covariance matrix implied by the model to the empirical covariance matrix of observed choices of firms.¹⁰ In contrast to single equation estimators (e.g., OLS), this structural estimator models outputs and inputs simultaneously (system approach) by deriving optimal output and factor demands from a profit maximization or cost minimization problem. In this section I explain the intuition behind the estimator, which I call the covariance estimator, and discuss identification and estimation.

3.1 INTUITION

To illustrate the workings and intuition of the estimator, consider model (1)-(2) and assume—for reasons discussed later—that $\phi=1$ and a_{it} and w_{it} have variances σ_a^2 and σ_w^2 with $\rho(a_{it}, w_{it})=0$. These assumptions are restrictive and later I will show that the estimator works under less stringent conditions.

Because a_{it} and w_{it} are not observed, one cannot run a regression of z_{it} and/or y_{it} on these shocks to estimate the RTS in the revenue function η . Note, however, that

¹⁰ This approach is also called structural equation modeling, MIMIC, LISREL and other names (see Bollen 1989 for a general discussion).

$\text{var}(z_{it}) = (\eta - 1)^{-2}(\sigma_w^2 + \sigma_a^2)$, $\text{var}(y_{it}) = (\eta - 1)^{-2}(\eta^2\sigma_w^2 + \sigma_a^2)$, and $\text{cov}(y_{it}, z_{it}) = (\eta - 1)^{-2}(\eta\sigma_w^2 + \sigma_a^2)$ with unknowns $\eta, \sigma_a^2, \sigma_w^2$. One can solve this system of equations for η :

$$\eta = \frac{\text{var}(y_{it}) - \text{cov}(y_{it}, z_{it})}{\text{cov}(y_{it}, z_{it}) - \text{var}(z_{it})}. \quad (6)$$

Thus, one can estimate η from the observed second moments of the data. This was the insight of the seminal paper by Marschak and Andrews (1944). I will call (6) and expressions analogous to (6) the covariance (COV) estimator. Why is the estimator working?

Equations (1)-(2) describe the optimal profit-maximizing behavior of firms and optimization imposes restrictions on how firms respond to shocks. Specifically, the assumption of Hicks-neutral technology and perfectly elastic factor supply curve result in the restriction that revenue and input demand respond equally strongly to an innovation in technology. In other words, the coefficient on the structural shock a_{it} is the same in equations (1) and (2). Furthermore, the assumption of the perfectly elastic factor supply curve implies the restriction that the response of revenue to a shock in the factor price w_{it} is η times stronger than the response of the factor demand z_{it} to the factor price shock. Put differently, the coefficient on w_{it} in equation (2) is equal to the coefficient on w_{it} in (1) multiplied by η . The economic restrictions of Hicks-neutral technology and perfect elasticity of the factor supply are complemented with the technical restriction $\rho(a_{it}, w_{it}) = 0$. This latter condition ensures that one can separate technology shocks and factor price shocks. If technology and factor prices are correlated, this simple model is not identified.

This estimator can have an instrumental variables interpretation. Equation (6) can be equivalently written as

$$\eta = \frac{\text{cov}(y_{it}, y_{it} - z_{it})}{\text{cov}(z_{it}, y_{it} - z_{it})} \quad (7)$$

and, hence, $y_{it} - z_{it}$ is an instrumental variable for z_{it} . Because of the Hicks-neutral technology and perfectly elastic factor supply, profit maximization imposes that revenue y_{it} and input z_{it} respond equally strongly to productivity shocks a_{it} and, hence, $y_{it} - z_{it} \propto w_{it}$. Given the assumption $\rho(a_{it}, w_{it}) = 0$, $y_{it} - z_{it}$ is correlated with z_{it} and uncorrelated with a_{it} . In this simple case, covariance and instrumental variable estimators are equivalent. However, as I will discuss below, explicit instrumental variables like $y_{it} - z_{it}$ are not always available and typically the instrument depends on an unknown parameter.

3.2 MODEL FRAMEWORK

The basic model (3) can be generalized along several dimensions. First, I specify the dynamics of unobserved technology and factor prices collected in the vector F_{it} . Second, measurement errors are salient in micro-level data sets. To address this important fact, I augment (3) with measurement errors. In summary, the general model is

$$X_{it} = \Lambda F_{it} + \bar{X}_i + \varepsilon_{it}, \quad (8)$$

$$F_{i,t+1} = \Pi F_{it} + v_{it}, \quad (9)$$

where X_{it} is the vector of n observed variables (inputs and revenue), F_{it} is the vector of m unobserved variables (factor prices, productivity), the matrix Λ summarizes the responses of observed variable to F_{it} , \bar{X}_i is a vector of unobserved permanent firm-specific effects for X_{it} , ε_{it} is a vector of i.i.d. zero-mean measurement or expectations errors, v_{it} is a vector of i.i.d. structural zero-mean innovations to F_{it} , and the matrix Π captures the dynamics of F_{it} .¹¹ The matrix Λ for the n -input case is given in equation (39), Appendix A. I collect parameters of the model in the vector θ and assume here and henceforth that the mapping from θ to $\Pi, \Lambda, \Omega \equiv E(v_{it}v'_{it})$ and $\Psi \equiv E(\varepsilon_{it}\varepsilon'_{it})$ is one-to-one in the admissible domain of θ .

This state space representation of the problem nests many important cases such as dynamic factor models ($m < n$), log-linearized rational expectations models in state-space form and serially correlated measurement errors.¹² I do not take a stand on time series properties of F_{it} and the contemporaneous correlation of innovations in v_{it} as economic theory may have few restrictions on how variables in F_{it} evolve over time or how v_{it} is correlated. Note that variables in F_{it} can be correlated because either Π or Ω is not diagonal. Likewise, I do not impose any structure on \bar{X}_i .

¹¹ Since dependence of X_{it} on observed exogenous variables (e.g., time dummies) can be easily eliminated by projection methods, I abstract from such dependence without loss of generality.

¹² The latter is important in practice because econometricians rarely have reliable estimates of capital stock, effort, etc. For example, there are two popular estimates of capital: 1) real capital stock computed by inventory methods; 2) book value of fixed assets. In either case, measurement error is likely to be serially correlated. Suppose that the econometrician uses a noisy measure of investment such that e_t , the error in true investment I_t^* , is classical (the measurement error can arise from using an investment price index to deflate firm-level investment expenditures). The true capital stock evolves according to $K_t^* = (1 - \delta)K_{t-1}^* + I_t^*$. Then the estimated capital stock is $K_t = (1 - \delta)K_{t-1} + I_t = K_t^* + \sum_{s=0}^{\infty} (1 - \delta)^s e_{t-s} = K_t^* + \varepsilon_t$ with $\varepsilon_t = (1 - \delta)\varepsilon_{t-1} + e_t$, that is, measurement error $\varepsilon_t \sim AR(1)$. Importantly, serially correlated measurement errors invalidate instrumental variables based on leads/lags of inputs/outputs or input mixes. Similarly, true labor input may be measured with serially correlated error because of labor hoarding.

The model (8)-(9) has much in common with dynamic factor models. However, in contrast to dynamic factor models, the factor loadings embodied in the matrix Λ can be identified under certain conditions and, thus, factors can have a structural interpretation. In the next section, I present the conditions under which θ is identified.

3.3 IDENTIFICATION

The key question for the COV estimator is the identification of parameters because many models can be consistent with observed covariances. Local identification of these parameters in the static model (8) and dynamic model (8)-(9) is discussed elsewhere (e.g., Hoch 1958, Maravall and Aigner 1977, Maravall 1979, Bollen 1989, Bekker, Merkens and Wansbeek 1994). In effect, local identification requires showing that the appropriate Jacobian has full rank. Obviously, the necessary condition for identification is that the number of parameters in θ is not greater than the number of unique moments in the considered covariance and autocovariance matrices.

Global identification is more subtle. In factor analysis terminology, global identification reduces to proving that there is no rotation matrix T producing $\{\tilde{\Lambda}, \tilde{\Pi}, \tilde{\Omega}, \tilde{\Psi}\} = \{\Lambda T, T^{-1}\Pi T, \Psi, T^{-1}\Omega T^{-1}\}$ that satisfies the restrictions imposed on $\{\Lambda, \Pi, \Omega, \Psi\}$ (see Theorem 5 in Tse and Anton 1972). Profit maximization imposes many restrictions on the matrix Λ and, thus, on admissible rotation matrices T . Yet, these restrictions do not eliminate rotational equivalence in (8)-(9). Further restrictions on Ω and Π can guarantee identification. The following proposition proves global identification for two important special cases.

Proposition 2

Assume that

- i) the matrix Π is invertible,*
- ii) the eigenvalues of Π are in the unit circle,*
- iii) the system in (8)-(9) is observable and controllable,*
- iv) $E(\boldsymbol{\varepsilon}_{it}) = E(\mathbf{v}_{it}) = 0$ and $E(\mathbf{v}_{it}\mathbf{v}'_{jt}) = E(\mathbf{v}_{it}\mathbf{v}'_{is}) = E(\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{jp}) = E(\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{is}) = E(\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{jt}) = 0$ for any t, i, p, j and $s \neq t$,*
- v) firms maximize profits so that the matrix of loadings Λ is as in (39),*
- vi) at least one of the factors is supplied in a competitive market.*

Then the model (8)-(9) is uniquely globally identified if

- a) innovations in \mathbf{v}_{it} are not correlated (i.e., the covariance matrix Ω is diagonal), or*

b) the matrix Π is diagonal (i.e., there are no dynamic cross-variable responses in F_{ii})

Proof: see Appendix B.

The first three assumptions are technical: the system does not have redundant (linearly dependent) shocks; unobservables are stationary; and one can back out the behavior of the system if one can control unobservables. These assumptions are easily satisfied for relevant economic models. The assumption that one of the factors is supplied in a perfectly competitive market fixes the elasticity of the factor supply curve for other inputs which, in turn, fixes the parameters of the revenue function. Note that factor price and productivity can be correlated in both (a) or (b). Identification is achieved by imposing restrictions on the correlation of innovations in factor prices and technology (Ω is diagonal) or by imposing restrictions on the propagation of the shocks to technology and factor prices (Π is diagonal). It is also possible to identify θ if combinations of restrictions on Π or Ω are available.¹³

Local identification of models with serially correlated measurement error is discussed in Maravall (1979) and Maravall and Aigner (1977). In the next proposition, I present conditions under which the model is globally identified.

Proposition 3

Suppose that (i) serially correlated measurement errors in observed inputs and outputs are not correlated across inputs and outputs at all leads and lags; (ii) measurement errors are not correlated with factor prices and productivity and the number of serially correlated measurement errors k cannot exceed the number of observed variables n ; (iii) serially correlated measurement errors are AR(1) and covariance stationary. Then Λ , Π , and Ω identified almost everywhere if and only if Λ , Π , and Ω are identified in the absence of measurement errors.

Proof: see Appendix B.

Note that in Proposition 2 and Proposition 3 I use only time series variation in factor prices and technology to identify parameters of the model. In other words, I do not use variation in \bar{X}_i . However, it is possible to use restrictions on the distribution of \bar{X}_i to achieve identification in otherwise underidentified models. For example, one may be willing to impose

¹³ Glover and Willems (1974) show that one needs to modify the conditions slightly if observed and latent variables can respond contemporaneously for the same set of shocks.

$\bar{X}_i = \Lambda \bar{F}_i$ with $Var(\bar{F}_i)$ being diagonal. Such restrictions can be particularly important if between variation is large relative to within variation.¹⁴

3.4 ESTIMATION AND INFERENCE

Since I do not model entry and exit decisions, I assume that the panel of the firms is balanced with $t=0, \dots, T$ observations for each cross-section.¹⁵ The number of i.i.d. cross-sections is N . I collect the parameters of interest in the vector θ , which I assume to be locally identified. I assume that X_{it} is stationary. The estimation strategy is to find θ that minimizes the distance between the appropriate sample covariance matrix and the covariance structure implied by θ . There are many possible ways to construct a metric of discrepancy between the sample and implied covariance matrices. I focus on maximum-likelihood methods since they tend to have somewhat better performance in finite samples because MLE does not use a weighting matrix that depends on unknown parameters (e.g., Clark 1996).

It is convenient for further derivations to stack observed choices for each firm in vector $X_i = [X'_{i0} \ X'_{i1} \ \dots \ X'_{iT}]'$ where $X_{it} = \bar{X}_i + \Lambda F_{it} + \varepsilon_{it}$ and $F_{it} = \Pi F_{i,t-1} + v_{it}$. Suppose that $v_i = [v'_{0i} \ v'_{1i} \ \dots \ v'_{Ti}]' \sim N(0, \Omega \otimes I_T)$ and $\varepsilon_i = [\varepsilon'_{0i} \ \varepsilon'_{1i} \ \dots \ \varepsilon'_{iT}]' \sim N(0, \Psi \otimes I_T)$ (i.e., measurement error ε_{it} and structural shocks v_{it} are normally distributed and serially uncorrelated) and $E(v_i \varepsilon_i') = 0$ (i.e., structural shocks and measurement errors are not correlated at all leads and lags). Provided $\bar{X}_i = 0$, one can find that $X_i \sim N(0, \Phi_T)$ where

$$\Phi_T \equiv E(X_i X_i') = \begin{bmatrix} \Sigma_0 & & & \\ \Sigma_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ \Sigma_T & \dots & \Sigma_1 & \Sigma_0 \end{bmatrix} = \begin{bmatrix} \Lambda \Gamma_0 \Lambda' + \Psi & & & \\ \Lambda \Gamma_0 \Pi \Lambda' & \ddots & & \\ \vdots & \ddots & \ddots & \\ \Lambda \Gamma_0 \Pi^T \Lambda' & \dots & \Lambda \Gamma_0 \Pi \Lambda' & \Lambda \Gamma_0 \Lambda' + \Psi \end{bmatrix}, \quad (10)$$

with $\Sigma_s = E(X_{it} X'_{i,t-s})$, $\Gamma_s = E(F_{it} F'_{i,t-s})$, $\Gamma_0 : \Gamma_0 = \Pi \Gamma_0 \Pi' + \Omega$. To simplify the notation, I use Φ instead of $\Phi_T(\theta)$, which explicitly indicates that Φ_T is a function of parameters collected in the vector θ . Hence, the likelihood function is given by

¹⁴ In applications, it may happen that η , returns to scale in the revenue function, is identified while other parameters in θ are not. In such cases, one can impose fairly arbitrary restrictions on unidentified parameters to have a well-defined estimation problem without affecting the identification of η (see Bollen 1989 for a discussion). If η is identified locally but not globally, it may be possible to rule out implausible cases, e.g., $\eta < 0$. If η is not locally identified, one can follow Marschak and Andrews (1944) and put economic bounds on possible values of η . This amounts to constructing the set of values that parameters can take for all admissible rotations.

¹⁵ One may use weighting techniques similar in spirit of Olley and Pakes (1995) to control for entry/exit decisions.

$$\sum_{i=1}^N l(\mathbf{X}_i, \theta) = \ln |\Phi_T| + \text{trace}\{\hat{\Phi}_T \Phi_T^{-1}\} - \ln |\hat{\Phi}_T| - Tn \quad (11)$$

where $\hat{\Phi}_T = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i'$, n is the number of observed choices of firms and the maximum likelihood estimate of θ maximizes (11). Since rational expectations models can be represented in a state-space form like (8)-(9), it is an easy step to extend (11) to estimation of rational expectations models (see Appendix A).

For the case where steady state levels of inputs and output are treated as random, suppose that $\bar{\mathbf{X}}_i \sim N(0, \Xi)$ and $E(\bar{\mathbf{X}}_i \mathbf{u}_i') = 0$ and observe that $\tilde{\Phi}_T = E(\mathbf{X}_i \mathbf{X}_i') = (\Xi \otimes J_T J_T') + \Phi_T$, where J_T is the $(T+1) \times 1$ vector of ones. It is straightforward to find that the associated likelihood satisfies $\sum_{i=1}^N l(\mathbf{X}_i, \theta) \propto -\ln |\tilde{\Phi}_T| - \text{trace}\{\hat{\Phi}_T \tilde{\Phi}_T^{-1}\}$. If $\bar{\mathbf{X}}_i$ is treated as a fixed parameter, one can transform the data to eliminate the incidental parameters $\bar{\mathbf{X}}_i$, e.g., apply first differencing as in Hsiao, Pesaran, and Tahmiscioglu (2002). The log-likelihood for first-differenced \mathbf{X}_i satisfies: $\sum_{i=1}^N l(D\mathbf{X}_i, \theta) \propto -\ln |D\Phi_T D'| - \text{trace}\{(D\hat{\Phi}_T D')(D\Phi_T D')^{-1}\}$ where D is the $nT \times n(T+1)$ first-difference matrix. Alternatively, one can use a conditional likelihood approach, which under certain conditions is equivalent to applying a transformation (e.g., Arellano 2003).

Since \mathbf{X}_i is not necessarily normally distributed, one may want to use the standard quasi-maximum likelihood tools to interpret estimates and construct standard errors, i.e.,

$\text{Var}(\hat{\theta}) = N^{-1} H^{-1} G H^{-1}$ where $H = N^{-1} \sum_{i=1}^N \nabla_{\theta}^2 l$ and $G = N^{-1} \sum_{i=1}^N \nabla_{\theta} l \cdot \nabla_{\theta} l'$.¹⁶ In the course of specification searches, one can use overidentifying restrictions tests since dynamic models such as (8)-(9) are typically overidentified. If the researcher is not satisfied with standard asymptotic inference, he or she can evaluate the distribution of a test statistic using bootstrap procedures (e.g., Horowitz 1998) or rely on a statistic that is robust to non-normality (e.g., Bollen 1989).¹⁷

¹⁶ A popular alternative to (Q)MLE is generalized method of moments (GMM), which does not require normality. GMM minimizes the following objective function $J = N[\hat{\Phi}_T^* - \Phi_T^*(\theta)]' W^{-1} [\hat{\Phi}_T^* - \Phi_T^*(\theta)]$ where $\Phi_T^* \equiv [\text{vech}(\Sigma_0)' \quad \text{vec}(\Sigma_1)' \quad \dots \quad \text{vec}(\Sigma_q)']$, $\hat{\Phi}_T^*$ is a sample estimate of Φ_T^* , W is a weighting matrix of conformable size. GMM and ML are asymptotically equivalent (Anderson and Amemiya 1988). If factor prices and productivity are uncorrelated, GMM and MLE are equivalent to IV estimator with (if necessary, leads or lags of) input ratios as instruments (Schmidt 1988).

¹⁷ Monte-Carlo experiments (not reported here) suggest that finite sample performance of the COV estimator can be improved if a relatively small number of moments (sufficient for identification) are used in estimation. This enhancement is possible because low-order autocovariances can be estimated more precisely than in the presented formulation. For example, the first-order autocovariance can be estimated using NT observations while in the presented formulation only N observations are used for the estimation. This issue is similar to choosing optimal number of moments in GMM application and is left for future research.

3.5 DISCUSSION

The structural approach embodied in the suggested estimator is built on earlier works by Marschak and Andrews (1944) and Schmidt (1988). I extend their static full-information maximum likelihood (FIML) estimators to dynamic settings and improve upon their FIML in several respects. First, I allow factor prices to be correlated with technology. This correlation can arise because of rent/profit sharing, complementarity of worker skills and technology, monopsony power, overtime premia, etc. In contrast, the static models considered in previous studies are not identified if a_{it} and factor prices are correlated. Furthermore, many popular dynamic estimators of returns to scale do not allow correlation between technology and factor prices or innovations in technology and factor prices. Second, my extension permits i.i.d. and serially correlated measurement errors while static FIML is not identified if there is any measurement error. Third, I show that static and dynamic models can be identified and estimated when factor markets are imperfectly competitive, i.e., factor supply curves are not perfectly elastic. Specifically, I show that having at least one input with a perfectly elastic factor supply is sufficient for identification. Furthermore, I show in Appendix A that the covariance estimator can be extended to cases where the profit-maximizing firm faces adjustment costs or production function is constant elasticity of substitution (CES).

