



DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

**UNTESTED ASSUMPTIONS AND DATA SLICING: A
CRITICAL REVIEW OF FIRM-LEVEL PRODUCTION
FUNCTION ESTIMATORS**

Markus Eberhardt and Christian Helmers

Number 513
November 2010

Manor Road Building, Oxford OX1 3UQ

Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators*

Markus EBERHARDT^a

`markus.eberhardt@economics.ox.ac.uk`

Christian HELMERS^b

`c.r.helmerts@lse.ac.uk`

^a *Department of Economics, University of Oxford*

^b *Centre for Economic Performance, London School of Economics*

17th November 2010

ABSTRACT

This paper surveys the most popular parametric and semi-parametric estimators for Cobb-Douglas production functions arising from the econometric literature of the past two decades. We focus on the different approaches dealing with ‘transmission bias’ in firm-level studies, which arises from firms’ reaction to unobservable productivity realisations when making input choices. The contribution of the paper is threefold: we provide applied economists with (i) an in-depth discussion of the estimation problem and the solutions suggested in the literature; (ii) a detailed empirical example using FAME data for UK high-tech firms, emphasising analytical tools to investigate data properties and the robustness of the empirical results; (iii) a powerful illustration of the impact of estimator choice on TFP estimates, using matched data on patents in ‘TFP regressions’. Our discussion concludes that while from a theoretical point of view the different estimators are conceptually very similar, in practice, the choice of the preferred estimator is far from arbitrary and instead requires in-depth analysis of the data properties rather than blind belief in asymptotic consistency.

KEYWORDS: Productivity, Production Function, UK firms, Panel data estimates

JEL Classification: D21, D24, L25, O23

*We thank Bronwyn Hall, Jonathan Haskel, Simon Quinn, Måns Söderbom and Nicolas van de Sijpe for helpful comments and suggestions. We also acknowledge substantive comments from Steve Bond and David Greenstreet on an earlier version of this paper. We benefited from comments following presentation at the 10th Comparative Analysis of Enterprise Data & COST Conference. All remaining errors are our own. We are further indebted to Måns Söderbom for making his **Stata** computer code for the MD estimator and COM-FAC test publicly available on the internet. In order to promote the application of the concepts and practices discussed in this article, we will provide detailed **Stata** code for our empirical application on a companion website (<https://sites.google.com/site/medevecon/publications-and-working-papers/code>). Eberhardt gratefully acknowledges financial support from the UK Economic and Social Research Council [grant number PTA-026-27-2048].

“[C]hoosing a method [for production function estimation] is a matter of asking oneself which assumptions one is comfortable making. Certainly, one cannot escape the fact that *some* assumptions must be made when estimating the production function.”

Syverson (2010: p.9, emphasis in original)

“[I]n trying to evade this problem [transmission bias], researchers have shifted their focus to thinner and thinner slices of data, leading to the exacerbation of other problems and misspecifications. Much of this line of work has been guided, unfortunately, by what ‘econometrics’ as a technology might be able to do for this problem, rather than focusing on the more important but technically less tractable problem of data quality and model specification.”

Griliches and Mairesse (1995, p.22)

1 Introduction

The estimation of firm-level productivity has attracted an enormous amount of attention in the economic literature.¹ Despite the recent emergence of Randomized Control Trials (RCT) for the analysis of (micro-)enterprise-level data (de Mel, McKenzie and Woodruff, 2008, 2009, 2010) and the domination of the Rubin Causal Model in much of microeconomic thinking (e.g. Angrist and Pischke, 2009), regression-based productivity estimates employing observational data are still very popular and either of direct interest or used to investigate a range of related issues: the effect of foreign direct investment on domestic firms (Javorcik, 2004; Haskel et al. 2007), international technology sourcing (Griffith et al., 2006), the effect of trade liberalisation (Tybout and Westbrook, 1995; Pavcnik, 2002), firm and industry restructuring (Disney et al., 2003), agglomeration externalities (Greenstone et al., 2010), the effect of exporting (Blalock and Gertler, 2004; Bernard and Jensen, 2004), returns to R&D (Hall et al., 2009) as well as firm location and market structure (Syverson, 2004) among other topics. In firm-level studies these estimates are commonly derived from simple Cobb-Douglas production function regressions, frequently with no more than passing mention of the various difficulties arising in this first step of the empirical analysis. Prime amongst these is the possible correlation between firms’ input choices and their idiosyncratic productivity shocks, arising from the fact that firms observe these idiosyncratic shocks and adapt their input choices accordingly. If such correlation is present in the data and given that firm-specific productivity shocks are unobserved by the econometrician, the resulting OLS estimates of input coefficients are biased and inconsistent.