There is a cost of using the covariance estimator. Like any other FIML estimator, the COV estimator is more sensitive to misspecification than single-equation methods (e.g., OLS). Since the COV estimator works with higher moments, it may be more sensitive to outliers.

4 ALTERNATIVE ESTIMATORS OF THE RETURNS TO SCALE

In this section I analyze alternative estimators. I start with OLS to highlight the problems of estimating production (revenue) functions and then proceed with the analysis of popular solutions to these problems. To contrast estimators, I use the dynamic model (8)-(9) with observed input z and output (revenue) y , measurement errors $\varepsilon_{it}^z, \varepsilon_{it}^y$, and unobserved factor price w_{it} and technology a_{it} :

$$z_{it} = \frac{1}{\eta-\phi} w_{it} - \frac{1}{\eta-\phi} a_{it} + \varepsilon_{it}^z, \quad (12)$$

$$y_{it} = \frac{\eta}{\eta-\phi} w_{it} - \frac{\phi}{\eta-\phi} a_{it} + \varepsilon_{it}^y, \quad (13)$$

$$w_{i,t+1} = \rho_w w_{it} + v_{it}^w, \quad (14)$$

$$a_{i,t+1} = \rho_a a_{it} + v_{it}^a. \quad (15)$$

To simplify the presentation, I abstract from firm-specific effects. The estimated equation is

$$y_{it} = \eta z_{it} + a_{it} + \varepsilon_{it}^y = \eta z_{it} + error. \quad (16)$$

This model makes exposition clear, yet my conclusions apply to more realistic cases as well.

4.1 OLS

Consider a firm characterized by (12)-(15) and assume that variables are measured without error and a_{it} and w_{it} are uncorrelated i.i.d. zero-mean shocks with $\rho_a = 0$ and $\rho_w = 0$ and variances σ_a^2 and σ_w^2 .¹⁸ Using the structural equations in (12)-(15), I find the probability limit of $\hat{\eta}_{OLS}$ in (16):

$$p \lim \hat{\eta}_{OLS} = \frac{\text{cov}(y_{it}, z_{it})}{\text{var}(z_{it})} = \frac{\phi \sigma_a^2 + \eta \sigma_w^2}{\sigma_a^2 + \sigma_w^2} = \eta + (\phi - \eta) \frac{\sigma_a^2 / \sigma_w^2}{1 + \sigma_a^2 / \sigma_w^2} > \eta.$$

The upward bias in the OLS estimates is “the transmission bias” identified by Marschak and Andrews (1944). The asymptotic bias is decreasing in the variance of factor prices and, if the only source of variation is productivity, the OLS estimate is ϕ , the elasticity of the cost, irrespective of the true η , the elasticity of the revenue function.

How big is the bias? If wage and productivity shocks are uncorrelated, then $bias = \frac{(\phi - \eta) \sigma_a^2}{\sigma_a^2 + \sigma_w^2} < (\phi - \eta) = \phi s_\pi$ because $\phi - \eta = \phi s_\pi$ by Proposition 1. Since the profit share is 3% or less (e.g., Basu and Fernald, 1997) and ϕ is likely to be no greater than 1.5, the bias is positive but likely to be smaller than 0.045. Intuitively, the OLS estimate is between η and ϕ . Because these two quantities are close to each other, there is only a narrow range in which the OLS estimate can fall. Even if wage and productivity shocks are correlated, the asymptotic bias is likely to be small.¹⁹ The same conclusion is likely to hold for cases with multiple inputs.²⁰ A

¹⁸ If firm-specific productivity is time invariant, then one can use panel data techniques to control productivity with firm-specific fixed effects (FE). This happy situation is not universally applicable and FE is not consistent if productivity is time varying. Furthermore, as Griliches and Mairesse (1995) observe, FE aggravates other problems (e.g., attenuation bias of measurement errors) precisely because of assuming time invariant differences in productivity across firms.

¹⁹ If wage and productivity shocks are correlated, the asymptotic bias is $(\phi - \eta)(\sigma_a^2 - \rho \sigma_a \sigma_w) / (\sigma_a^2 + \sigma_w^2 - 2\rho \sigma_a \sigma_w)$ where $\rho \equiv \rho(a_{it}, w_{it})$. The OLS estimate of η can exceed ϕ if and only if $-\rho > \sigma_w / \sigma_a$ or fall below η if and only if $\rho > \sigma_a / \sigma_w$. The first case requires a negative correlation between productivity and factor price, which is a possible but a

relatively small bias in RTS, however, does not imply a small bias in the OLS estimate of the coefficient for a given input. Put differently, an upward bias in one of the coefficients is offset by a downward bias in other coefficients. This result, however, can be distorted by measurement errors.

It is also instructive to consider the correlation between true productivity shocks and measured productivity shocks. Using (12)-(13), one can show that the measured productivity shock is

$$\hat{a}_{it} = y_{it} - \hat{\eta}z_{it} = a_{it} + (\eta - \phi) \frac{\sigma_a^2 / \sigma_w^2}{1 + \sigma_a^2 / \sigma_w^2} \left(\frac{1}{\eta - \phi} w_{it} - \frac{1}{\eta - \phi} a_{it} \right) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_a^2} a_{it} + \frac{\sigma_a^2}{\sigma_w^2 + \sigma_a^2} w_{it}$$

and therefore

$$\begin{aligned} \text{var}(\hat{a}_{it}) &= \sigma_w^2 \sigma_a^2 / (\sigma_w^2 + \sigma_a^2) < \text{var}(a_{it}), \\ \text{cov}(a_{it}, \hat{a}_{it}) &= \sigma_a^2 \sigma_w^2 / (\sigma_w^2 + \sigma_a^2) < \text{var}(a_{it}), \\ \rho(a_{it}, \hat{a}_{it}) &= \frac{1}{\sqrt{1 + (\sigma_a / \sigma_w)^2}} < 1. \end{aligned}$$

Apparently, the measured productivity is necessarily less volatile than and imperfectly correlated with the true productivity. Note that, for example, correlation does not depend on $(\phi - \eta)$ as long as it is non-zero because productivity and factor price shocks are amplified in the same proportion and comovement between measured and true shocks depends only on the ratio of volatilities of productivity and factor price shocks.

4.2 IV/GMM ESTIMATORS

The transmission bias can be eliminated if the researcher has an instrumental variable (IV) explaining variation in z_{it} unrelated to productivity shocks a_{it} . In the simple setup of uncorrelated w_{it} and a_{it} , the best instrument is w_{it} , the price of z_{it} . The problem is that factor prices w_{it} are almost never collected and therefore such an IV is infeasible in the vast majority of cases. To

less likely case. The second case is more plausible but it still requires that productivity be less volatile than base wage. Bartelsman and Doms (2000) report that the ratio of the ninth decile of the distribution of a_{it} across firms to the first decile is typically about two to three. Juhn, Murphy and Pierce (1993) report that the log wage differential between the 90th and 10th percentiles in the private sector is about 1.5 and the differential is about 1.1 after controlling for observed labor force characteristics. Hence, σ_a / σ_w and σ_w / σ_a are likely to be large. If the correlation between a_{it} and w_{it} is in the range $(-\sigma_w / \sigma_a, \sigma_a / \sigma_w)$, the bound presented above is still appropriate, i.e., the bias is likely to be less than 0.045.

²⁰ The expression for the bias becomes complicated with multiple inputs and the upper bound for the bias depends on the elasticities of cost for specific inputs and relative variability of factor prices.

rectify this problem, Schmidt (1988) suggests using input/output ratios as instruments, e.g., $y_{it} - z_{it}$ in (7). If the production function is Cobb-Douglas, then Schmidt's IV (SIV) is identical to the IV estimator with factor prices as instruments. However, SIV is not consistent if factor prices and productivity are correlated, supply curve for at least one of the factors is not perfectly elastic (i.e., $\phi \neq 1$), or if either the output (revenue) or inputs are measured with a serially correlated error. All of these cases are empirically important and, hence, although SIV can help in certain circumstance, its assumptions may be too restrictive.

Alternatively, Blundell and Bond (1998, 1999, henceforth BB) suggest using (i) transformations of the variables to eliminate a_{it} from (16) and (ii) lags of inputs and outputs as instruments. Specifically, BB suggest two types of moment conditions: levels and differences.²¹ Define $\mathcal{G}_{it} \equiv y_{it} - \rho y_{i,t-1} - \eta z_{it} + \rho \eta z_{i,t-1}$, the serially uncorrelated residual from the quasi-differenced production (revenue) function (16). The differences moment condition is $E(\Delta \mathcal{G}_{it} s_{it}) = 0$, where s_{it} is any combination of $y_{i,t-j}, z_{i,t-j}, j \geq 3$. The levels moment condition is $E(\mathcal{G}_{it} s_{it}) = 0$, where s_{it} is any combination of $\Delta y_{i,t-j}, \Delta z_{i,t-j}, j \geq 2$. Since technology is an AR(1) process by assumption, the two sets of moment conditions are valid. Note that, in contrast to SIV, \mathcal{G}_{it} can be correlated with serially uncorrelated innovations in factor prices. Two options for estimation are available. First, estimate the unrestricted model (i.e., let $\mathcal{G}_{it} \equiv y_{it} - b_1 y_{i,t-1} - b_2 z_{it} - b_3 z_{i,t-1}$ with b_1, b_2, b_3 being free parameters) and take the coefficient on z_{it} as $\hat{\eta}$. Second, estimate the restricted model.

The following result can be proven for the restricted specification:

Proposition 4

Consider profit-maximizing firms as in (8)-(9) and estimate the production (revenue) function using the restricted specification of the BB estimator. Then the model is not globally identified. In particular, the model has multiple locally-identified solutions, provided that the matrix Π has distinct eigenvalues. The number of solutions is no greater than $n+1$ where n is the number of inputs. If the matrix Π has repeated eigenvalues, then the model is not identified.

Proof: see Appendix B.

²¹ Since Arellano and Bond (1991) consider a subset of moments in Blundell-Bond estimator the subsequent discussion applies to Arellano-Bond estimator as well.

To get the intuition behind this result, consider, without loss of generality, the “levels” moments $E\{(y_{it} - \rho y_{i,t-1} - \eta z_{it} + \rho \eta z_{i,t-1})s_{it}\} = 0$ where s_{it} is a subset of $\Delta y_{i,t-2}, \Delta y_{i,t-3}, \dots, \Delta z_{i,t-2}, \Delta z_{i,t-3}, \dots$. Use (12)-(13) to eliminate y_{it} and z_{it} from the moment condition and observe that two sets of parameter values satisfy the moment condition:

$$\text{Solution \#1: } \hat{\rho} = \rho_a, \hat{\eta} = \eta \text{ which yields } (y_{it} - \hat{\rho} y_{i,t-1} - \hat{\eta} z_{it} + \hat{\rho} \hat{\eta} z_{i,t-1}) = v_{it}^a,$$

$$\text{Solution \#2: } \hat{\rho} = \rho_w, \hat{\eta} = 1 \text{ which yields } (y_{it} - \hat{\rho} y_{i,t-1} - \hat{\eta} z_{it} + \hat{\rho} \hat{\eta} z_{i,t-1}) = v_{it}^w.$$

In this simple case technology is Hicks-neutral and factor supply is perfectly elastic. Under these assumptions, profit maximization imposes that $y_{it} - z_{it} \propto w_{it}$ and $y_{it} - \eta z_{it} \propto a_{it}$. After appropriate quasi-differencing, each of these expressions is proportional to a serially uncorrelated shock. Thus, the objective function of the estimator in this simple case has two local minima.

In principle, the standard prescription is to choose a solution that gives the global minimum of some objective function (e.g., residual sum of squares), yet this heuristic may choose the incorrect solution #2. It may be hard to rule out some of the solutions on *economic* grounds. For example, in the presented one-input/one-output case, both solutions can be appealing. Furthermore, since the empirically observed profit share is small, η is likely to be close to unity (given perfect competition in factor markets) and, hence, the estimator may be poorly identified even locally.

The consequences of having multiple solutions become particularly acute in the unrestricted specification since it is possible to take linear combinations of solutions such as above so that the model is not identified locally. The following proposition shows this formally.

Proposition 5

Consider profit-maximizing firms as in (8)-(9) or in a modification of (8)-(9) that allows for a contemporaneous response of observed variables to innovations in F_{it} . Then in the unrestricted specification, the Jacobian of the moment conditions (either in levels or differences or both) based on lags of inputs or revenue or their differences does not have full rank.

Proof: see Appendix B.

This proposition demonstrates that the rank of the Jacobian for the moment conditions is smaller than the number of parameters to be estimated in the unrestricted specification and, hence, the model is not identified. Note that the problem is not in the weak correlation of lags of variables with their current values (which is the point addressed by using level moment conditions). The reduced rank problem arises because profit maximization imposes restrictions on how inputs and outputs comove over time so that some moments are collinear. In other words, one cannot treat choices of firms as independent. Optimization not only ensures the simultaneity in the choice of inputs and output but also it puts a precise structure on how the outcomes are related to each other. Such relationship in the context of profit-maximizing firms is so tight that observed choices (moment conditions) are linearly dependent. Because of the weak identification, the estimator is likely to have a flat density. Furthermore, I show in Appendix A that, to a first order approximation, BB can be poorly identified even when it is costly to adjust inputs. It is critical to use the restricted specification to attenuate the problem of weak identification. Mavroeidis (2004) notes a similar problem with using lags as instruments in estimation of rational-expectations macroeconomic models.

Note that Proposition 4 and Proposition 5 do not show poor identification of system GMM rather they show that poor identification can be a serious problem when the estimator is applied to estimating returns to scale for optimizing firms.²²

4.3 *INVERSION (CONTROL FUNCTION) ESTIMATORS*

In this section I consider inversion estimators that use demands for inputs, investment or other observable choices of firms to construct a proxy for firm's productivity and condition inputs in the production (revenue) function on the proxy. A typical regression in this control-function approach is

$$y_{it} = \eta z_{it} + \lambda \tilde{a}_{it} + \mathcal{G}_{it} = \eta z_{it} + \lambda \tilde{a}_{it} + error, \quad (17)$$

where \tilde{a} is the proxy for the productivity of a firm. The critical assumption of these estimators is that the mapping (inversion function) from observed characteristics to productivity (or its proxy)

²² The BB estimator can be identified from nonlinearities in decision rules captured by second-order effects. In addition, one may expect a better performance of the BB estimator if shocks to factor prices have higher orders of correlation than shocks to productivity. For example, factor prices with AR(2) structure are sufficient to guarantee identification of the BB estimator if productivity is AR(1). However, if the roots (other than the largest root) of the lag polynomial for factor prices are small, the BB moments can be almost collinear in finite samples and the estimator can behave erratically. Furthermore, there is no a priori reason to believe that wage shocks have a higher order of autocorrelation than productivity shocks. Likewise, identification from second-order effects can be fragile. BB can be identified if it is costly to adjust *all* inputs.

is non-stochastic. I focus on the Levinsohn-Petrin (2003, henceforth LP) estimator but my conclusions are also relevant to similar estimators (e.g., Olley and Pakes 1996, Pavcnik 2002).

Following LP, consider the Cobb-Douglas revenue function with capital, labor and material inputs, that is, $Y_{it} = \exp(a_{it})K_{it}^{\beta_K}L_{it}^{\beta_L}M_{it}^{\beta_M}$ where the productivity shock a_{it} is an AR(1) process: $a_{it} = \rho_a a_{i,t-1} + v_{it}^a$ and $v_{it}^a \sim iid(0, \sigma_{va}^2)$. In the notation of LP, $\omega_{it} = a_{it}$ and $\xi_{it} = v_{it}^a$ and, for convenience, define $\tau_{it} = E(\omega_{it} | \omega_{i,t-1}) = \rho_a a_{i,t-1}$. Capital is chosen in the beginning of period t when ξ_{it} is not observed but τ_{it} and factor prices are observed. Labor and materials are chosen when ξ_{it} is known, that is, variable inputs can be adjusted when more information is available. I denote (log) factor prices for capital, labor and materials with r_{it} , w_{it} , and p_{it}^M . All factors are supplied in perfectly competitive markets. There is no measurement error. The rest of the problem is unchanged and the estimated production (revenue) function is

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + \beta_M m_{it} + \omega_{it}, \quad (18)$$

where $\eta = \beta_K + \beta_L + \beta_M$ is RTS in the revenue function.

The idea of the LP estimator is to invert demands for capital and materials to infer productivity shocks ω_{it} and then use the estimated productivity shock as a regressor in the production (revenue) function—that is, condition (18) on ω_{it} . The problem, however, is in the poor quality of the estimates of the productivity shocks.

Note from profit maximization that the observed variables k_{it} , l_{it} , m_{it} and y_{it} can be expressed in terms of unobserved variables r_{it} , w_{it} , p_{it}^M , τ_{it} , and ξ_{it} :

$$(\eta - 1)k_{it} = (1 + \beta_K - \eta)r_{it} + \beta_L w_{it} + \beta_M p_{it}^M - \tau_{it}, \quad (19)$$

$$(\eta - 1)l_{it} = \beta_K r_{it} + (1 + \beta_L - \eta)w_{it} + \beta_M p_{it}^M - \tau_{it} - (\eta - 1)(\eta - 1 - \beta_K)^{-1} \xi_{it}, \quad (20)$$

$$(\eta - 1)m_{it} = \beta_K r_{it} + \beta_M w_{it} + (1 + \beta_M - \eta)p_{it}^M - \tau_{it} - (\eta - 1)(\eta - 1 - \beta_K)^{-1} \xi_{it}, \quad (21)$$

$$(\eta - 1)y_{it} = \beta_K r_{it} + \beta_L w_{it} + \beta_M p_{it}^M - \tau_{it} - (\eta - 1)(\eta - 1 - \beta_K)^{-1} \xi_{it}. \quad (22)$$

It is straightforward to invert factor demands to firm's productivity $\omega_{it} = \tau_{it} + \xi_{it}$:

$$\tau_{it} = (1 - \eta)k_{it} + (1 + \beta_K - \eta)r_{it} + \beta_L w_{it} + \beta_M p_{it}^M, \quad (23)$$

$$\xi_{it} = -(1 + \beta_K - \eta)(k_{it} - m_{it} + r_{it} - p_{it}^M). \quad (24)$$

There is a one-to-one non-stochastic mapping between factor demands and productivity shocks if and only if factor prices are the same across firms. But if factor prices are the same

across firms then labor and materials are collinear. To see this point, suppose that factor prices are the same across firms in any given period t . Because inversion of factor demands is indexed by time, one can conveniently set $r_{it}=r_t=0$, $w_{it}=w_t=0$, and $p_{it}^M = p_t^M = 0$. Clearly, this leads to $m_{it} = l_{it} = \frac{-1}{\eta-1} \tau_{it} + \frac{-1}{\eta-1-\beta_K} \xi_{it}$ and, thus, m_{it} and l_{it} are collinear. Even if the responses of l_{it} and m_{it} to τ_{it} and ξ_{it} are different (e.g., factor supply curves for labor and materials have different slopes), there is no unexplained variation in l_{it} after it is conditioned on m_{it} and k_{it} : $l_{it} - E(l_{it} | k_{it}, m_{it}) = l_{it} - E(l_{it} | \tau_{it}, \xi_{it}) = l_{it} - l_{it} = 0$. Put differently, once (18) is conditioned on ω_{it} there is no variation in labor/materials ratio and coefficients β_L, β_M are not identified. This point is raised by Basu (1999) and further discussed in Akerberg and Caves (2003) and Bond and Soderbom (2005).²³

On the other hand, if factor prices are not the same across firms, then the assumption of a non-stochastic inversion function is violated. Therefore, inversion of factor demands and conditioning (18) on estimated productivity shocks are internally inconsistent. In applications, identification of LP must come from misspecification of the model.

What happens if volatility in factor prices is ignored? After all, LP indeed moves the estimates in the direction predicted by the theory, e.g., $\beta_L^{OLS} > \beta_L^{LP}$ and $\beta_K^{OLS} < \beta_K^{LP}$. To understand why LP can improve upon OLS, observe that the control function that combines k_{it} , m_{it} and ω_{it} in (18) is

$$\zeta(k_{it}, m_{it}) = \beta_K k_{it} + \beta_M m_{it} + \omega_{it} = (1 + \beta_M + \beta_K - \eta)(m_{it} + p_{it}^M) + \beta_L w_{it}, \quad (25)$$

which is correlated with prices w_{it} and p_{it}^M . What are the consequences? Consider a simple case of one input with perfectly elastic supply and $\tilde{a}_{it} = a_{it} + \chi w_{it}$ as a control function in (17) where $\rho(\tilde{a}_{it}, w_{it}) \neq 0$ if $\chi \neq 0$. Using projection methods, one can find

$$p \lim \hat{\eta}'_{OLS} = \frac{\text{cov}(y_{it}, z_{it}) \text{var}(\tilde{a}_{it}) - \text{cov}(y_{it}, \tilde{a}_{it}) \text{cov}(z_{it}, \tilde{a}_{it})}{\text{var}(\tilde{a}_{it}) \text{var}(z_{it}) - \text{cov}(z_{it}, \tilde{a}_{it})^2} = \eta + (1 - \eta) \frac{\chi}{1 + \chi}.$$

Clearly, the estimated coefficient is inconsistent unless $\chi = 0$. From (25), it is likely that χ is positive and $\hat{\eta}'_{OLS}$ is biased upward. The performance of LP depends critically on the parameter χ . Specifically, as χ increases, the bias in LP estimate increases. However, for small χ , LP is

²³ Akerberg and Caves (2003) propose alternative timing assumptions to remove this identification problem. Bond and Soderbom (2005) propose to utilize non-linearities and to introduce shocks to adjustment costs to identify parameters.

likely to have large standard errors since variation of z_{it} condition on a_{it} is small. (To reiterate, if $\chi = 0$, LP is not identified.) Hence, for reasonably small χ , LP can have a smaller bias than OLS does.²⁴ Using nonparametric techniques or polynomials does not resolve the misspecification in (25) and the subsequent identification problem in (18) because identification of LP does not depend on the functional forms. Furthermore, this exercise shows that LP cannot have technology shocks correlated with factor prices.

Measurement errors present another problem in the inversion estimators because the assumption of non-stochastic inversion of observable choices does not hold and upward biases are likely to arise. More generally, conditioning on a proxy variable contaminated with measurement error leads to inconsistent estimates. To get intuition, consider a simple case of one input and $\tilde{a}_{it} = a_{it} + \zeta_{it}$ as a control function in (17), where $\zeta_{it} \sim iid(0, \sigma_\zeta^2)$ is a classical measurement error. It follows that

$$p \lim \hat{\eta}_{OLS}'' = \frac{\text{cov}(y_{it}, z_{it}) \text{var}(\tilde{a}_{it}) - \text{cov}(y_{it}, \tilde{a}_{it}) \text{cov}(z_{it}, \tilde{a}_{it})}{\text{var}(\tilde{a}_{it}) \text{var}(z_{it}) - \text{cov}(z_{it}, \tilde{a}_{it})^2} = \eta + (1 - \eta) \frac{\sigma_a^2 \sigma_\zeta^2}{\sigma_a^2 \sigma_w^2 + (\sigma_a^2 + \sigma_w^2) \sigma_\zeta^2}.$$

Clearly, this estimate is not consistent unless $\sigma_\zeta^2 = 0$. Intuitively, because z_{it} is correlated with a_{it} , the attenuation bias in the estimate λ translates into upward bias in the estimate of η . Note that the bias in $\hat{\eta}_{OLS}''$ is strictly increasing in σ_ζ^2 and, as informativeness of \tilde{a}_{it} falls (i.e., $\sigma_\zeta^2 \rightarrow \infty$), the probability limit of $\hat{\eta}_{OLS}''$ converges to the probability limit of $\hat{\eta}_{OLS}$. Thus, measurement error in the productivity proxy leads to inconsistent estimates of η although the bias is smaller than in the case of OLS estimates.

Overall, LP estimates of RTS are likely to be biased upward, although the bias is likely to be smaller than in OLS. The same problems can arise in other inversion-based estimators (e.g., Olley-Pakes 1996, Pavcnik 2002) because the dispersion of factor prices across firms does not allow non-stochastic inversion of firm's observed choices into firm's unobserved productivity. Inversion estimators can be a tenuous solution to the transmission bias problem because they ignore the variation in input mixes and/or measurement errors in inputs.

5 MONTE CARLO EXPERIMENTS

In this section I run a series of Monte Carlo experiments to evaluate the performance of the COV estimator and its alternatives. In each of these experiments, I draw factor prices, productivity and

²⁴ If χ is very large, the OLS bias $(1 - \eta)(\sigma_a^2 / \sigma_w^2) / (1 + \sigma_a^2 / \sigma_w^2)$ can be smaller than the LP bias $(1 - \eta)\chi / (1 + \chi)$.

other shocks from the normal distribution and for given realizations of the shocks I compute profit maximizing choices of revenue and inputs. Starting values of shocks are drawn from the corresponding unconditional distributions. For each replication, I generate a panel of 1,000 firms observed for 10 periods, which is close to typical sizes in applied work. I feed the generated data into various estimators and compute the estimates for structural parameters. I repeat the procedure 1,000 times and report median bias, standard deviation and root mean squared error (MSE) for each of the considered estimators.

In all experiments, RTS is $\gamma = 1.1$, which is consistent with the estimates of RTS from reports compiled by engineers (e.g., Pratten 1988), and the markup is $\mu = 2$. RTS in the revenue function is $\eta = \gamma/\mu = 0.55$. Factors are supplied in perfectly competitive markets unless otherwise specified. I choose a large markup to contrast the performance of the estimators (recall that bias increases in profit share which is $1-\eta$ in the considered case).

I consider the following estimators: OLS, fixed effects (FE), Schmidt's instrumental variables (SIV), Blundell-Bond (BB), COV and, where possible, Levinsohn-Petrin (LP). I use STATA's commands **xtabond2** and **levpet** for the BB and LP estimators, respectively. Schmidt's (1988) IV estimator uses (if necessary leads or lags of) input ratios as instruments for inputs. In designing COV estimator, I impose restrictions that are relevant for the given data generating process. Identification of the covariance estimator is ensured by Proposition 2 and Proposition 3.

5.1 ONE-INPUT/ONE-OUTPUT

The data generating process (DGP) for this set of experiments is given in (12)-(15). I start with the simplest calibration that allows no measurement error (Panel A, Table 2). SIV with $(y_{it} - z_{it})$ as the instrument for z_{it} is consistent. Note, however, that COV is overidentified while SIV is exactly identified and thus SIV has larger variance than COV.²⁵ OLS and FE have a predictably large bias in the estimated RTS. Although BB has a smaller bias than OLS, the reduction in the bias is small and the standard deviation of the estimates increases substantially. Figure 1 presents the kernel densities of the estimates. In agreement with my theoretical predictions, the density of the BB estimator is essentially flat, which is typical for all experiments and parameterizations that I consider.

²⁵ In fact, SIV does not have even first moments because it is exactly identified (Kinal 1980). It is an easy extension to make SIV overidentified by using leads or lags of input ratios wherever appropriate.

Next, I add measurement error to y and z (Panel B, Table 2). SIV with $(y_{i,t-1} - z_{i,t-1})$ as the instrument is consistent but it has standard deviation larger than that of COV. BB is considerably worse than FE in terms of MSE. Even OLS has a smaller MSE than BB. The somewhat better performance of OLS and FE can be explained by the fact that the measurement error attenuation (downward bias) partially offsets the upward transmission bias. This is particularly important for FE because the signal to noise ratio for FE falls more than that for OLS (see Griliches and Hausman 1986).

In the next experiment, I add serially correlated measurement errors to the input z_{it} (Panel C, Table 2). In particular, I assume that the measurement error is $\varepsilon_{it}^z = \rho_z \varepsilon_{i,t-1}^z + e_{it}$ with $\rho_z = 0.8, \sigma_e^2 = 1$. Note that SIV is not consistent because the input/output ratio is correlated with the measurement error and, consequently, the SIV's instrument is correlated with the error term in the revenue function at all leads and lags. Serial correlation of the measurement error deteriorates the signal to noise ratio and the attenuation bias becomes stronger. The performance of BB (quasi-differenced twice) remains poor: the bias and standard error are large.

Finally I consider the case when the factor price and productivity shocks are positively correlated (Panel D, Table 2). Specifically, I set $\rho(a_{it}, w_{it}) = 0.7$. This correlation invalidates the SIV estimator because any lead/lag of $(y_{it} - z_{it})$ is correlated with the residual in the revenue function. To highlight the consequences of the correlation, I assume no measurement errors. Because $\rho(w_{it}, a_{it}) \neq 0$, SIV has a very large downward bias so that the estimate of RTS is negative. COV is the only consistent estimator. Note that the bias in OLS, FE and BB estimates of RTS increases considerably because there is less exogenous variation in factor prices. FE is more biased than BB but FE dominates BB in terms of MSE.

Note that in all simulations productivity measured according to OLS, FE or BB estimates is poorly correlated with the true productivity. SIV provides a good correlation only when it is consistent. In contrast, COV estimates yield productivity measures that are highly correlated with true productivity.

5.2 MULTI-INPUT/ONE-OUTPUT

In this section I consider a more realistic setup with multiple inputs (capital, labor and materials) as in Section 4.3. The data generating process is given by (19)-(22) and the estimated revenue function is (18). I set $\beta_K = 0.1\eta, \beta_L = 0.2\eta, \beta_M = 0.7\eta$. I assume diagonal Ω and Π , i.e., factor

price and productivity are uncorrelated. Because capital is predetermined at time t , the appropriate instrument for capital in SIV is $(y_{i,t+1} - k_{i,t+1}) = -(\eta - 1 - \beta_K)^{-1} v_{i,t+1}^a + r_{i,t+1}$ that is uncorrelated with $a_{i,t-1}$ and v_{it}^a .²⁶

In the first experiment, I consider the case with no measurement error (Panel A, Table 3). COV has the smallest median bias and MSE. SIV has no bias in estimated RTS in the revenue function η but the variance of the estimate is large (recall that SIV is exactly identified). The LP estimator does better than OLS but LP still has a sizable upward bias, which is consistent with my theoretical predictions. Furthermore, computationally simpler FE has performance very close to that of the LP estimator. The BB estimator has a large negative bias in the coefficient on materials and a large upward bias on the coefficient on labor. Nonetheless, BB has a relatively small bias in the estimated RTS. I plot the kernel density of the estimators in Figure 2. Observe that the density of the LP estimator almost coincides with FE's density. Also note the flat density of the BB estimator.

To show the importance of the small profit share for the estimate of the bias, I vary the demand elasticity so that profit share ranges from 50% to 0.1%. Figure 3 plots the bias as a function of the profit share. Note that BB, SIV, and LP reduce the bias relative to OLS but as profit share falls these estimators yield only a minor reduction in the bias. Interestingly, LP only marginally improves upon FE. Given that LP and BB tend to have larger variance than OLS, it is not clear if popular solutions to the transmission bias are better in terms of MSE than the OLS estimate.

In the next experiment, I add a small measurement error to inputs and revenue to assess the sensitivity of BB and inversion-based LP to measurement errors (Panel B, Table 3). Predictably, the attenuation bias partially offsets the transmission bias and, thus, the estimates of RTS are less biased than in the absence of measurement errors. Nonetheless, BB has an increased bias because one has to take more distant lags in the moment conditions. This greatly deteriorates the performance of the estimator. Although the LP estimator has a smaller bias in the estimated RTS η , the upward bias in the coefficient on materials is reallocated to the upward bias in the coefficients on capital and labor. Overall, LP is very similar to FE. Only SIV and COV yield consistent estimates in this experiment.

²⁶ Timing of shocks modifies the moments used in the BB estimator. However, the fact that output, labor and materials are determined simultaneously and the response of labor, materials and revenue to v^a is identical leads to the reduced rank problem for the BB estimator (see proof of Proposition 5).

Next I examine the case with an upward-sloping labor supply curve. I set the elasticity of the labor cost to $\phi_L = 1.5$ and I assume no measurement error (Panel C, Table 3). Importantly, although the base wage $\log(W_{it})$ is uncorrelated with productivity a_{it} , the log of wage $W_{it}L_{it}^{\phi-1}$ is correlated with a_{it} . Since the log wage is correlated with a_{it} , SIV is not consistent. Note that the OLS, BB and FE estimates exceed unity although the true RTS in the revenue function is 0.55. Only COV estimate the parameters consistently. Note that in this scenario the correlation between true productivity and productivity measured according to SIV estimates becomes negative.

In the next series of experiments, I assume quadratic costs of adjustment for capital and keep the rest of the assumptions unchanged. In brief, the firm solves the following dynamic problem:

$$\max E_0 \sum_{t=0}^{\infty} (Y_{it} - R_{it}I_{it} - W_{it}L_{it} - P_{it}^M M_{it} - \frac{1}{2}\psi(I_{it}/K_{i,t-1} - \delta)^2 K_{i,t-1}) \quad (26)$$

$$\begin{aligned} s.t. \quad Y_{it} &= A_{it}K_{i,t-1}^{\beta_K}L_{it}^{\beta_L}M_{it}^{\beta_M}, \\ K_{it} &= (1 - \delta)K_{i,t-1} + I_{it}, \\ W_{it} &= W_{it}^0L_{it}^{\phi-1}, \\ r_{it} &= \rho_r r_{i,t-1} + v_{it}^r, \\ w_{it}^0 &= \rho_w w_{i,t-1}^0 + v_{it}^w, \\ p_{it}^M &= \rho_{p^M} p_{i,t-1}^M + v_{it}^M, \\ a &= \rho_a a_{i,t-1} + v_{it}^a, \end{aligned}$$

where I is investment, ψ is the adjustment cost parameter, small letter denote logs of the respective variables. In all simulations I set $\psi = 6$, which is consistent with the available estimates of adjustment costs (e.g., Gordon 1992), and estimate other parameters of the model. I log-linearize the first-order conditions and constraints given perfect foresight. Because the analytical solution to the above problem is complicated, it is hard to establish that the covariance estimator is uniquely globally identified. However, since the numeric solution can be readily written in the state-space form (see Appendix A), it is straightforward to establish local identification of the parameters by checking the rank of the Jacobian.

Using the log-linearized solution to the model, I generate artificial data sets and feed them into various estimators. Table 4 presents the results for the cases with perfectly and imperfectly elastic factor supply curves and with/without measurement errors. In the baseline

experiment with perfect competition in factor markets ($\phi = 1$) and no measurement errors (panel A, Table 4), OLS, FE and BB estimates are biased so much that the estimated RTS are increasing (recall that the true RTS in the revenue function is equal to 0.55). Consistent with the argument in section 4.2, BB estimates have large standard errors, which indicate poor identification of the estimator. Although the LP estimator has a smaller bias than other estimators, the size of the bias remains very large. SIV produces implausible estimates because the shadow price of capital is correlated with technology and, hence, no lead or lag of the output-to-capital ratio is a valid instrument for the level of the capital stock. This correlation of the shadow price of capital and technology is the key to understanding why the conventional estimators yield increasing RTS even when true returns are well below unity. Because of the attenuation bias, adding measurement error (panel B, Table 4) reduces the bias in the estimated RTS. In the case with an upward-sloping labor supply curve ($\phi = 1.5$, panel C, Table 4) the bias tends to increase in the estimate of β_L and decrease in the estimate of β_K . Nonetheless, because β_L and β_K have a small contribution to the RTS (recall that the elasticity of output with respect to material is 0.7), the bias in the estimate of RTS barely changes.

Note that in all simulations with popular multiple-input estimators lead to inferior productivity measures. Specifically, small departures from the assumptions required by SIV can produce a negative correlation between true productivity and productivity measured according to SIV estimates. In all simulations, other estimators (OLS, LP, FE, BB) yield low correlation between true and measured productivity. Typically, the correlation for these estimators is well below 0.4. The covariance estimator performs well in terms of measuring productivity and capturing returns to scale, although the standard error of the coefficient on capital is somewhat large.

5.3 DISCUSSION

The results of Monte Carlo experiments are in agreement with my theoretical predictions that LP is biased upwards and BB is poorly identified. SIV is extremely sensitive to serially correlated measurement errors and (shadow) factor prices being correlated with technology. The experiments show that simpler OLS and FE have performance comparable to that of BB and LP. If the profit share is small, the reduction in the bias from using BB and LP is offset by an increase in the variance of the estimates. Hence, in empirically plausible settings with small profit shares, it is useful to compare estimates from sophisticated techniques with OLS estimates.

Furthermore, the researcher can reach incorrect conclusions about relative productivity of firms and magnitudes of productivity differences across firms when he or she uses inconsistent estimates. Specifically, the experiments suggest that OLS, FE, BB, and LP tend to yield productivity measures that are poorly correlated with true productivity.²⁷ SIV's performance varies from good to disastrous. If the assumptions of SIV are satisfied, the correlation is close to one. Slight departures from SIV's assumptions can lead to a negative correlation between measured and true productivity. These results are true irrespective of the absolute size of the bias in the estimates because productivity and factor price shocks are amplified in the same proportion for observed optimized choices of firms. In contrast, productivity measures constructed on the basis of COV estimates are highly correlated with true productivity.

Importantly, the Monte Carlo experiments suggest that the puzzling estimates of RTS in the revenue function can arise because statistical estimators fail to provide consistent estimates. In the next section, I contrast the estimates of competing techniques when applied to real data.

6 APPLICATION

In this section, I apply the COV estimator to a well-known data set of Chilean manufacturing plants. Lui (1991, 1993), Lui and Tybout (1996), Pavcnik (2002) and Petrin and Levinsohn (2005) describe the data in detail. To illustrate the estimator, I focus on a balanced sample of firms in the SIC-3240 industry (manufacture of footwear).²⁸ The annual data spans from 1982 to 1996. Descriptive statistics for logs of real value added, real capital stock and labor are presented in Table 5 and Table 6.

I assume that the inverse demand function is given by $P_{it} = P(Q_{it})$ with demand elasticity σ and the markup is $\mu = \sigma/(\sigma - 1)$. The production function is described by $Q_{it} = A_{it}^{\mu} \min\{M_{it}, cF(K_{it}, L_{it})\}$, where Q_{it} is output in physical units, M_{it} is the input of materials, K_{it} is capital, L_{it} is the number of employees, A_{it} is the level of Hicks-neutral technology (A_{it} to the power of μ is a normalization), and c is a constant of proportionality. This functional form imposes zero substitution between materials and combined capital/labor inputs. Since at the optimum no resources are wasted, $M_{it} = cF(K_{it}, L_{it})$ and, hence, the profit function is given by $\pi_{it} = P_{it}Q_{it} - P_{it}^M M_{it} - R_{it}K_{it} - W_{it}L_{it} = VA_{it} - R_{it}K_{it} - W_{it}L_{it} = A_{it}cF(K_{it}, L_{it}) - R_{it}K_{it} - W_{it}L_{it}$, where

²⁷ Productivity measured as a cost-based Solow residual can be poorly correlated with the true productivity when factor supply curves are upward sloping.

²⁸ I am grateful to Jim Levinsohn for providing me with the data.

VA_{it} is the value added, R_{it} and W_{it} are the cost of capital and labor for firm i at time t .²⁹ The elasticity of value added with respect to capital and labor is β_K and β_L . In the data, the share of materials in total cost is 0.66.