While economists have been well aware of this so called ‘transmission bias’ for considerable time, only the development of new estimation techniques over the past two decades has allowed for the issue to be tackled in empirical analysis. A number of solutions to the problem have been introduced and the most popular of these, the dynamic panel estimators by Arellano and Bond (1991) [henceforth AB] and Blundell and Bond (1998) [BB] on the one hand and the ‘structural estimators’ (more on the appropriateness of this common label below) by Olley and Pakes (1996) [OP], Levinsohn and Petrin (2003) [LP] and Akerberg, Caves and Frazer (2006) [ACF] on the other, have since been applied in literally thousands of empirical papers.² While initially

¹Note that we use the terms productivity and TFP interchangeably throughout this article.

²Merely as an illustration of this popularity we report the *Google Scholar* citation count for each of these articles as of April 2010 (in order of citations): Arellano and Bond (1991) — 6,900; Blundell and Bond (1998) — 3,111 [these counts include references to applications of the GMM estimators to macro data as well as other econometric theory papers; Blundell and Bond (2000), which focuses on an application of the Blundell and Bond (1998) estimator to the estimation of Cobb-Douglas production functions, received 622 citations]; Olley and Pakes

confined to the skilled econometrician, the application of these methods has been boosted substantially by the availability of pre-coded routines in standard statistical programmes,³ most prominently in **Stata**.

Unlike the suggestions in Zvi Griliches' and Jacques Mairesse's seminal review of the literature and Chad Syverson's more recent survey paper quoted above, many of these applications do not question the validity of the empirical estimator of choice *in the particular data environment studied*. Frequently the superiority of the chosen estimation approach over alternatives other than OLS or Fixed Effects estimation is implicitly assumed, but not questioned or investigated. This is particular the case with regard to the 'structural' assumptions made in the various novel estimation techniques, whose validity or failure have a crucial bearing on performance. Discussion of the empirical implementation is too frequently limited to a carbon copy of the econometric equations that make up the complex regression technique rather than a concern for the intuition and validity of the assumptions made in the process. Instead, the different dynamic panel and 'structural' estimators are considered to be equally suitable solutions, and the specific choice among them is reduced to a personal preference with little reference to the properties of the data of interest.⁴

The aims and contributions of this paper are threefold: first, we outline and discuss the econometric problems related to the estimation of Cobb-Douglas production functions at an intuitive level, focusing on the transmission bias and the solutions proposed in the recent literature. Our contribution is to provide a comprehensive discussion of all up-to-date existing estimators for Cobb-Douglas micro production functions, spelling out in detail the crucial assumptions made in the process. We attempt to provide a much less technical treatment of the different empirical approaches than is common in the literature, which allows applied researchers with limited interest/insights into the econometric theory to make a sensible, data-dependent choice over the most appropriate empirical implementation.

Ideally, the conceptual issues raised in this discussion would be investigated in detail using simulated firm-level data — however, we feel that in contrast to earlier studies focused on dynamic panel estimators, this strategy is no longer viable in simulations which can incorporate *all* of the recent structural estimators. Monte Carlo simulations are most insightful if the data generating process (DGP) employed is as general as possible; in the present case this would include specifying not only observable net output (sales, value-added), labour and capital stock, but also gross output, material inputs and investment, all of which are involved in a firm-level period-specific optimization process which is responsive to productivity shocks. Ultimately we feel that the idiosyncracies of the necessary choices over the DGP and the controversy these might entail would jeopardize the purpose and informativeness of this (furthermore computationally complex) exercise, while also shifting attention from our aim to inform applied work.

(1996) — 1,563; Levinsohn and Petrin (2003) — 885; Akerberg, Caves and Frazer (2006) — 177. For comparison, the seminal Mankiw, Romer and Weil (1992) paper on cross-country growth empirics has 6,672 citations.