I assume that capital is supplied in perfectly competitive markets. The slope of the labor supply curve is a free parameter. In particular, I assume that the elasticity of the wage function $W_{it}(L) = W_{it}^0 C(L)$ with respect to labor is $\phi - 1$ so that the elasticity of the wage bill $W_{it}(L_{it})L_{it}$ with respect to labor is ϕ . I further assume that capital, labor, and revenue are chosen simultaneously. I allow serially correlated errors in all observed variables, which are capital, labor and revenue. Unobserved technology and factor prices are serially correlated and there could be feedback from technology to factor prices and vice versa. I assume that innovations to technology and factor prices are uncorrelated. In summary, the estimated log-linearized model is

$$y_{it}^* - k_{it}^* = r_{it}, \quad (27)$$

$$y_{it}^* - l_{it}^* = w_{it} + (\phi - 1)l_{it}^*, \quad (28)$$

$$y_{it}^* = a_{it} + \beta_K k_{it}^* + \beta_L l_{it}^*, \quad (29)$$

$$a_{it} = \rho_{aa} a_{i,t-1} + \rho_{aw} w_{i,t-1} + \rho_{ar} r_{i,t-1} + v_{it}^a, \quad (30)$$

$$w_{it}^0 = \rho_{wa} a_{i,t-1} + \rho_{ww} w_{i,t-1}^0 + \rho_{wr} r_{i,t-1} + v_{it}^w, \quad (31)$$

$$r_{it} = \rho_{ra} a_{i,t-1} + \rho_{rw} w_{i,t-1} + \rho_{rr} r_{i,t-1} + v_{it}^r, \quad (32)$$

$$y_{it} = y_{it}^* + u_{it}^y, \quad (33)$$

$$k_{it} = k_{it}^* + u_{it}^k, \quad (34)$$

$$l_{it} = l_{it}^* + u_{it}^l, \quad (35)$$

$$u_{it}^y = \rho_y u_{i,t-1}^y + \varepsilon_{it}^y, \quad (36)$$

$$u_{it}^k = \rho_k u_{i,t-1}^k + \varepsilon_{it}^k, \quad (37)$$

$$u_{it}^l = \rho_l u_{i,t-1}^l + \varepsilon_{it}^l, \quad (38)$$

where small letters denote logs of the respective variables with $y_{it} = \ln VA_{it}$, stars denote true values, and $\{v_{it}^a, v_{it}^w, v_{it}^r, \varepsilon_{it}^y, \varepsilon_{it}^k, \varepsilon_{it}^l\}$ are uncorrelated i.i.d. innovations. Parameters of interest are

²⁹ Since I analyze value added, I do not have the problem of measuring the quantity of the materials input. Note that in the vast majority of cases the researcher knows only the nominal spending on materials and the quantity of the material input is obtained by deflating the nominal spending with the industry-level material price index. Since the mix of intermediate inputs varies across firms and the price index is the same for all firms in any given period, the computed quantity of the material input can be poorly correlated with the true quantity of the material input. In the case of the Cobb-Douglas production function, nominal spending on materials is proportional to revenue and, hence, including the deflated expenditures on materials should yield perfect collinearity. Stochastic errors (e.g., optimization errors, measurement errors) can break the collinearity but the coefficient is still likely to be close to unity, which is often the case in applications. In this case, identification of materials elasticity can require additional assumptions.

β_K , β_L and returns to scale in the value-added function $\eta = \beta_K + \beta_L$. Equations (27) and (28) are the first order conditions for capital and labor. Equation (29) is the value-added function (analogue to the revenue function). Equations (30)-(32) describe the evolution of structural shocks to productivity and factor prices. Measurement equations are collected in (33)-(35). Dynamics of the measurement errors are in equations (36)-(38). The equations can be succinctly rewritten in the matrix form that corresponds to the state space representation in (8)-(9):

$$\mathbf{X}_{it} \equiv \begin{bmatrix} y_{it} \\ k_{it} \\ l_{it} \end{bmatrix} = \begin{bmatrix} \frac{-\phi}{\beta_L + \beta_K \phi - \phi} & \frac{\beta_L}{\beta_L + \beta_K \phi - \phi} & \frac{\beta_K}{\beta_L + \beta_K \phi - \phi} & | & 1 & 0 & 0 \\ \frac{-\phi}{\beta_L + \beta_K \phi - \phi} & \frac{\beta_L}{\beta_L + \beta_K \phi - \phi} & \frac{\beta_L - \phi}{\beta_L + \beta_K \phi - \phi} & | & 0 & 1 & 0 \\ \frac{-1}{\beta_L + \beta_K \phi - \phi} & \frac{1 - \beta_K}{\beta_L + \beta_K \phi - \phi} & \frac{\beta_L}{\beta_L + \beta_K \phi - \phi} & | & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{it} \\ w_{it} \\ r_{it} \\ u_{it}^y \\ u_{it}^k \\ u_{it}^l \end{bmatrix} = [\Lambda \ | \ I_3] \mathbf{F}_{it},$$

$$\mathbf{F}_{it} \equiv \begin{bmatrix} a_{it} \\ w_{it} \\ r_{it} \\ u_{it}^y \\ u_{it}^k \\ u_{it}^l \end{bmatrix} = \begin{bmatrix} \rho_{aa} & \rho_{aw} & \rho_{ar} & | & 0 & 0 & 0 \\ \rho_{wa} & \rho_{ww} & \rho_{wr} & | & 0 & 0 & 0 \\ \rho_{ra} & \rho_{rw} & \rho_{rr} & | & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & | & \rho_y & 0 & 0 \\ 0 & 0 & 0 & | & 0 & \rho_k & 0 \\ 0 & 0 & 0 & | & 0 & 0 & \rho_l \end{bmatrix} \begin{bmatrix} a_{i,t-1} \\ w_{i,t-1} \\ r_{i,t-1} \\ u_{i,t-1}^y \\ u_{i,t-1}^k \\ u_{i,t-1}^l \end{bmatrix} + \begin{bmatrix} v_{it}^a \\ v_{it}^w \\ v_{it}^r \\ \varepsilon_{it}^y \\ \varepsilon_{it}^k \\ \varepsilon_{it}^l \end{bmatrix} = \begin{bmatrix} \Pi & | & 0 \\ \hline 0 & | & R \end{bmatrix} \mathbf{F}_{it} + \begin{bmatrix} v_{it} \\ \varepsilon_{it} \end{bmatrix}.$$

The model has 21 parameters:

$$\theta = \{\beta_K, \beta_L, \phi, \rho_{aa}, \rho_{aw}, \rho_{ar}, \rho_{wa}, \rho_{ww}, \rho_{wr}, \rho_{ra}, \rho_{rw}, \rho_{rr}, \sigma_{va}, \sigma_{vw}, \sigma_{vr}, \sigma_{\varepsilon y}, \sigma_{\varepsilon k}, \sigma_{\varepsilon l}, \rho_y, \rho_k, \rho_l\},$$

where $\sigma_{va}, \sigma_{vw}, \sigma_{vr}$ are standard deviations of innovations to productivity (v_{it}^a), wages (v_{it}^w) and the interest rate (v_{it}^r), and $\sigma_{\varepsilon y}, \sigma_{\varepsilon k}, \sigma_{\varepsilon l}$ are innovations to measurement errors in value added (ε_{it}^y), capital stock (ε_{it}^k), and labor input (ε_{it}^l). Global identification is guaranteed by Proposition 2 and Proposition 3. I use MLE given in (11) to estimate the model with $T=5$.

I report the estimation results in Table 7, column 1. Since the data are not normally distributed, I bootstrap the estimates to correct the bias and improve the confidence intervals.³⁰ Using bootstrapped critical values, I do not reject the model at any conventional significance level (p-value=0.4). To contrast the results, I estimate the value-added function $y_{it} = \beta_K k_{it} + \beta_L l_{it} + \xi_{it}$ by OLS, FE, LP, and BB estimators and report these results in columns 2

³⁰ I use non-parametric bootstrap with resampling firms. See Horowitz (1998) for the discussion of bootstrap for covariance structures.

to 7 in Table 7. I report two versions of the BB estimator: quasi-differenced (column 6) and twice-quasi-differenced (column 7).

BB, LP and FE estimators yield RTS in a 0.62 to 0.9 range. These estimates suggest a very large 10-38% profit share in value added if factor markets are perfectly competitive. In contrast, the observed (*accounting*) profit share in value added is 2%.³¹ Also observe that, consistent with our theoretical results and Monte Carlo experiments, the BB estimator has very large standard errors and LP estimates are close to FE estimates. On the other hand, the OLS estimate (RTS=1.30) is inconsistent with profit maximization if factor markets are perfectly competitive. In addition, the OLS estimate of β_L implies increasing returns in labor. The SIV estimator yields implausibly large RTS. This cacophony in the estimates can be reconciled by the COV estimates.

First, note that the COV estimates RTS in the revenue function to be 1.17, which is in line with our argument that the bias in the OLS estimate of RTS is likely to be relatively small. Second, the estimate of ϕ is greater than unity and, thus, firms face an upward-sloping labor supply curve. Since the OLS estimate is biased to ϕ , the OLS estimate of RTS is greater than COV estimate of RTS. Third, I find relatively large measurement errors. These errors tend to attenuate the estimates toward zero, especially when estimates are from within variation. This can explain why FE, BB, and LP produce low RTS. Note that the small coefficient on capital in BB is consistent with strong downward bias in β_K in my Monte Carlo experiments with serially correlated measurement errors. Finally, since the SIV estimator uses output to input ratios as instruments and measurement error is present in inputs and factor prices are correlated with technology, the instruments used in the SIV are correlated with the error term in the revenue function so that the estimates of β_L and β_K behave wildly.

Increasing returns in the revenue function do not contradict profit maximization because the labor supply curve is upward sloping. Specifically, the elasticity of the labor cost $\phi = 1.42$ (i.e., the wage premium is 42%) is generally in agreement with the estimates from previous studies. For example, Shapiro (1986, 1996) and Bils (1987) estimate from aggregate US data that the shift premium is about 25-40%. Other factors (monopsony, rent-sharing, etc.) can further increase the slope of the labor supply curve. Manning (2004) observes that a plausible elasticity

³¹ The profit share is computed as the ratio of aggregate gross profit to aggregate value added. Although it is hard to sign the bias of the accounting profit as a measure of economic profit, the small magnitude of the profit share is consistent with the discussion in section 2. Alternative definitions of the profit share are in the range of 0.2% to 2.5%.

of labor supply is between 2 and 10. In the present case, the implied elasticity of the labor supply curve $1/(\phi - 1) = 2.4$ falls nicely in this interval.

According to (41) in Appendix A, I find that the implied elasticity of the cost of capital and labor is 1.21 and elasticity of the cost of all inputs is 1.07. Using (42) in Appendix A to compute the profit share from the COV estimates, I find that the profit share in value added is 1.3%, which is a significant improvement in comparison to other estimators.

The correlation between productivity measured according to COV estimates and productivity measured according to OLS, FE, SIV, LP, BB and BB2 estimates is 0.884, 0.637, 0.248, 0.361, 0.666, and 0.671 respectively. Therefore, one can reach strikingly different conclusions about which firms are relatively productive when he or she uses different estimates to construct productivity measures. Given the results in Monte Carlo simulations one may trust COV measures of productivity more than measures of productivity constructed on the basis of other estimators.³²

Note that variation in factor prices is comparable to variation in productivity a_{it} . Specifically, the point estimates in Table 7 imply that $\sigma(a_{it}) = 0.332$, $\sigma(w_{it}) = 0.513$, $\sigma(r_{it}) = 0.230$. This supports other evidence on the dispersion of prices even in narrowly defined industries. I also conclude that ignoring variation in factor prices across firms can lead to serious identification problems for the inversion-based estimators. Finally, since the markup $\mu \geq 1$ and returns to scale in the production function $\gamma = \mu\eta \geq \eta$, one can expect sizable increasing returns to scale in production.

7 CONCLUSION

This paper has critical and constructive parts. In the critical part, I demonstrate that under weak assumptions estimates from production function regressions using firm-level data are often inconsistent with profit maximization or imply implausibly large profits. Specifically, I argue that firm-level data limitations lead to estimating returns to scale in the revenue function, not production function. On the other hand, I show that returns to scale in the revenue function cannot be greater than unity or significantly less than unity as long as the profit share in revenue is non-negative and factor supplies are perfectly elastic. This prediction sharply contrasts with the frequent finding that returns to scale in the revenue function at the firm level exceed unity or

³² It would be highly informative to compare various productivity measures with measurements based on detailed case studies or expert assessments of relative performance. Unfortunately, this external information is not available.

are well below unity. On the econometric front, I point out that inversion-based estimators (e.g., Olley and Pakes 1996, Pavcnik 2002, Levinsohn and Petrin 2003) lead to inconsistent estimates because they ignore variation in input mixes (factor prices). I also show that GMM/IV estimators using lags of endogenous variables as instruments (e.g., Blundell and Bond 1998, 1999) can be poorly identified in the context of estimating production (revenue) functions because of economic restrictions on the comovement of inputs and output. Furthermore, I show that these misspecifications greatly distort measures of productivity so that the researcher using these estimators can be led to grossly incorrect conclusions about the relative productivity of firms and magnitudes of productivity differences across firms. In summary, puzzling estimates can stem from applying misspecified or poorly identified estimators.

In the constructive part, I show that under weak assumptions the elasticity of the factor supply can reconcile increasing or large decreasing returns in the revenue function and a small non-negative profit share. Furthermore, I argue that simple structural estimators that model the cost and the revenue function simultaneously and treat unobserved heterogeneity in productivity and factor prices symmetrically can resolve many of the problems I identify above. I provide an example and illustrate the strength of the suggested estimator in Monte Carlo simulations and in an empirical application. The paper also provides a link between revenue function parameters and other structural parameters describing the cost side, production and markup. So even when we estimate revenue functions, we can still obtain useful information about objects we do not observe directly (e.g., returns to scale in production).

The paper has broader implications. First, I argue that the profit share can be used as a robust non-parametric *economic* diagnostic for estimates of returns to scale. Second, although I analyze only one industry, it is clear that variation in product and factor prices at the firm-level is not trivial. This entails important consequences for aggregating firm-level data (and devastating effects on the inversion-based estimators). Specifically, reallocation effects due to heterogeneity in factor prices are likely to be of first-order importance. Furthermore, productivity aggregates measure revenue-generating ability in the industry rather than technical efficiency. Third, since it is fairly common to find constant returns to scale in the revenue function at the firm level and the markup is not less than unity, returns to scale in production at the firm level can be sizeable. Hence, business cycle and trade models appropriately calibrated to capture increasing returns to scale in production (not constant returns to scale in revenue!) can produce qualitatively different results. In addition, the gap between RTS in production at the aggregate level and RTS in

production at the firm level is smaller than thought before because RTS in production at the aggregate level were compared to RTS in revenue at the micro level. Fourth, factor supply curves are likely to be upward-sloping at the firm level. This means that the cost-weighted composite input does not measure total input correctly and, hence, cost-based Solow residual can be procyclical.

Table 1. Profit share s_π as a function of returns to scale in the revenue function and the elasticity of the cost with respect to inputs

Returns to scale in the revenue function	Elasticity of the cost with respect to inputs		
	$\phi < 1$	$\phi \approx 1$	$\phi > 1$
$\eta < 1$	small s_π	large s_π	large s_π
$\eta \approx 1$	negative s_π	small s_π	large s_π
$\eta > 1$	negative s_π	negative s_π	small s_π

Table 2. Estimates of returns to scale: One-output/one-input

		OLS	FE	BB	SIV	COV
		(1)	(2)	(3)	(4)	(5)
Panel A: no measurement error						
η	Median bias	0.359	0.272	0.221	-0.001	0.001
	Standard deviation	0.003	0.003	0.112	0.013	0.006
	Root MSE	0.359	0.272	0.251	0.013	0.006
$\rho(\hat{a}_{it}, a_{it})$	Median estimate	0.658	0.454	0.381	1.000	1.000
Panel B: i.i.d. measurement error						
η	Median bias	0.332	0.217	0.225	-0.001	0.001
	Standard deviation	0.004	0.005	0.259	0.039	0.007
	Root MSE	0.332	0.217	0.348	0.039	0.008
$\rho(\hat{a}_{it}, a_{it})$	Median estimate	0.450	0.341	0.281	1.000	1.000
Panel C: serially correlated measurement error						
η	Median bias	0.288	0.192	0.145	-0.267	0.000
	Standard deviation	0.006	0.005	0.200	0.024	0.018
	Root MSE	0.288	0.193	0.247	0.269	0.018
$\rho(\hat{a}_{it}, a_{it})$	Median estimate	0.501	0.412	0.339	0.916	0.997
Panel D: correlated factor prices and productivity						
η	Median bias	0.423	0.313	0.223	-1.773	0.001
	Standard deviation	0.004	0.005	0.413	0.323	0.008
	Root MSE	0.423	0.313	0.475	1.844	0.008
$\rho(\hat{a}_{it}, a_{it})$	Median estimate	0.465	0.322	0.240	0.975	0.987

Note: The table reports median bias, standard deviation and root MSE for OLS, Schmidt's instrumental variables (SIV), covariance (COV), fixed effects (FE), and Blundell-Bond (BB) estimators. The data generating process is (12)-(15): one input and one output. Each experiment is simulated 1,000 times. In all experiments, $\rho_a = 0.9, \rho_w = 0.5, \sigma_{va} = \sigma_{vw} = 1$. In panel A, $\rho(v_{it}^w, v_{it}^a) = 0, \sigma_{\varepsilon z} = \sigma_{\varepsilon y} = 0$. In panel B, $\rho(v_{it}^w, v_{it}^a) = 0, \sigma_{\varepsilon z} = \sigma_{\varepsilon y} = 1$. In panel C, $\rho(v_{it}^w, v_{it}^a) = 0, \sigma_{\varepsilon y} = 0, \varepsilon_{it}^z = \rho_z \varepsilon_{i,t-e}^z + e_{it}, \sigma_e^2 = 1, \rho_z = 0.8$. In panel D, $\rho(v_{it}^w, v_{it}^a) = 0.7, \sigma_{\varepsilon z} = \sigma_{\varepsilon y} = 0$. $\rho(\hat{a}_{it}, a_{it})$ is the correlation between true productivity a_{it} and measured productivity \hat{a}_{it} given the estimate of η . See text for further details.