³DPD estimators: Arellano and Bond's DPD for **Gauss** and in particular the **Stata** commands by David Roodman (`xtabond` and `xtabond2` — see detailed documentation in Roodman, 2008); Olley and Pakes estimator: implementation in **Stata** by Yasar, Raciborski and Poi (`opreg`); Levinsohn and Petrin estimator: implementation in **Stata** by Levinsohn, Petrin, and Poi (`levpet`). Arnold (2005) provides an informal introduction to the OP and LP approaches in **Stata**.

⁴Data availability may serve as a choice criterion in the sense that the OP estimator requires investment data.

Instead, and this is the second contribution of the paper, we carry out a practical illustration of the most popular estimators using ‘typical’ observational firm-level data from a source which is easily accessible and thus widely used. Our approach here is to investigate the observable properties of the data, to compare empirical results across different estimators, to interpret the results against the background of the DGP assumptions made by each estimator, and to consult recent empirical work using similar data to find commonalities in the patterns observed across estimation techniques. Although our discussion will naturally be informed by the insights from existing simulation studies and treatments of the topic (e.g. Griliches and Mairesse, 1995; Blundell and Bond, 2000; Akerberg, Benkard, Berry and Pakes, 2006; van Biesebroeck, 2007; van Beveren, 2007; Del Gatto, Di Liberto and Petraglia, 2010), we prefer to emphasise the practical experience with real data much more than these previous discussions: our treatment highlights data construction and preliminary descriptive analysis, which is typically relegated to the technical appendix of an applied paper or omitted entirely. We emphasise that the data properties, in combination with the data requirements of standard empirical estimators raise considerable concerns over sample selection, and we conclude that the ‘external validity’ of results obtained is at best tenuous and in many cases not given.

Our third contribution could be seen as a reminder to the applied literature rather than an original thought: TFP is not an observable variable that can be found on the accounting books of an enterprise but represents an *estimate*, the computation of which makes a host of assumptions and generalisations with ample scope for misspecification and inconsistency. Too many empirical studies begin by regressing TFP on some variable of interest, with the production function results on which these TFP estimates are based relegated to a footnote, the technical appendix or omitted entirely. This treatment neglects the important point that the magnitudes of the technology coefficients for labour and capital (and material) inputs actually *mean something in their own right*: based on data for labour-compensation we know that in a typical firm around 60 to 80% of income share accrues to this factor, so that we would expect empirical labour and capital coefficients of around .7 and .3 respectively. Or at least non-negative coefficients, bounded by zero and unity, with the two (three) coefficients adding up to *somewhere* near unity. In the absence of a standard set of residual diagnostic tests for panel regression, we feel that the analysis of finite samples must pay close attention to the magnitudes of these coefficient estimates and we therefore make them the centre of attention in our empirical illustration. We furthermore provide some illustrative analysis of the ‘determinants of TFP’ by regressing TFP estimates on patent data. This exercise reveals substantial differences in the results depending on which estimator was used to construct the TFP estimates and thus vindicates the concerns we raise in this survey. We further highlight that a pairwise correlation analysis of the various TFP estimates would suggest a high degree of commonality, thus creating a false sense of security over the choice of estimator which, as we illustrate, may not be warranted in practice.

The remainder of this paper is organized as follows. Section 2 provides a detailed overview and discussion of the existing literature on Cobb-Douglas production function estimation with firm-level panel data. Section 3 describes the application of the estimators to observational firm-level data and comments on their relative performance. Section 4 provides some concluding remarks, further pointing to some relevant issues deserving more attention in applied work.

2 Theory

A standard Cobb-Douglas production function is given by

$$Y_j = F(A_j, K_j, L_j) = A_j K_j^{\beta_k} L_j^{\beta_l} \quad (1)$$

where Y_j denotes firm j 's output, commonly measured as value-added,⁵ L_j denotes labour input (commonly headcount or a measure of hours worked – more on actual measurement in Section 3.2), K_j physical capital stock. A_j is a measure of firm j 's level of efficiency which is unobserved by the econometrician in the available data. β_i with $i \in (k, l)$ denotes output elasticities with respect to factor inputs.⁶ The theoretical foundations for this highly parsimonious framework for productivity analysis are rather compelling (Solow, 1956): unobserved determinants of output are contained in the productivity term A_j , commonly referred to as Total Factor Productivity (TFP). The specification in (1) assumes that the effect of A_j on Y_j is Hicks-neutral. This implies that in log-linearized form, TFP is additively separable from the other production factors.