Table 3. Estimates of returns to scale: One-output/multi-input

Parameter values		OLS (1)	FE (2)	BB (3)	LP (4)	SIV (5)	COV (6)
Panel A: no measurement error							
β_K	Median bias	0.033	-0.011	-0.019	0.001	-0.009	0.000
	St. Dev.	0.004	0.003	0.054	0.003	0.100	0.007
β_L	Median bias	0.265	0.270	0.482	0.265	-0.001	-0.001
	St. Dev.	0.006	0.006	0.206	0.006	0.034	0.010
β_M	Median bias	0.123	0.096	-0.302	0.087	0.004	0.000
	St. Dev.	0.006	0.006	0.130	0.006	0.049	0.007
η	Median bias	0.421	0.356	0.161	0.353	-0.006	0.000
	St. Dev.	0.002	0.003	0.074	0.004	0.054	0.007
	Root MSE	0.421	0.356	0.177	0.353	0.054	0.007
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.505	0.389	0.342	0.666	1.000	1.000
Panel B: i.i.d. measurement error							
β_K	Median bias	0.077	0.032	0.042	0.048	-0.017	0.000
	St. Dev.	0.006	0.006	0.230	0.007	0.144	0.011
β_L	Median bias	0.269	0.259	0.254	0.269	-0.007	0.001
	St. Dev.	0.007	0.008	0.274	0.007	0.077	0.025
β_M	Median bias	0.069	0.042	0.092	0.014	0.000	0.000
	St. Dev.	0.007	0.008	0.264	0.010	0.099	0.038
η	Median bias	0.415	0.334	0.388	0.331	-0.024	0.001
	St. Dev.	0.003	0.005	0.260	0.008	0.126	0.025
	Root MSE	0.415	0.334	0.467	0.331	0.129	0.025
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.612	0.420	0.391	0.709	1.000	1.000

(continued on next page)

Table 3 (continued)

Parameter values		OLS	FE	BB	LP	SIV	COV
		(1)	(2)	(3)	(4)	(5)	(6)
Panel C: upward sloping labor supply curve							
β_K	Median bias	0.044	-0.002	-0.017	0.010	-0.075	0.000
	St. Dev.	0.004	0.004	0.052	0.003	0.069	0.008
β_L	Median bias	0.443	0.446	0.791	0.443	1.810	-0.001
	St. Dev.	0.009	0.009	0.285	0.009	0.120	0.017
β_M	Median bias	0.117	0.093	-0.297	0.083	-0.500	0.000
	St. Dev.	0.006	0.006	0.108	0.006	0.026	0.008
η	Median bias	0.604	0.536	0.477	0.536	1.235	-0.001
	St. Dev.	0.004	0.005	0.178	0.006	0.045	0.013
	Root MSE	0.604	0.536	0.509	0.536	1.236	0.013
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.310	0.253	0.203	0.552	-0.414	0.999

Note: The table reports median bias, st. dev. and MSE of OLS, Schmidt's instrumental variables (SIV), covariance (COV), fixed effects (FE), Blundell-Bond (BB), and Levinsohn-Petrin (LP) estimators. The data generating process is (19)-(22): three inputs and one output. The estimated revenue function is (18). Each experiment is simulated 1,000 times. In all parameterizations, $\beta_K=0.1\eta$, $\beta_L=0.1\eta$, $\beta_M=0.1\eta$, $\eta=0.55$, $\rho_r=0.5$, $\rho_w=0.6$, $\rho_{p^M}=0.4$, $\rho_a=0.9$, $\sigma_{vr} = \sigma_{vw} = \sigma_{vp^M} = \sigma_{va} = 1$. In panel A, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 0$, $\phi = 1$. In panel B, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 1$, $\phi = 1$. In panel C, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 0$, $\phi = 1.5$. $\rho(\hat{a}_{it}, a_{it})$ is the correlation between true productivity a_{it} and measured productivity \hat{a}_{it} given the estimate of β_K , β_L , and β_M . See text for further details.

Table 4. Estimates of returns to scale: One-output/multi-input with adjustment costs

Parameter values		OLS (1)	FE (2)	BB (3)	LP (4)	SIV (5)	COV (6)
Panel A: no measurement error							
β_K	Median bias	0.169	0.250	0.285	0.187	3.714	-0.002
	St. Dev.	0.011	0.025	0.495	0.014	0.512	0.044
β_L	Median bias	0.442	0.389	0.345	0.442	-0.093	0.000
	St. Dev.	0.007	0.006	0.170	0.007	0.084	0.018
β_M	Median bias	-0.096	-0.087	0.106	-0.186	-0.079	0.000
	St. Dev.	0.005	0.005	0.163	0.007	0.035	0.011
η	Median bias	0.516	0.552	0.735	0.443	3.542	-0.002
	St. Dev.	0.009	0.023	0.454	0.011	0.444	0.045
	Root MSE	0.266	0.305	0.747	0.197	12.741	0.002
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.361	0.276	0.121	0.487	-0.486	1.000
Panel B: i.i.d. measurement error							
β_K	Median bias	0.153	0.059	0.032	0.154	-0.055	0.008
	St. Dev.	0.011	0.014	0.178	0.011	61.911	0.061
β_L	Median bias	0.369	0.317	1.105	0.369	2.385	-0.001
	St. Dev.	0.008	0.008	0.339	0.008	51.015	0.034
β_M	Median bias	-0.037	-0.047	-0.696	-0.086	-1.893	-0.002
	St. Dev.	0.008	0.007	0.312	0.010	35.505	0.018
η	Median bias	0.486	0.329	0.441	0.437	0.437	0.005
	St. Dev.	0.009	0.015	0.230	0.014	46.387	0.057
	Root MSE	0.236	0.108	0.248	0.191	2151.950	0.003
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.294	0.206	0.065	0.218	-0.436	0.999

(continued on next page)

Table 4 continued

Parameter values		OLS	FE	BB	LP	SIV	COV
		(1)	(2)	(3)	(4)	(5)	(6)
Panel C: upward sloping labor supply curve							
β_K	Median bias	-0.014	-0.021	0.339	-0.004	3.026	0.000
	St. Dev.	0.005	0.013	0.454	0.016	0.369	0.046
β_L	Median bias	0.455	0.398	0.628	0.455	-0.014	-0.001
	St. Dev.	0.005	0.005	0.228	0.005	0.059	0.028
β_M	Median bias	0.039	0.062	-0.140	0.021	-0.012	0.000
	St. Dev.	0.004	0.004	0.074	0.010	0.032	0.011
η	Median bias	0.481	0.439	0.827	0.472	2.999	-0.001
	St. Dev.	0.003	0.011	0.442	0.011	0.315	0.048
	Root MSE	0.231	0.193	0.880	0.223	9.096	0.002
$\rho(\hat{a}_{it}, a_{it})$	Median est.	0.267	0.237	0.349	0.154	-0.492	0.989

Note: The table reports median bias, st. dev. and MSE of OLS, Schmidt's instrumental variables (SIV), covariance (COV), fixed effects (FE), Blundell-Bond (BB), and Levinsohn-Petrin (LP) estimators. The data generating process is (26). The estimated revenue function is (18). Each experiment is simulated 1,000 times. In all parameterizations, $\beta_K=0.1\eta$, $\beta_L=0.1\eta$, $\beta_M=0.1\eta$, $\eta=0.55$, $\rho_r=0.5$, $\rho_w=0.6$, $\rho_{p^M}=0.4$, $\rho_a=0.9$, $\sigma_{vr} = \sigma_{vw} = \sigma_{vp^M} = \sigma_{va} = 1$, $\psi = 6$. In panel A, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 0$, $\phi = 1$. In panel B, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 1$, $\phi = 1$. In panel C, $\sigma_{vy} = \sigma_{ek} = \sigma_{el} = 0$, $\phi = 1.5$. $\rho(\hat{a}_{it}, a_{it})$ is the correlation between true productivity a_{it} and measured productivity \hat{a}_{it} given the estimate of β_K , β_L , and β_M . See text for further details.

Table 5. Descriptive statistics

Variable	variation	Mean	Std. Dev.	Min	Max
Ln(real value added)	overall	4.190	1.636	0.269	9.437
	between		1.548	0.793	8.816
	within		0.528	1.790	6.355
Ln(real capital stock)	overall	7.649	1.799	3.432	12.492
	between		1.701	3.935	12.330
	within		0.448	4.973	9.627
Ln(number of employees)	overall	3.763	1.078	2.303	7.145
	between		0.986	2.303	6.709
	within		0.325	1.982	5.095

Note: This table reports descriptive statistics for Chilean manufacturing plants in SIC 3240 industry (Manufacture of footwear). The time span is from 1982 to 1996. Real value added is nominal value added deflated by the industry price index. Employment includes production and non-production workers. Capital stock, which includes machines and structures, is constructed by perpetual inventory method. See references cited in the text for further information.

Table 6. Covariance and autocovariance matrices

	Y_t	K_t	L_t
Y_t	261.5	256.3	168.0
K_t	256.3	344.8	176.3
L_t	168.0	176.3	126.0
Y_{t-1}	244.8	251.3	163.7
K_{t-1}	249.2	333.4	171.9
L_{t-1}	163.5	172.5	120.0
Y_{t-2}	239.3	248.0	160.7
K_{t-2}	244.0	324.0	167.8
L_{t-2}	159.5	169.2	116.0
Y_{t-3}	233.3	245.1	157.3
K_{t-3}	239.5	316.1	164.2
L_{t-3}	155.6	166.5	112.3
Y_{t-4}	230.4	243.6	155.3
K_{t-4}	234.1	308.2	159.9
L_{t-4}	152.4	163.5	109.2

Note: This table presents covariance and autocovariance matrices for logs of value added (Y_t), capital stock (K_t) and labor (L_t) after projecting these variables on the complete set of time dummies. See note to Table 5 for further details.

Table 7. Estimation results

	COV	OLS	FE	SIV	LP	BB	BB-2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_K	0.498 [0.423, 0.514]	0.198 (0.017)	0.146 (0.029)	-0.398 (0.050)	0.135 (0.054)	0.197 (0.130)	0.2099 (0.123)
β_L	0.697 [0.510, 0.730]	1.105 (0.029)	0.677 (0.047)	2.952 (0.131)	0.672 (0.073)	0.676 (0.132)	0.6897 (0.128)
ϕ	1.420 [1.307, 1.578]						
η	1.172 [1.008, 1.226]	1.302 (0.017)	0.822 (0.043)	2.554 (0.089)	0.807 (0.112)	0.874 (0.161)	0.899 (0.160)
Factor prices and technology: standard deviation of innovations and serial correlation							
σ_{va}	0.0306 [0.002, 0.408]	ρ_{aa}	0.9059 [0.732, 0.961]	ρ_{aw}	-0.0602 [-0.163, 0.031]	ρ_{ar}	-0.2719 [-0.463, -0.042]
σ_{vw}	0.0163 [0.001, 0.442]	ρ_{wa}	0.3118 [-0.315, 0.533]	ρ_{ww}	0.8177 [0.074, 0.915]	ρ_{wr}	0.0398 [-0.090, 0.264]
σ_{vr}	0.0657 [0.001, 0.520]	ρ_{ra}	-0.5024 [-0.795, -0.147]	ρ_{wa}	-0.0359 [-0.181, 0.097]	ρ_{rr}	0.1579 [-0.443, 0.408]

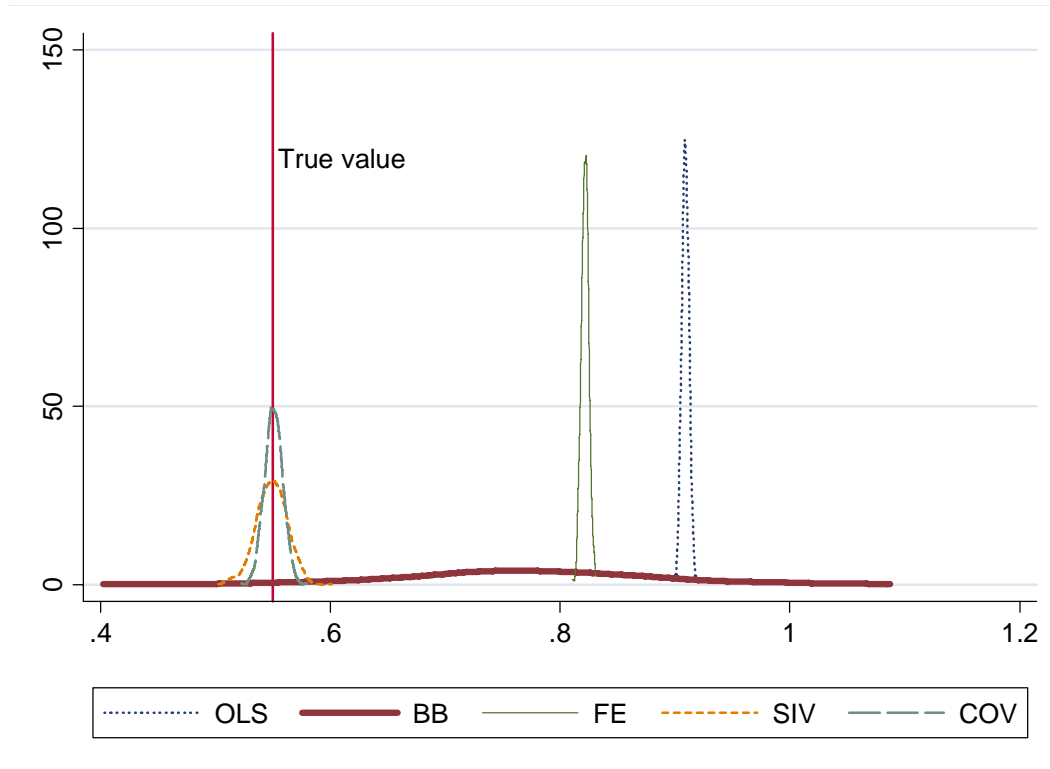
(continued on next page)

Table 7 continued

Measurement errors: standard deviation of innovations and serial correlation			
σ_{ev}	1.862	ρ_v	0.143
	[1.665, 2.008]		[-0.110, 0.439]
σ_{ek}	1.8052	ρ_k	0.9305
	[1.581, 1.972]		[0.903, 0.957]
σ_{kl}	1.5222	ρ_l	0.7609
	[1.331, 1.668]		[0.651, 0.844]

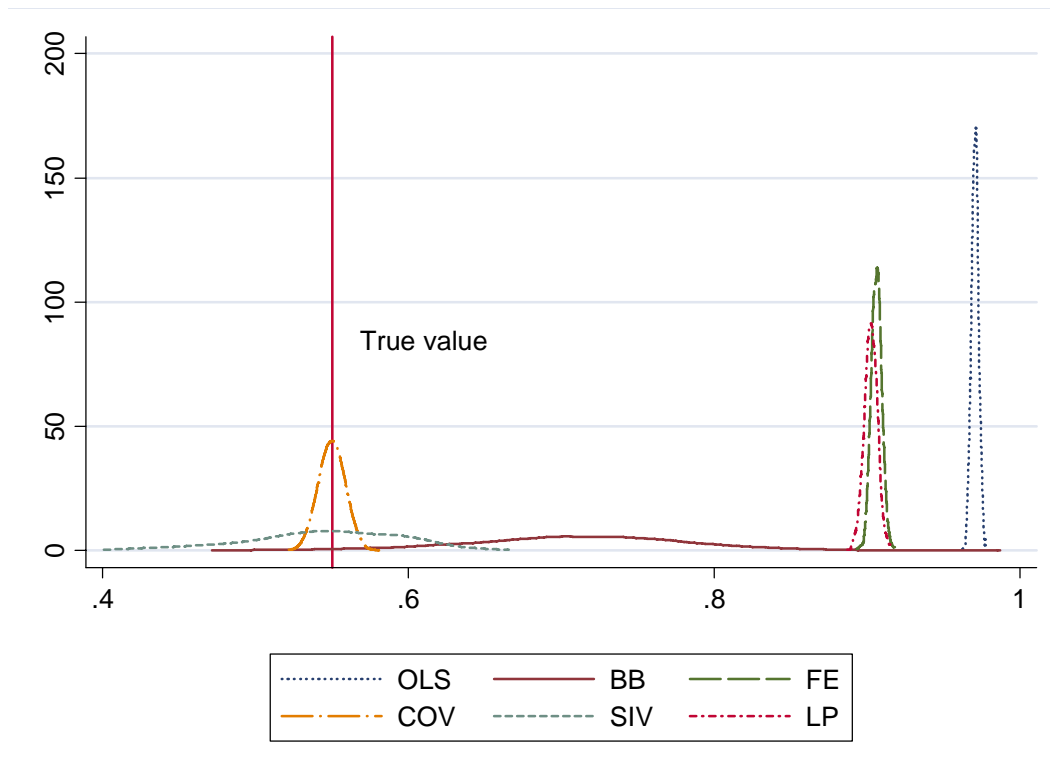
Note: The COV model is described in (27)-(38). 95% bootstrap confidence interval is in square parentheses. FE is fixed effects, LP is Levinsohn-Petrin estimator, BB is Blundell-Bond estimator, SIV is Schmidt's instrumental variables estimator. BB estimator is unrestricted and the reported coefficients are on the current k_{it} and l_{it} . Standard errors are in parentheses. R^2 in OLS regression is 0.92. The LM test does not reject AR(1) model for the error term in the BB estimator.

Figure 1. Kernel density of estimates for returns to scale: One input



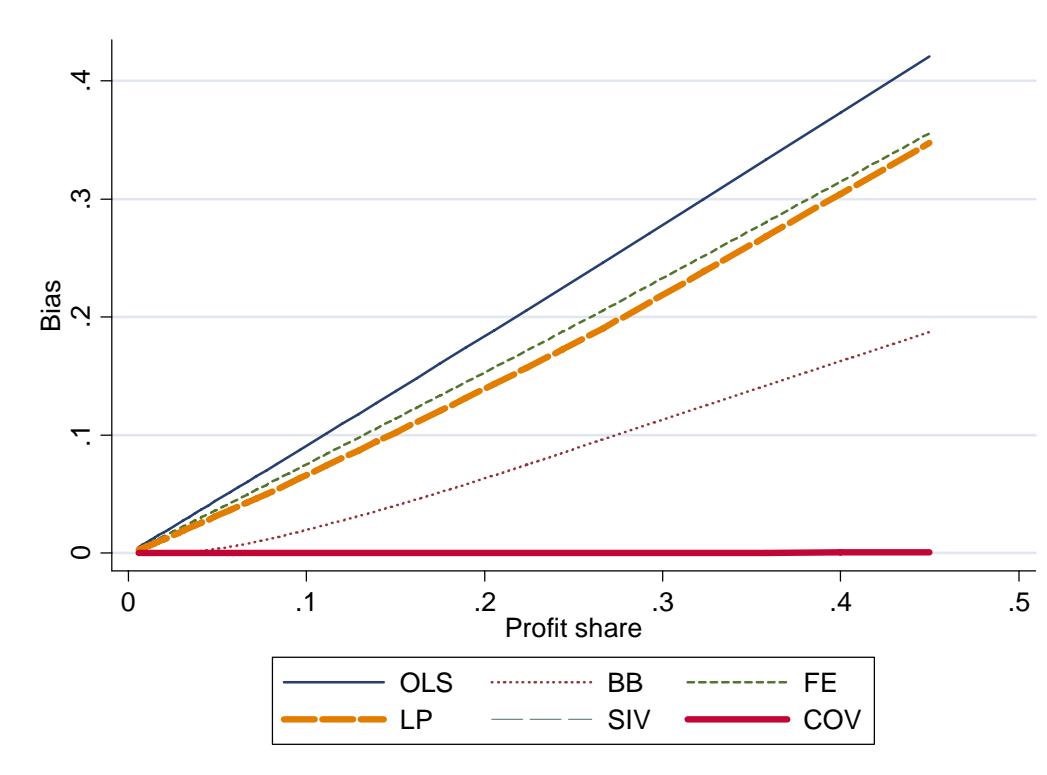
Note: The figure plots Epanechnikov kernel densities of the returns to scale in the revenue function for OLS, Schmidt's instrumental variables (SIV), covariance (COV), fixed effects (FE), and Bond-Blundell (BB) estimators. Parameter values of the data generating process are for the scenario reported in Panel A, Table 2. Returns to scale are on horizontal axis. The data generating process is (12)-(15). The estimated revenue function is (16). Each experiment is simulated 1,000 times. See text and Table 2 for further details.

Figure 2. Kernel density of estimates for returns to scale: Multiple inputs



Note: The figure plots Epanechnikov kernel densities of the returns to scale in the revenue function for OLS, Schmidt's instrumental variables (SIV), covariance (COV), fixed effects (FE), and Bond-Blundell (BB), and Levinsohn-Petrin (LP) estimators. Parameter values of the data generating process are for the scenario reported in Panel A, Table 3. Returns to scale are on horizontal axis. The data generating process is (19)-(22): three inputs and one output. The estimated revenue function is (18). Each experiment is simulated 1,000 times. See text and Table 3 for further details.

Figure 3. Profit share and bias in returns to scale



Note: The figure reports the bias in the estimated returns to scale in the revenue functions for various values of the profit share. The lines are from lowess which smoothes over 100 replications for each value of the profit share. Parameterization is as in Panel A of Table 3. The data generating process is (19)-(22): three inputs and one output. The estimated revenue function is (18). BB is Blundell-Bond estimator, FE is fixed effects, SIV is Schmidt's IV, LP is Levinsohn-Petrin estimator, COV is the covariance estimator. SIV essentially coincides with COV in this figure.

Appendix A: Extensions

Multi-input case

This section presents the n -input analogue for the model considered in Section 2.1. One could derive results as first order approximations but without loss of generality it is convenient to work with specific functional forms. The production function is assumed to be Cobb-Douglas:

$Q_{it} = A_{it}^\mu \prod_{j=1}^n L_{j,it}^{\alpha_j}$ where i, t, j index firms, time, and inputs, $\gamma = \sum_{j=1}^n \alpha_j$ is returns to scale in

production, A_{it} is Hicks-neutral firm-specific productivity, and $L_{j,it}$ is j^{th} input. The inverse demand function is isoelastic $P_{it} = G_{it} \cdot Q_{it}^{-1/\sigma}$ where P_{it} is the price of the good, Q_{it} is the quantity of the good, G is a demand shifter, and σ is the elasticity of demand. The markup is $\mu = \sigma / (\sigma - 1)$. Hence, the revenue function is

$Y_{it} = P_{it} Q_{it} = G_{it} (A_{it}^\mu \prod_{j=1}^n L_{j,it}^{\alpha_j})^{1-1/\sigma} = G_{it} A_{it} \prod_{j=1}^n L_{j,it}^{\beta_j}$, where $\beta_i = \alpha_i / \mu$, and $\eta = \sum_{j=1}^n \beta_j$ is RTS in

the revenue function. Also note that A_{it} and G_{it} are isomorphic in the revenue function so that the econometrician cannot separate these shocks. Hence, I drop G_{it} from the analysis and concentrate

on A_{it} only. The cost for input j is given by $C_j(L_j) = W_{j,it} L_{j,it}^{\phi_j}$ where ϕ_j is the elasticity of the cost of input j . The profit maximization problem is then $\max_{\{L_{1,it}, \dots, L_{n,it}\}} \pi_{it}$ where

$$\pi_{it} \equiv Y_{it} - \sum_{j=1}^n C_j(L_j) = A_{it} \prod_{j=1}^n L_{j,it}^{\beta_j} - \sum_{j=1}^n W_{j,it} L_{j,it}^{\phi_j}.$$

I take logs of the first order conditions, suppress uninteresting constants, partial out industry-wide shocks, and get the following expressions for optimal input choices and revenue

$$\begin{bmatrix} -\phi_1 & 0 & 0 & \cdots & 0 & 1 \\ 0 & -\phi_2 & 0 & \cdots & 0 & 1 \\ 0 & 0 & -\phi_3 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi_n & 1 \\ -\beta_1 & -\beta_2 & -\beta_3 & \cdots & -\beta_n & 1 \end{bmatrix} \begin{bmatrix} l_{1,it} \\ \vdots \\ l_{n,it} \\ y_{it} \end{bmatrix} = \begin{bmatrix} w_{1,it} \\ \vdots \\ w_{n,it} \\ a_{it} \end{bmatrix} \Rightarrow \mathbf{X}_{it} \equiv \begin{bmatrix} l_{1,it} \\ \vdots \\ l_{n,it} \\ y_{it} \end{bmatrix} = \Lambda \begin{bmatrix} w_{1,it} \\ \vdots \\ w_{n,it} \\ a_{it} \end{bmatrix} = \Lambda \mathbf{F}_{it},$$

where

$$\Lambda = \frac{1}{\sum_{j=1}^n \beta_j \psi_{jj} - \prod_{i=1}^n \phi_i} \begin{bmatrix} \psi_{11} - \sum_{i=1}^n \beta_i \psi_{1i} & \psi_{12} \beta_2 & \psi_{13} \beta_3 & \cdots & \psi_{1,n-1} \beta_{n-1} & -\psi_{11} \\ \psi_{21} \beta_1 & \psi_{22} - \sum_{i=2}^n \beta_i \psi_{2i} & \psi_{23} \beta_3 & \cdots & \psi_{2,n-1} \beta_{n-1} & -\psi_{22} \\ \psi_{31} \beta_1 & \psi_{32} \beta_2 & \psi_{33} - \sum_{i=3}^n \beta_i \psi_{3i} & \cdots & \psi_{3,n} \beta_n & -\psi_{33} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \psi_{n,1} \beta_1 & \psi_{n,2} \beta_2 & \psi_{n,3} \beta_3 & \cdots & \psi_{n,n} - \sum_{i=n}^n \beta_i \psi_{n,i} & -\psi_{n,n} \\ \psi_{11} \beta_1 & \psi_{22} \beta_2 & \psi_{33} \beta_3 & \cdots & \psi_{n,n} \beta_n & -\prod_{i=1}^n \phi_i \end{bmatrix}, \quad (39)$$

$\psi_{ij} = \prod_{s \neq i, j} \phi_s$ and $\psi_{ij} = 1$ if $s \neq i, j$ is an empty set. Observe that $\det \Lambda = \sum_{j=1}^n \beta_j \psi_{jj} - \prod_{i=1}^n \phi_i < 0$ which is the necessary and sufficient condition for the profit function to be concave.

One can use information from the first order conditions to compute the cost shares. Observe that for each input j , the first order condition is $\beta_j Y_{it} / L_{j,it} = \phi_j W_{j,it} L_{j,it}^{\phi_j - 1}$. Hence, $C_j(L_j) = W_{j,it} L_{j,it}^{\phi_j} = \beta_j Y_{it} / \phi_j$. It follows that the cost share for input j is given by

$$\omega_j = \frac{C_j(L_j)}{\sum_{h=1}^n C_h(L_h)} = \frac{\beta_j / \phi_j}{\sum_{h=1}^n \beta_h / \phi_h}. \quad (40)$$

The elasticity of the cost with respect to all inputs is

$$\phi = \sum_{h=1}^n \omega_h \phi_h = \frac{\sum_{h=1}^n \beta_h}{\sum_{h=1}^n \beta_h / \phi_h}. \quad (41)$$

Using this expression one can find the profit share in terms of cost and revenue elasticities:

$$s_\pi = 1 - \frac{\gamma}{\mu \phi} = 1 - \sum_{h=1}^n \beta_h / \phi_h. \quad (42)$$

Constant elasticity of substitution (CES) production function

Consider the CES production function $Q = A^\mu (\omega_K K^\rho + \omega_L L^\rho)^{\gamma/\rho}$ where $\frac{1}{1-\rho}$ is the elasticity of substitution. In this example, I assume that productivity and factor prices are mutually uncorrelated. Otherwise the structure is the same as in the Cobb-Douglas case. The profit function is given by: $\pi = A(\omega_K K^\rho + \omega_L L^\rho)^{\gamma/(\mu\rho)} - RK - WL$. The first order conditions with respect to capital and labor are: $\eta s_K Y / K = R$, $\eta s_L Y / L = W$ where $s_K = \omega_K K^\rho / (\omega_K K^\rho + \omega_L L^\rho)$, $s_L = 1 - s_K$. After log-linearizing first-order conditions and the revenue function, one has the following structural equations: $y = \eta s_K k + \eta s_L l + a$, $y - k + \rho s_L (k - l) = r$, $y - l + \rho s_K (l - k) = w$. The reduced form is

$$X = \begin{bmatrix} y \\ k \\ l \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\eta} & -\frac{\eta s_K}{1-\eta} & -\frac{\eta(1-s_K)}{1-\eta} \\ \frac{1}{1-\eta} & -\frac{1-\eta+s_K(\eta-\rho)}{(1-\rho)(1-\eta)} & -\frac{(\eta-\rho)(1-s_K)}{(1-\rho)(1-\eta)} \\ \frac{1}{1-\eta} & -\frac{s_K(\eta-\rho)}{(1-\rho)(1-\eta)} & -\frac{1-\rho-s_K(\eta-\rho)}{(1-\rho)(1-\eta)} \end{bmatrix} \begin{bmatrix} a \\ r \\ w \end{bmatrix}.$$

The model has six parameters: $\theta = \{\eta, s_K, \rho, \sigma_a, \sigma_w, \sigma_r\}$. It is straightforward (but tedious) to show that $\nabla_{\theta} E(XX')$ has full rank and, hence, the model is locally identified almost everywhere.

Rational Expectations models

Following Blanchard and Kahn (1980), one can show that, after log-linearization, rational profit-maximizing producer behavior can be summarized as follows:

$$\mathbf{S}_t \equiv \begin{bmatrix} \mathbf{G}_t \\ \mathbf{H}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} 0 & \Pi_{12} & \Pi_{13} \\ 0 & \Pi_{22} & \Pi_{23} \\ 0 & 0 & \Pi_{33} \end{bmatrix} \begin{bmatrix} \mathbf{G}_{t-1} \\ \mathbf{H}_{t-1} \\ \mathbf{Z}_{t-1} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \mathbf{v}_t = \Pi \mathbf{S}_{t-1} + B \mathbf{v}_t, \quad (43)$$

where \mathbf{G}_t is $px1$ vector of endogenous non-predetermined variables (e.g., materials), \mathbf{H}_t is $mx1$ vector of endogenous predetermined variables (e.g., capital), \mathbf{Z}_t is $nx1$ vector of exogenous variables, \mathbf{v}_t is vector of serially uncorrelated innovations. The econometrician observes only \mathbf{G}_t and \mathbf{H}_t . The number of shocks is not less than the number of observed variables, i.e., $n \geq m + p$. No assumptions are made about the sources of shocks \mathbf{v}_t , which can be shocks to adjustment costs, factor prices, productivity, etc. Hence, the setup is very general.

The autocovariance matrix of the observed variables collected in $\mathbf{X}_t = [\mathbf{G}_t' \ \mathbf{H}_t']' = \Upsilon \mathbf{S}_t$ with $\Upsilon \equiv [I_{p+m} \ | \ 0]$ is given by $\tilde{\Gamma}_k \equiv E(\mathbf{X}_t \mathbf{X}_{t-k}') = \Upsilon \Pi^k \Gamma_0 \Upsilon'$, $k=0,1,\dots$, where $\Gamma_0 \equiv E(\mathbf{S}_t \mathbf{S}_t')$, $\text{vec}(\Gamma_0) = (I_{m+p+n} - \Pi \otimes \Pi)^{-1} \text{vec}(B \Omega B')$, and $\Omega \equiv E(\mathbf{v}_t \mathbf{v}_t')$. Given that $\tilde{\Gamma}_k$, $k=0,1,\dots$, and matrices Π, B are deterministic one-to-one functions of structural parameters, one can use methods presented in section 3.4 to set up likelihood function for MLE. Specifically, the log-likelihood function for the no-firm-specific-effects and no-measurement-error case is given by $\sum_{i=1}^N l(\mathbf{X}_i, \theta) = \ln |\Phi_T| + \text{trace}\{\hat{\Phi}_T \Phi_T^{-1}\} - \ln |\hat{\Phi}_T| - Tn$ where $\hat{\Phi}_T = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i'$, n is the number of observed choices of firms, and

$$\Phi_T \equiv \begin{bmatrix} \tilde{\Gamma}_0 & & & \\ \tilde{\Gamma}_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ \tilde{\Gamma}_T & \cdots & \tilde{\Gamma}_1 & \tilde{\Gamma}_0 \end{bmatrix} = \begin{bmatrix} \Upsilon \Gamma_0 \Upsilon' & & & \\ \Upsilon \Pi \Gamma_0 \Upsilon' & \ddots & & \\ \vdots & \ddots & \ddots & \\ \Upsilon \Pi^T \Gamma_0 \Upsilon' & \cdots & \Upsilon \Pi \Gamma_0 \Upsilon' & \Upsilon \Gamma_0 \Upsilon' \end{bmatrix}.$$

The likelihood can be easily extended to cases with measurement errors and firm-specific effects.

Note that model (43) is highly nonlinear in structural parameters. Hence, global identification is hard to prove in general. Local identification is easy to verify (numerically) by checking the rank of the relevant Jacobian.

Identification in models with adjustment costs

In this section I show that, under certain assumptions, the Blundell-Bond estimator is not identified even in the presence of adjustment costs. In the spirit to the results in Section 4.2, poor identification arises because profit maximization imposes restrictions on the dynamic and contemporaneous comovement of inputs and output. The following proposition gives the *necessary* condition for identification of the BB estimator for any rational expectation model described by (43).

Proposition 6

Consider rational profit-maximizing firm characterized by the reduced-form dynamics as in (43). Then if $n < 2p+m-1$, the unconstrained Blundell-Bond estimator is not identified.

Proof: see Appendix B.

Profit maximization can impose further restrictions on the dynamic and contemporaneous correlation between variables so that the estimator is not identified even when the presented *necessary* condition is satisfied. The following proposition provides an important example where BB is not identified although the necessary condition is satisfied.

Proposition 7

Consider rational profit-maximizing firm characterized by the reduced-form dynamics as in (43). Suppose that Π_{33} is diagonal and that output and one of the inputs are free to adjust contemporaneously in response to shocks. Then, irrespective of the number of the structural shocks, the unconstrained Blundell-Bond estimator is not identified to a first-order log-linear approximation.

Proof: see Appendix B.

If production function is Cobb-Douglas, then log-linear approximation is exact and, thus, even higher order approximation cannot identify the parameters. Using the argument of Proposition 4 it is straightforward to show that even when BB is locally identified, there could be several solutions.

Appendix B: Proofs

Proof of Proposition 1.

Consider cost minimization problem, which is implied by profit maximization: $L \equiv (L_1, \dots, L_n) = \arg \min_L \{ \sum_{j=1}^n w_j(L_j) : f(L) = Q \}$, where f is the production function, $w_j(L_j)$ is the cost of input L_j , Q is output.

The first order condition gives $w'_j(L_j) = \lambda f'_j(L)$ for $j = 1, \dots, n$ where λ is the Lagrange multiplier and f is the production function. Multiply both sides by L_j for each j and sum over j to get

$$\begin{aligned} \sum_{j=1}^n w'_j(L_j) L_j &= \lambda \sum_{j=1}^n f'_j(L) L_j && \Leftrightarrow \\ Q \cdot AC(Q) \cdot \sum_{j=1}^n \frac{w'_j(L_j) L_j}{w_j(L_j)} \frac{w_j(L_j)}{TC(Q)} &= MC(Q) \cdot \sum_{j=1}^n f'_j(L) L_j && \Leftrightarrow \\ \frac{AC(Q)}{MC(Q)} \left(\sum_{j=1}^n \phi_j \omega_j \right) &= \frac{1}{Q} \sum_{j=1}^n f'_j(L) L_j && \Leftrightarrow \text{(by Euler's theorem)} \frac{AC(Q)}{MC(Q)} \phi = \gamma, \end{aligned}$$

where ϕ_j is the elasticity of j^{th} factor price, ω_j is the share of factor i in total cost $TC(Q)$, $\phi = \sum_{j=1}^n \phi_j \omega_j$ is the elasticity of the cost with respect to inputs, $AC(Q)$ and $MC(Q)$ are average and marginal costs. If factor markets are competitive, $\phi_j = 1$ for all j and hence $\phi = 1$. Now observe that profit share is equal to $s_\pi = \frac{PQ - AC(Q) \cdot Q}{PQ} = 1 - \frac{AC(Q)}{P}$. It follows that $\gamma = \phi AC(Q)/MC(Q) = \phi(1 - s_\pi)\mu \Leftrightarrow \gamma/\mu\phi = (1 - s_\pi)$, where $\mu = P/MC(Q)$ is the markup.

Since marginal revenue (MR) is equal to marginal cost for a profit-maximizing firm, one has $\partial TR/\partial L_j = MR(\partial Q/\partial L_j) \Rightarrow \sum_{j=1}^n (\partial TR/\partial L_j) L_j = MC \sum_{j=1}^n (\partial Q/\partial L_j) L_j = (MC/P)\gamma PQ$ and

$$\text{hence } \eta = \frac{\sum_{j=1}^n (\partial TR/\partial L_j) L_j}{PQ} = \frac{\gamma}{\mu}. \blacksquare$$

Proof of Corollary 1.

Consider cost minimization problem:

$$L \equiv (L_1, \dots, L_k, \bar{L}_{k+1}, \dots, \bar{L}_n) = \arg \min_{L_1, \dots, L_k} \{ \sum_{j=1}^n w_j(L_j) : f(L) = Q \}, \text{ where } w_j(L_j) \text{ is the cost of input}$$

L_j , f is the production function and Q is output and $k+1, \dots, n$ inputs are fixed. Using the arguments of Proposition 1, one can show that the first order condition with respect to variable $w'_j(L_j) = \lambda f'_j(L)$ for $j = 1, \dots, k$ (λ is the Lagrange multiplier) yield:

$$\sum_{j=1}^k w'_j(L_j)L_j = \lambda \sum_{j=1}^k f'_j(L)L_j \Leftrightarrow$$

$$\frac{AC(Q)}{MC(Q)} \frac{\sum_{i=1}^k w_i(L_i)L_i}{TC(Q)} \left(\sum_{j=1}^k \phi_j \omega_j \right) = \frac{1}{Q} \sum_{j=1}^k f'_j(L)L_j \Leftrightarrow \frac{AC(Q)}{MC(Q)} \phi^* \omega^* = \gamma^*,$$

where ω^* is the cost share of variable inputs in total cost, $\phi^* = \sum_{j=1}^k \phi_j \omega_j$ is the elasticity of cost with respect to variable inputs, γ^* is RTS in production with respect to variable inputs, $AC(Q)$ and $MC(Q)$ are average and marginal costs. Now observe that profit share is equal to $s_\pi^* = 1 - AC(Q)/P$. It follows that

$$\frac{AC(Q)}{MC(Q)} \phi^* \omega^* = \gamma^* \Leftrightarrow \frac{AC(Q)}{P} \frac{P}{MC(Q)} \phi^* \omega^* = \gamma^* \Leftrightarrow (1 - s_\pi^*) \mu \phi^* \omega^* = \gamma^*$$

where $\mu = P/MC(Q)$ is the markup.

Since marginal revenue (MR) is equal to marginal cost for a profit-maximizing firm, one has

$$\begin{aligned} \partial TR / \partial L_j &= MR(\partial Q / \partial L_j) \Rightarrow \\ \sum_{j=1}^k (\partial TR / \partial L_j) L_j &= MC \sum_{j=1}^k (\partial Q / \partial L_j) L_j = (MC/P) \hat{\gamma} P Q \Rightarrow \\ \eta^* &= \frac{\sum_{j=1}^k (\partial TR / \partial L_j) L_j}{P Q} = \frac{\gamma^*}{\mu}, \end{aligned}$$

where η^* is RTS in the revenue function with respect to variable inputs. By combining the results, one can find: $\eta^* = \gamma^* / \mu = (1 - s_\pi^*) \phi^* \omega^*$. ■

Proof of Proposition 2

Without loss of generality assume that there are two inputs and one output, the first input is supplied in a competitive market. Suppose there are two solutions θ and $\tilde{\theta}$. To satisfy restrictions imposed by profit maximization, the matrix $\tilde{\Lambda}$ must possess the same structure and properties as Λ .

Because $\Lambda, \tilde{\Lambda}, T$ are invertible, $\Lambda = \tilde{\Lambda} T^{-1}$ implies that

$$\begin{aligned} T &= \Lambda^{-1} \tilde{\Lambda} \\ &= \frac{1}{\tilde{\beta}_1 \tilde{\phi}_2 + \tilde{\beta}_2 - \tilde{\phi}_2} \begin{bmatrix} \tilde{\beta}_1 \tilde{\phi}_2 + \tilde{\beta}_2 - \tilde{\phi}_2 & 0 & 0 \\ \tilde{\beta}_1 (\tilde{\phi}_2 - \phi_2) & \tilde{\beta}_2 - \phi_2 (1 - \tilde{\beta}_1) & \phi_2 - \tilde{\phi}_2 \\ -\beta_1 (\tilde{\phi}_1 - \tilde{\beta}_1) - \beta_2 \tilde{\beta}_1 + \tilde{\phi}_2 \tilde{\beta}_1 & -\beta_1 \tilde{\beta}_2 - \beta_2 (1 - \tilde{\beta}_1) + \beta_1 & \beta_2 - \tilde{\phi}_2 (1 - \beta_1) \end{bmatrix} \quad (44) \end{aligned}$$

Note that $\det(T) = (\beta_1 \phi_2 + \beta_2 - \phi_2) / (\tilde{\beta}_1 \tilde{\phi}_2 + \tilde{\beta}_2 - \tilde{\phi}_2) \neq 0$ and the solution $\tilde{\theta}$ must have $\tilde{\beta}_1 \tilde{\phi}_2 + \tilde{\beta}_2 - \tilde{\phi}_2 < 0$. Thus, the model is not identified unless further restrictions are imposed.

Now consider

$$\Omega = T\tilde{\Omega}T' = \frac{1}{(\tilde{\beta}_1\tilde{\phi}_2 + \tilde{\beta}_2 - \tilde{\phi}_2)^2} \begin{bmatrix} D_{11} & & \\ D_{21} & D_{22} & \\ D_{31} & D_{32} & D_{33} \end{bmatrix},$$

where D_{11} , D_{22} , D_{33} are positive quantities and

$$\begin{aligned} D_{21} &= (\tilde{\beta}_2 + \tilde{\beta}_1\tilde{\phi}_2 - \tilde{\phi}_2)\tilde{\beta}_1(\tilde{\phi}_2 - \phi_2)\tilde{\sigma}_{11}, \quad D_{31} = (\tilde{\beta}_2 + \tilde{\beta}_1\tilde{\phi}_2 - \tilde{\phi}_2)[- \beta_1(\tilde{\phi}_2 - \tilde{\beta}_2) + \tilde{\beta}_1(\tilde{\phi}_2 - \beta_2)]\tilde{\sigma}_{11}, \\ D_{32} &= \tilde{\beta}_1(\tilde{\phi}_2 - \phi_2)[\tilde{\beta}_1(\tilde{\phi}_2 - \beta_2) - \beta_1(\tilde{\phi}_2 - \tilde{\beta}_2)]\tilde{\sigma}_{11} + [\tilde{\beta}_2 - \phi_2(1 - \tilde{\beta}_1)][\tilde{\beta}_2 - \beta_1\tilde{\beta}_2 - \beta_2(1 - \tilde{\beta}_1)]\tilde{\sigma}_{22} + \\ &\quad + (\tilde{\phi}_2 - \phi_2)(\beta_1\tilde{\phi}_2 + \beta_2 - \tilde{\phi}_2)\tilde{\sigma}_{33}. \end{aligned}$$

The restriction that Ω is diagonal implies that $D_{21} = D_{31} = D_{32} = 0$. From $D_{21}=0$ it follows that $\tilde{\phi}_2 = \phi_2$ since $\tilde{\beta}_1 \neq 0$. Hence, $D_{31} = D_{32} = 0$ implies that

$$\tilde{\beta}_1(\phi_2 - \beta_2) - \beta_1(\phi_2 - \tilde{\beta}_2) = 0 \quad (45)$$

$$\tilde{\beta}_2(1 - \beta_1) - \beta_2(1 - \tilde{\beta}_1) = 0 \quad (46)$$

The only solution to this system of equations is $\beta_1 = \tilde{\beta}_1$ and $\beta_2 = \tilde{\beta}_2$ implying that $T=I$ and, thus, the model is uniquely globally identified.

For a general model with a productivity shock and n inputs and associated factor prices, the first entry of the first row of T in (44) will continue to be non-zero while other entries of the row are zeros. This fixes $\tilde{\phi}_j = \phi_j$ for $j=2, \dots, n$ and then it is an easy step to show that n -input analogue of (45)-(46) has unique solution $\tilde{\beta}_j = \beta_j$ for $j=1, \dots, n$. This proves part *a*.

To prove part *b*, again, without loss of generality, assume that there are two inputs and one output and that the first input is supplied in a competitive market. Suppose there are two solutions θ and $\tilde{\theta}$. Then by assumptions of the proposition, the following matrix must be diagonal

$$\tilde{\Pi} = T^{-1}\Pi T = \frac{1}{|T| \|\tilde{\Lambda}\|} \begin{bmatrix} D_{11} & 0 & 0 \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix}$$

where D_{11} , D_{22} , D_{33} are non-zero quantities and

$$\begin{aligned} D_{21} &= (\phi_2 - \tilde{\phi}_2)[\tilde{\Lambda} | \beta_1\Pi_{11} - (\beta_1\tilde{\phi}_2 + \beta_2 - \tilde{\phi}_2)\tilde{\beta}_1\Pi_{22} - (\tilde{\beta}_1(\tilde{\phi}_2 - \beta_2) - \beta_1(\tilde{\phi}_2 - \tilde{\beta}_2))\Pi_{33}], \\ D_{31} &= \tilde{\Lambda} | [\beta_1(\phi_2 - \tilde{\beta}_2) - \tilde{\beta}_1(\phi_2 - \beta_2)]\Pi_{11} - [\tilde{\beta}_1(1 - \beta_2) - \beta_1(1 - \tilde{\beta}_2)]\tilde{\beta}_1(\tilde{\phi}_2 - \phi_2)\Pi_{22} + \\ &\quad + (\tilde{\beta}_1\phi_2 + \tilde{\beta}_2 - \phi_2)[\tilde{\beta}_1(\tilde{\phi}_2 - \beta_2) - \beta_1(\tilde{\phi}_2 - \tilde{\beta}_2)]\Pi_{33}, \\ D_{32} &= [\tilde{\beta}_2(1 - \beta_1) - \beta_2(1 - \tilde{\beta}_1)](\tilde{\beta}_1\phi_2 + \tilde{\beta}_2 - \phi_2)(\Pi_{33} - \Pi_{22}), \\ D_{23} &= (\beta_1\tilde{\phi}_2 + \beta_2 - \tilde{\phi}_2)(\phi_2 - \tilde{\phi}_2)(\Pi_{22} - \Pi_{33}). \end{aligned}$$

The restriction that $T^{-1}\Pi T$ is diagonal, implies that $D_{21} = D_{23} = D_{31} = D_{32} = 0$. Suppose that $\Pi_{22} \neq \Pi_{33}$. From $D_{23}=0$, $(\phi_2 - \tilde{\phi}_2)(\beta_1\tilde{\phi}_2 + \beta_2 - \tilde{\phi}_2) = 0$. Suppose that $\phi_2 = \tilde{\phi}_2$. Then $D_{21}=0$ and $D_{31}=D_{32}=0$ imply that

$$\beta_1(\phi_2 - \tilde{\beta}_2) - \tilde{\beta}_1(\phi_2 - \beta_2) = 0 \quad (47)$$

$$\tilde{\beta}_2(1 - \beta_1) - \beta_2(1 - \tilde{\beta}_1) = 0 \quad (48)$$

provided that $\tilde{\Pi}_{33} - \tilde{\Pi}_{11} \neq 0$. The only solution to (47) and (48) is $\beta_1 = \tilde{\beta}_1$ and $\beta_2 = \tilde{\beta}_2$ implying that $T=I$ and, thus, the model is uniquely globally identified almost everywhere.

Now suppose that $\phi_2 \neq \tilde{\phi}_2$ so that $\beta_1\tilde{\phi}_2 + \beta_2 - \tilde{\phi}_2 = 0 \Leftrightarrow \tilde{\phi}_2 = \beta_2/(1 - \beta_1)$. Suppose that $\tilde{\beta}_2(1 - \beta_1) - \beta_2(1 - \tilde{\beta}_1) = 0 \Leftrightarrow \tilde{\beta}_2 = \beta_2(1 - \tilde{\beta}_1)/(1 - \beta_1)$ from $D_{32}=0$. Substitute $\tilde{\phi}_2, \tilde{\beta}_2$ in $D_{21}=0$ and reach the contradiction that $\tilde{\beta}_2 = 0$. Now suppose that $\tilde{\beta}_1\phi_2 + \tilde{\beta}_2 - \phi_2 = 0 \Leftrightarrow \tilde{\beta}_2 = \phi_2(1 - \tilde{\beta}_1)$ from $D_{32}=0$. Substitute $\tilde{\phi}_2, \tilde{\beta}_2$ into $|\tilde{\Lambda}|$ and find that $|\tilde{\Lambda}| = 0$ which contradicts $|T| \neq 0$. Hence, $\phi_2 \neq \tilde{\phi}_2$ leads to contradiction.

For a general case with n inputs, one again uses the fact that $\phi_1 = 1$ to fix $\tilde{\phi}_j = \phi_j$ for $j=2, \dots, n$ and then it is a tedious but straightforward step to show that n -input analogue of (47)-(48) has unique solution is $\tilde{\beta}_j = \beta_j$ for $j=1, \dots, n$ almost everywhere. This proves part *b*. ■

Proof of Proposition 3.

Under assumptions of the proposition, system (8)-(9) can be re-formulated as follows:

$$\mathbf{X}_{it} = [\Lambda \mid B] \begin{bmatrix} \mathbf{F}_{it} \\ \mathbf{M}_{it} \end{bmatrix} + \boldsymbol{\varepsilon}_{it}, \quad (49)$$

$$\begin{bmatrix} \mathbf{F}_{it} \\ \mathbf{M}_{it} \end{bmatrix} = \begin{bmatrix} \Pi & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \mathbf{F}_{i,t-1} \\ \mathbf{M}_{i,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_{it} \\ \boldsymbol{\omega}_{it} \end{bmatrix}, \quad (50)$$

where B is the known matrix whose columns are selection vectors e_i (that is e_i is the i^{th} column of matrix I_n) with unity in the row corresponding to the variable with a serially correlated measurement error, \mathbf{M}_{it} is a vector of measurement errors, R is a diagonal nonsingular matrix with entries less than unity in absolute value (stationarity of measurement errors), $E(\boldsymbol{\omega}_{it}\boldsymbol{\omega}'_{it}) = \Omega_1$ is a diagonal nonsingular matrix, $E(\boldsymbol{\omega}_{it}\boldsymbol{\omega}'_{js}) = 0$ for all i, j and $s \neq 0$, and $E(\boldsymbol{\varepsilon}_{it}\boldsymbol{\omega}'_{js}) = E(\mathbf{v}_{it}\boldsymbol{\omega}'_{js}) = 0$ for all i, j, s .

To prove global identification, it is sufficient to show that there is no rotation matrix T that preserves the structure of the model. Suppose that such T exists. Then a rotationally equivalent solution must satisfy

$$[\tilde{\Lambda} \mid B] = [\Lambda \mid B] \left[\begin{array}{c|c} T_{11} & T_{12} \\ \hline T_{21} & T_{22} \end{array} \right] = [\Lambda T_{11} + BT_{21} \mid \Lambda T_{12} + BT_{22}].$$

Hence,

$$\tilde{\Lambda} = \Lambda T_{11} + BT_{21} \Leftrightarrow T_{11} = \Lambda^{-1} \tilde{\Lambda} - \Lambda^{-1} BT_{21} \quad (51)$$

$$B = \Lambda T_{12} + BT_{22} \Leftrightarrow T_{12} = \Lambda^{-1} B (I - T_{22}) \quad (52)$$

Furthermore, there are nonlinear restrictions imposed by uncorrelatedness of \mathbf{v}_t and $\boldsymbol{\omega}_t$ and

block diagonal structure of $\left[\begin{array}{c|c} \Pi & 0 \\ \hline 0 & R \end{array} \right]$. In particular,

$$n \times k \text{ restrictions: } T_{21} \Omega T'_{11} + T_{22} \Omega_1 T'_{12} = 0 \quad (53)$$

$$\frac{1}{2} k(k-1) \text{ restrictions: } T_{21} \Omega T'_{21} + T_{22} \Omega_1 T'_{22} \text{ is a diagonal matrix} \quad (54)$$

$$n \times k \text{ restrictions: } \Pi T_{12} - T_{12} T_{22}^{-1} R T_{22} = 0 \quad (55)$$

$$n \times k \text{ restrictions: } T_{21} (T_{11} - T_{12} T_{22}^{-1} T_{21})^{-1} \Pi (T_{11} - T_{12} T_{22}^{-1} T_{21}) = R T_{21} \quad (56)$$

$$\text{one restriction: } \det |T_{22}| \neq 0 \quad (57)$$

$$\text{one restriction: } \det |T_{11}| \neq 0 \quad (58)$$

Because matrices Λ, Π, B, R have full rank, (51)-(58) form an overidentified system of quadratic equations. It is easy to verify that $T_{11} = I_n, T_{12} = 0, T_{21} = 0, T_{22} = I_k$ is a solution to the system for any Ω, Ω_1, Π, R . It is straightforward to verify for low dimensional systems (i.e., $k \leq n \leq 3$) that $T_{12} = 0, T_{21} = 0, T_{22} = I_k$ is the unique real solution. For higher dimensional cases, consider the worst case when $n=k$ so that $B=I$. Substitute T_{12} from (52) into (55) so that after rearranging terms

$$A \bar{T}_{22} = \bar{T}_{22} R \quad (59)$$

where $A \equiv \Lambda \Pi \Lambda^{-1}$ and $\bar{T}_{22} \equiv (I - T_{22}^{-1})$. Note that A has full rank. It is convenient to treat matrices as linear operators. Note that the space X on which a linear operator Q is defined is given by $X = \text{Im}(Q) \oplus \text{Ker}(Q)$, where $\text{Im}(Q)$ is the image of Q and $\text{Ker}(Q)$ is the core of Q .

Suppose that $x \in \text{Ker}(\bar{T}_{22})$. Then by the definition of the core, $0 = A \bar{T}_{22} x = \bar{T}_{22} R x \Rightarrow R x \in \text{Ker}(\bar{T}_{22}) \Rightarrow R(\text{Ker}(\bar{T}_{22})) \subset \text{Ker}(\bar{T}_{22})$. Since R is invertible by assumption, $R(\text{Ker}(\bar{T}_{22})) = \text{Ker}(\bar{T}_{22})$ and, consequently, $R(\text{Im}(\bar{T}_{22})) = \text{Im}(\bar{T}_{22})$ from the orthogonal decomposition.

On the other hand, for $x \in \text{Im}(\bar{T}_{22})$ one has $\bar{T}_{22} x = y \in \text{Im}(\bar{T}_{22})$. Then $Ay = \bar{T}_{22} z, Rx = z \in \text{Im}(\bar{T}_{22}) \Rightarrow Ay \in \text{Im}(\bar{T}_{22})$. Since A is invertible, it follows that $A(\text{Im}(\bar{T}_{22})) = \text{Im}(\bar{T}_{22})$.

Note that by the properties of the core, the operator T_{22} on $\text{Ker}(\bar{T}_{22})$ is equal to I . Consider equation (59) on $\text{Im}(\bar{T}_{22})$. As has been established, there is a \bar{T}_{22}^{-1} operator defined on $\text{Im}(\bar{T}_{22})$. Apply this operator to both sides of (59) and get $R = \bar{T}_{22}^{-1}A\bar{T}_{22}$. Suppose that the dimension of $\text{Im}(\bar{T}_{22})$ is at least one. In this case, operators A and R are equal on $\text{Im}(\bar{T}_{22})$ (Korn and Korn, 1968, §14.6-2). The measure of this event, however, is zero. Hence, almost everywhere the dimension of $\text{Im}(\bar{T}_{22})$ is zero, $\text{Ker}(\bar{T}_{22}) = X$ and $T_{22} = I$ on X . Since Ω has full rank and Ω_1 is diagonal, it follows that $T_{12} = I$ and $T_{21} = 0$ and, hence, $T_{12} = 0, T_{21} = 0, T_{22} = I_k$ is the unique solution almost everywhere.

Restrictions on T_{22} , T_{21} and T_{12} do not pin down the matrix T_{11} . Even if $T_{12} = 0, T_{21} = 0, T_{22} = I_k$, $T_{11} = \Lambda^{-1}\tilde{\Lambda}$ and, therefore, the model is identified if model (8)-(9) is uniquely identified. The “only if” direction follows trivially. ■

Proof of Proposition 4.

Without loss of generality, consider the system without firm specific effects and measurement error ε_{it} . Collect inputs in vector \mathbf{L}_{it} and partition matrix $\Lambda = [\Lambda'_1 \Lambda'_2]'$ so that Λ_1 and Λ_2 correspond to \mathbf{L}_{it} and y_{it} respectively. The residual of the quasi-differenced production (revenue) function is

$$\mathcal{G}_{it} = y_{it} - \hat{\rho}y_{i,t-1} - \hat{b}\mathbf{L}_{it} + \hat{\rho}\hat{b}\mathbf{L}_{i,t-1} = (\Lambda_2 - \hat{b}\Lambda_1)\mathbf{v}_{it} + (\Lambda_2 - \hat{b}\Lambda_1)(\Pi - \hat{\rho}I)\mathbf{F}_{i,t-1},$$

where $\hat{\rho}, \hat{b}$ are “candidate” parameter values for the serial correlation of technology and elasticities of output with respect to inputs. This residual is orthogonal to inputs and output lagged two or more periods if and only if $(\Lambda_2 - \hat{b}\Lambda_1)(\Pi - \hat{\rho}I) = 0$ (\mathbf{F}_{it} is serially correlated while \mathbf{v}_{it} is not).

Note that \hat{b} is a $1 \times (n-1)$ vector and $\hat{\rho}$ is a scalar. Hence, both $\Lambda_2 - \hat{b}\Lambda_1 = 0$ and $\Pi - \hat{\rho}I = 0$ are overidentified because each system has n equations. However, some rows of $\Pi - \hat{\rho}I$ can be non-zero when the corresponding columns of $\Lambda_2 - \hat{b}\Lambda_1$ are equal to zero and vice versa.

Consider first a simple case where the matrix Π is diagonal. If $\hat{\rho}$ is equal to Π_{jj} , one of the diagonal entries of Π , one of the equations in $\Lambda_2 - \hat{b}\Lambda_1 = 0$ can be eliminated, the system becomes just identified and $\hat{b} = \Lambda_{1j}^{-1}\Lambda_{2j}$ where Λ_{ij} is the matrix Λ_i without the j^{th} column. The Blundell-Bond estimator assumes that the $\hat{\rho}$ is equal to the autocorrelation coefficient for productivity ρ_a so that $\hat{b} = \beta$. However, there are other solutions. For example, the above logic suggests that $\hat{\rho}$ can be equal to the autocorrelation coefficient for wage shocks ρ_w and this

choice of $\hat{\rho}$ gives a different solution for \hat{b} . It is straightforward to verify that these solutions are locally identified, i.e., the rank of the Jacobian is full:

$$\text{rank} \left\{ E \left(\begin{bmatrix} -y_{i,t-1} + bL_{i,t-1} \\ -L_{it} + \rho L_{i,t-1} \end{bmatrix} \mathbf{S}'_{it} \right) \right\}_{\rho=a_{jj}, b=\tilde{\Lambda}_1^{-1}\tilde{\Lambda}_2} = \text{rank} \left\{ \begin{bmatrix} \Lambda_2 - \tilde{\Lambda}_1^{-1}\tilde{\Lambda}_2\Lambda_1 \\ \Lambda_1(\Pi - \Pi_{jj}I) \end{bmatrix} E(\mathbf{F}_{i,t-1}\mathbf{S}'_{it}) \right\} = n,$$

where \mathbf{S}_{it} is the vector of appropriately transformed lags of right hand side variables. It follows that there can be n different solutions to $(\Lambda_2 - \hat{b}\Lambda_1)(\Pi - \hat{\rho}I) = 0$ for the case with n inputs.

Now suppose that Π is not diagonal. Let $\hat{\rho}$ be equal to an eigenvalue of Π . Then $\text{rank}(\Pi - \hat{\rho}I) = n - 1$ and, thus, one is back to the case with a diagonal Π , i.e., multiply $\Lambda_2 - \hat{b}\Lambda_1$ by a singular matrix. Hence, for each eigenvalue of Π there is a unique locally-identified solution for \hat{b} . Since Π can have n distinct eigenvalues (for $n-1$ inputs), there can be n solutions for \hat{b} .

To prove the last result, note that if $\hat{\rho}$ is equal to a repeated eigenvalue, the rank of $(\Pi - \hat{\rho}I)$ is at most $n-2$. Hence, at least two columns in Λ_1, Λ_2 can be deleted and $\tilde{\Lambda}_2 - \hat{b}\tilde{\Lambda}_1 = 0$ is underidentified so that there are infinitely many solutions for \hat{b} . ■

Proof of Proposition 5.

This proof is for the case with multiple inputs which are collected in the vector \mathbf{L}_{it} . Partition matrix Λ so that $\begin{bmatrix} \mathbf{L}_{it} \\ \mathbf{Y}_{it} \end{bmatrix} = \Lambda \mathbf{F}_{it} + \begin{bmatrix} \bar{\mathbf{L}}_i \\ \bar{\mathbf{Y}}_i \end{bmatrix} + \boldsymbol{\varepsilon}_{it} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{w}_{it} \\ a_{it} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{L}}_i \\ \bar{\mathbf{Y}}_i \end{bmatrix} + \boldsymbol{\varepsilon}_{it}$, where $\bar{\mathbf{L}}_i, \bar{\mathbf{Y}}_i$ are time invariant effects. For convenience, I define $E(\mathbf{F}_{it}\mathbf{F}'_{it}) = \Sigma$. It is sufficient to show that the rank of the Jacobian for the moment conditions does not have full rank, i.e., the rank of the Jacobian is smaller than the number of parameters in the model. Define $\mathcal{G}_{it} \equiv \mathbf{Y}_{it} - \gamma[\mathbf{L}'_{it} \quad \mathbf{L}'_{i,t-1} \quad \mathbf{Y}'_{i,t-1}]' = \mathbf{Y}_{it} - \gamma\mathbf{V}_{it}$, which corresponds to the residual from quasi-differenced production (revenue) function. Apart from having a permanent component, the error term \mathcal{G}_{it} has MA(1) structure because of the error term $\boldsymbol{\varepsilon}_{it}$.

Consider the level moments $E(\mathcal{G}_{it}\mathbf{S}'_{it}) = 0$ where $\mathbf{S}_{it} = [\Delta\mathbf{L}'_{i,t-2} \quad \dots \quad \Delta\mathbf{L}'_{i,t-k} \quad \Delta\mathbf{Y}'_{i,t-2} \quad \dots \quad \Delta\mathbf{Y}'_{i,t-p}]'$. The expected value of the Jacobian of the moment conditions is

$$E(-\mathbf{V}'_{it}\mathbf{S}'_{it}) = E \left(\begin{array}{ccc|ccc} \mathbf{L}_{it}(\mathbf{L}_{i,t-3} - \mathbf{L}_{i,t-2})' & \dots & \mathbf{L}_{it}(\mathbf{L}_{i,t-k+1} - \mathbf{L}_{i,t-k})' & \mathbf{L}_{it}(\mathbf{Y}_{i,t-3} - \mathbf{Y}_{i,t-2})' & \dots & \mathbf{L}_{it}(\mathbf{Y}_{i,t-p+1} - \mathbf{Y}_{i,t-p})' \\ \mathbf{L}_{i,t-1}(\mathbf{L}_{i,t-3} - \mathbf{L}_{i,t-2})' & \dots & \mathbf{L}_{i,t-1}(\mathbf{L}_{i,t-k+1} - \mathbf{L}_{i,t-k})' & \mathbf{L}_{i,t-1}(\mathbf{Y}_{i,t-3} - \mathbf{Y}_{i,t-2})' & \dots & \mathbf{L}_{i,t-1}(\mathbf{Y}_{i,t-p+1} - \mathbf{Y}_{i,t-p})' \\ \mathbf{Y}_{i,t-1}(\mathbf{L}_{i,t-3} - \mathbf{L}_{i,t-2})' & \dots & \mathbf{Y}_{i,t-1}(\mathbf{L}_{i,t-k+1} - \mathbf{L}_{i,t-k})' & \mathbf{Y}_{i,t-1}(\mathbf{Y}_{i,t-3} - \mathbf{Y}_{i,t-2})' & \dots & \mathbf{Y}_{i,t-1}(\mathbf{Y}_{i,t-p+1} - \mathbf{Y}_{i,t-p})' \end{array} \right) =$$

$$\begin{aligned}
&= - \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi^2 D_1 & \dots & \Lambda_1 \Pi^k D_1 & \Lambda_1 \Pi^2 D_2 & \dots & \Lambda_1 \Pi^p D_2 \\ \Lambda_1 \Pi D_1 & \dots & \Lambda_1 \Pi^{k-1} D_1 & \Lambda_1 \Pi D_2 & \dots & \Lambda_1 \Pi^{p-1} D_2 \\ \Lambda_2 \Pi D_1 & \dots & \Lambda_2 \Pi^{k-1} D_1 & \Lambda_2 \Pi D_2 & \dots & \Lambda_2 \Pi^{p-1} D_2 \end{array} \right] = \\
&= - \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi^2 D_1 & \dots & \Lambda_1 \Pi^k D_1 & \Lambda_1 \Pi^2 D_2 & \dots & \Lambda_1 \Pi^p D_2 \\ \Lambda \Pi D_1 & \dots & \Lambda \Pi^{k-1} D_1 & \Lambda \Pi D_2 & \dots & \Lambda \Pi^{p-1} D_2 \end{array} \right],
\end{aligned}$$

where $D_1 = (I - \Pi)\Sigma\Lambda'_1$, $D_2 = (I - \Pi)\Sigma\Lambda'_2$. Observe that the first row is $\Lambda_1\Pi\Lambda^{-1}$ times the second row; hence, the matrix $E(-\mathbf{V}_{it}\mathbf{S}'_{it})$ does not have full rank and parameters of the model are not identified.

Now consider the difference moment conditions $E(\Delta\mathcal{G}_{it}\mathbf{S}'_{it}) = E\{(\Delta\mathbf{Y}_{it} - \gamma\Delta\mathbf{V}_{it})\mathbf{S}'_{it}\} = 0$ where $\mathbf{S}_{it} = [\mathbf{L}'_{i,t-3} \dots \mathbf{L}'_{i,t-k} \mathbf{Y}'_{i,t-3} \dots \mathbf{Y}'_{i,t-p}]'$. Find that the Jacobian is

$$\begin{aligned}
E(-\Delta\mathbf{V}_{it}\mathbf{S}'_{it}) &= -E \left(\begin{array}{ccc|ccc} (\mathbf{L}_{i,t-1} - \mathbf{L}_{i,t})\mathbf{L}'_{i,t-3} & \dots & (\mathbf{L}_{i,t-1} - \mathbf{L}_{i,t})\mathbf{L}'_{i,t-k} & (\mathbf{L}_{i,t-1} - \mathbf{L}_{i,t})\mathbf{Y}'_{i,t-3} & \dots & (\mathbf{L}_{i,t-1} - \mathbf{L}_{i,t})\mathbf{Y}'_{i,t-p} \\ (\mathbf{L}_{i,t-2} - \mathbf{L}_{i,t-1})\mathbf{L}'_{i,t-3} & \dots & (\mathbf{L}_{i,t-2} - \mathbf{L}_{i,t-1})\mathbf{L}'_{i,t-k} & (\mathbf{L}_{i,t-2} - \mathbf{L}_{i,t-1})\mathbf{Y}'_{i,t-3} & \dots & (\mathbf{L}_{i,t-2} - \mathbf{L}_{i,t-1})\mathbf{Y}'_{i,t-p} \\ (\mathbf{Y}_{i,t-2} - \mathbf{Y}_{i,t-1})\mathbf{L}'_{i,t-3} & \dots & (\mathbf{Y}_{i,t-2} - \mathbf{Y}_{i,t-1})\mathbf{L}'_{i,t-k} & (\mathbf{Y}_{i,t-2} - \mathbf{Y}_{i,t-1})\mathbf{Y}'_{i,t-3} & \dots & (\mathbf{Y}_{i,t-2} - \mathbf{Y}_{i,t-1})\mathbf{Y}'_{i,t-p} \end{array} \right) = \\
&= \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi^2 D_1 & \dots & \Lambda_1 \Pi^k D_1 & \Lambda_1 \Pi^2 D_2 & \dots & \Lambda_1 \Pi^p D_2 \\ \Lambda \Pi D_1 & \dots & \Lambda \Pi^{k-1} D_1 & \Lambda \Pi D_2 & \dots & \Lambda \Pi^{p-1} D_2 \end{array} \right].
\end{aligned}$$

Hence, the difference moment conditions do not have full rank either because the first row is $\Lambda_1\Pi\Lambda^{-1}$ times the second row. The same conclusion follows for the case without measurement errors, i.e., $\boldsymbol{\varepsilon}_{it} = 0$.

Now suppose that there is no firm-specific effect. Then $E\{(\mathbf{Y}_{it} - \gamma\mathbf{V}_{it})\mathbf{S}'_{it}\} = 0$ with $\mathbf{S}_{it} = [\mathbf{L}'_{i,t-2} \dots \mathbf{L}'_{i,t-k} \mathbf{Y}'_{i,t-2} \dots \mathbf{Y}'_{i,t-p}]'$ is a set of valid moment conditions. However, the reduced rank is still a problem as the Jacobian does not have full rank:

$$E(\mathbf{V}_{it}\mathbf{S}'_{it}) = \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi^2 \Sigma \Lambda'_1 & \dots & \Lambda_1 \Pi^k \Sigma \Lambda'_1 & \Lambda_1 \Pi^2 \Sigma \Lambda'_2 & \dots & \Lambda_1 \Pi^p \Sigma \Lambda'_2 \\ \Lambda \Pi \Sigma \Lambda'_1 & \dots & \Lambda \Pi^{k-1} \Sigma \Lambda'_1 & \Lambda \Pi \Sigma \Lambda'_2 & \dots & \Lambda \Pi^{p-1} \Sigma \Lambda'_2 \end{array} \right]$$

where the first row is $\Lambda_1\Pi\Lambda^{-1}$ times the second row.

Now consider $\begin{bmatrix} \mathbf{L}_{it} \\ \mathbf{Y}_{it} \end{bmatrix} = \Lambda \mathbf{F}_{it} + \begin{bmatrix} \bar{\mathbf{L}}_i \\ \bar{\mathbf{Y}}_i \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathbf{v}_{it} + \boldsymbol{\varepsilon}_{it} = \Lambda \mathbf{F}_{it} + \begin{bmatrix} \bar{\mathbf{L}}_i \\ \bar{\mathbf{Y}}_i \end{bmatrix} + B \mathbf{v}_{it} + \boldsymbol{\varepsilon}_{it}$ that nests models

where some of the inputs can response contemporaneously to changes in productivity (the matrix B is square). This modification also results in level and difference moments not having full rank because the structure of the moment conditions is not changed. For example, consider the difference moment conditions and find that the Jacobian is:

$$E(-\Delta\mathbf{V}_{it}\mathbf{S}'_{it}) = \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi^2 D_1 & \dots & \Lambda_1 \Pi^k D_1 & \Lambda_1 \Pi^2 D_2 & \dots & \Lambda_1 \Pi^p D_2 \\ \Lambda \Pi D_1 & \dots & \Lambda \Pi^{k-1} D_1 & \Lambda \Pi D_2 & \dots & \Lambda \Pi^{p-1} D_2 \end{array} \right] +$$

$$+ \left[\begin{array}{ccc|ccc} \Lambda_1 \Pi D_3 & \dots & \Lambda_1 \Pi^{k-1} D_3 & \Lambda_1 \Pi D_4 & \dots & \Lambda_1 \Pi^{p-1} D_4 \\ \hline \Lambda D_3 & \dots & \Lambda \Pi^{k-2} D_3 & \Lambda D_4 & \dots & \Lambda \Pi^{p-2} D_4 \end{array} \right],$$

where $\Sigma_v = E(\mathbf{v}_i \mathbf{v}_i')$, $D_3 = (I - \Pi)\Sigma_v B_1'$ and $D_4 = (I - \Pi)\Sigma_v B_2'$. This matrix does not have full rank because the first row is equal to $\Lambda_1 \Pi \Lambda^{-1}$ times the second row. ■

Proof of Proposition 6.

It is shown in Proposition 5 that level and difference moment conditions yield the same Jacobian matrix:

$$D = \begin{bmatrix} \Psi(\tilde{\Gamma}_2 - \tilde{\Gamma}_3) & \Psi(\tilde{\Gamma}_3 - \tilde{\Gamma}_4) & \dots & \Psi(\tilde{\Gamma}_d - \tilde{\Gamma}_{d+1}) \\ \tilde{\Gamma}_1 - \tilde{\Gamma}_2 & \tilde{\Gamma}_2 - \tilde{\Gamma}_3 & \dots & \tilde{\Gamma}_{d-1} - \tilde{\Gamma}_d \end{bmatrix},$$

where $\Psi \equiv [0 \mid I_{p+m-1}]$. Given assumptions of the problem, identification of the Blundell-Bond estimator requires that $\text{rank}(D) = 2(m + p) - 1$.

Observe that $\tilde{\Gamma}_k - \tilde{\Gamma}_{k+1} = \Upsilon \Pi^k (I - \Pi) \Gamma_0 \Upsilon'$. Consider matrix $P = \Psi \begin{bmatrix} 0 & \Pi_{12} & 0 \\ 0 & \Pi_{22} & 0 \end{bmatrix}$.

Multiply the second row of D by P and subtract from the first row of D . Denote the resulting matrix with D_1 :

$$\begin{aligned} D_1 &= \begin{bmatrix} \Phi \Pi (I - \Pi) \Gamma_0 \Upsilon' & \Phi \Pi^2 (I - \Pi) \Gamma_0 \Upsilon' & \dots & \Phi \Pi^{d-1} (I - \Pi) \Gamma_0 \Upsilon' \\ \Upsilon \Pi (I - \Pi) \Gamma_0 \Upsilon' & \Upsilon \Pi^2 (I - \Pi) \Gamma_0 \Upsilon' & \dots & \Upsilon \Pi^{d-1} (I - \Pi) \Gamma_0 \Upsilon' \end{bmatrix} = \\ &= \underbrace{\begin{bmatrix} \Phi \Pi \\ \Upsilon \Pi \end{bmatrix}}_{D_1^*} \underbrace{\left[(I - \Pi) \Gamma_0 \Upsilon' \mid \Pi (I - \Pi) \Gamma_0 \Upsilon' \mid \dots \mid \Pi^{d-2} (I - \Pi) \Gamma_0 \Upsilon' \right]}_{D_1^\dagger} = D_1^* D_1^\dagger \end{aligned}$$

where $\Phi = \Psi \begin{bmatrix} 0 & 0 & \Pi_{13} \\ 0 & 0 & \Pi_{23} \end{bmatrix}$. Observe that

$\text{rank}(D) = \text{rank}(D_1) \leq \min\{\text{rank}(D_1^*), \text{rank}(D_1^\dagger)\}$ and

$$\text{rank}(D_1^*) = \text{rank} \left(\begin{bmatrix} \Phi \Pi \\ \Upsilon \Pi \end{bmatrix} \right) = \text{rank} \left(\begin{bmatrix} \Psi \begin{bmatrix} 0 & 0 & \Pi_{13} \Pi_{33} \\ 0 & 0 & \Pi_{23} \Pi_{33} \end{bmatrix} \\ \begin{bmatrix} 0 & \Pi_{12} & \Pi_{13} \\ 0 & \Pi_{22} & \Pi_{23} \end{bmatrix} \end{bmatrix} \right) \leq m + n.$$

Hence the model can be identified if $m + n \geq 2(p + m) - 1 \Leftrightarrow n \geq 2p + m - 1$. ■

Proof of Proposition 7.

Order entries of \mathbf{X}_t so that the first element in \mathbf{X}_t (and \mathbf{G}_t) is output y . Without loss of generality suppose that there is only one freely-adjusted input l (labor) such that the first-order condition

with respect to this input is $y_t - \phi l_t = w_t$ where ϕ is some constant, $w_t = e_w X_t$ is an exogenous shock, and e_w is the selection vector (i.e., e_w is equal to one at the position of w_t in X_t and zero otherwise). Also, without loss of generality, assume that all other inputs are predetermined. Define Π^y and Π^l as rows of the matrix Π that correspond to the output y_t and the freely adjusted input l_t . By (43),

$$(\Pi^y - \phi \Pi^l) X_{t-1} + (B^y - \phi B^l) v_t = y_t - \phi l_t = w_t = e_w X_t = e_w \Pi X_{t-1} + e_w B v_t.$$

Since this holds for any X_t and v_t , $(\Pi^y - \phi \Pi^l) = e_w$ and $(B^y - \phi B^l) = e_w$. It follows that $\Pi_{12}^y - \phi \Pi_{12}^l = 0$ and $\Pi_{13}^y - \phi \Pi_{13}^l = e_w \Pi_{33}$. To a first-order log-linear approximation, production (revenue) function $y_t = \alpha_L l_t + \alpha K_t + a_t$ imposes another restriction on Π : $\Pi_{13}^y - \alpha_L \Pi_{13}^l - e_a \Pi_{33} = \alpha \Pi_{23}$ and $\Pi_{12}^y - \alpha_L \Pi_{12}^l = \alpha \Pi_{22}$. Using these restrictions and the proof of Proposition 6, one finds that the rank of the Jacobian matrix D is:

$$\begin{aligned} \text{rank}(D) &\leq \text{rank}(D_1^*) = \text{rank} \left(\begin{bmatrix} \Psi \\ \begin{bmatrix} 0 & 0 & \Pi_{13} \Pi_{33} \\ 0 & 0 & \Pi_{23} \Pi_{33} \end{bmatrix} \\ \begin{bmatrix} 0 & \Pi_{12} & \Pi_{13} \\ 0 & \Pi_{22} & \Pi_{23} \end{bmatrix} \end{bmatrix} \right) \\ &= \text{rank} \begin{pmatrix} 0 & \Pi_{13}^l \Pi_{33} \\ 0 & \Pi_{23} \Pi_{33} \\ \Pi_{12}^y & \Pi_{13}^y \\ \Pi_{12}^l & \Pi_{13}^l \\ \Pi_{22} & \Pi_{23} \end{pmatrix} = \text{rank} \begin{pmatrix} 0 & (\Pi_{13}^l - \frac{1}{\phi - \alpha_L} \alpha \Pi_{23}) \Pi_{33} \\ 0 & \Pi_{23} \Pi_{33} \\ 0 & \Pi_{13}^y - \frac{\phi}{\phi - \alpha_L} \alpha \Pi_{23} \\ 0 & \Pi_{13}^l - \frac{1}{\phi - \alpha_L} \alpha \Pi_{23} \\ \Pi_{22} & \Pi_{23} \end{pmatrix} \\ &= \text{rank} \begin{pmatrix} 0 & (\frac{1}{\phi - \alpha_L} e_a - \frac{1}{\phi - \alpha_L} e_w) \Pi_{33}^2 \\ 0 & \Pi_{23} \Pi_{33} \\ 0 & e_w \Pi_{33} \\ 0 & e_a \Pi_{33} \\ \Pi_{22} & \Pi_{23} \end{pmatrix} = \text{rank} \begin{pmatrix} 0 & 0 \\ 0 & \Pi_{23} \Pi_{33} \\ 0 & e_w \Pi_{33} \\ 0 & e_a \Pi_{33} \\ \Pi_{22} & \Pi_{23} \end{pmatrix} \leq 2(p+m) - 2 \end{aligned}$$

The last equality follows from Π_{33} being diagonal. Since the rank is less than $2(p+m)-1$, the estimator is not identified. ■