Hence, an empirical equivalent of this Cobb-Douglas function can be represented as

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_o + \varsigma_{jt} \quad (2)$$

where we have introduced time subscripts $t = 1, \dots, T$ and denote log values with lower case letters. In Equation (2), $\ln(A_j)$ is decomposed into two elements, β_o and ς_{jt} , where the constant β_o represents mean efficiency across all firms and the time-variant ς_{jt} can be regarded as representing deviations from this mean, capturing (i) unobserved factors affecting firm output, (ii) measurement error in output and inputs, and (iii) random noise. In the words of Griliches and Mairesse (p. 6: 1995), “ ς_{jt} is the econometrician’s problem, not the producer’s”. We now highlight this issue in greater detail.

2.1 The Transmission Bias Problem

The estimation of (2) encounters several well-known problems. Define ς_{jt} as

$$\varsigma_{jt} = \omega_{jt}^* + v_{jt} = \eta_j + \omega_{jt} + \gamma_t + v_{jt} \quad (3)$$

which indicates that ς_{jt} contains measurement error v_{jt} and a productivity term ω_{jt}^* (TFP), which is known to the firm but unobserved by the researcher. Measurement error is commonly assumed to be serially uncorrelated, but we allow for the possibility of correlation between

⁵A sizeable literature discusses the misspecification of production functions of value-added, capital stock and labour, as opposed to adopting gross output-based specifications which incorporate material inputs (including energy as well as raw materials and components). A discussion of this issue would go beyond the scope of this paper; instead we refer to the work by Bruno (1984) and Basu and Fernald (1995, 1997).

⁶An index approach to productivity determination would simply construct

$$A_j = \frac{Y_j}{K_j^{\beta_k} L_j^{\beta_l}}$$

where β_k and β_l are taken directly from the observed data: if a firm is minimising cost, then it will set β_k and β_l equal to the product of the scale elasticity and the respective input’s cost-share. If the researcher is willing to assume constant returns to scale and perfect competition the elasticities equal the revenue-shares paid to each input, with β_l readily available from data on labour compensation and $\beta_k = 1 - \beta_l$. The fact that these assumptions are far from innocuous leads researchers to estimate elasticities in the first place.

v_{jt} and the observable inputs (l, k) . A motivation for the latter may be the suggestion that larger firms may have more complex accounts, thus making it more difficult (and more error-prone) to establish their ‘true’ capital stock.⁷ The productivity shock ω_{jt}^* can be further split into an element common to all firms and firm-specific elements. The former represents for instance macroeconomic shocks which affect all firms and industries *in the same way* or overall ‘technological progress’ which improves productivity in *all* firms, ‘a common tide that lifts all boats’, such as the widespread introduction and improvement of ICT. For simplicity we use γ_t to represent these common shocks and average processes and specify that this is defined for $t = 2, \dots, T$ (i.e. we set $\gamma_1 = 0$), which implies that β_o represents average productivity in the base period $t = 1$.⁸ Finally, the firm-specific productivity term can be further divided into time-invariant and time-variant components: if certain firms have permanently higher productivity levels, for instance due to their industrial sector of operation or their historical geographic location, then these time-invariant productivity effects can be empirically represented with firm fixed effects, say η_j , which are defined for firms $j = 2, \dots, N$.⁹ In the latter case, η_j represents the *permanent* deviation of firm j from the reference firm productivity level in the base year β_o , γ_t represents the average technological progress (productivity increase) in the sample over time, and ω_{jt} represents the combined effect of firm-specific deviation from its own TFP level in the base period and from the common or average technological progress in period t . This firm- and time-specific effect or shock ω_{jt} can be attributed to “the technical knowledge, the will, effort and luck of a given entrepreneur” (Marschak and Andrews, p.145: 1944) in a given time period. We can now rewrite equation (2) to yield

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_o + \eta_j + \gamma_t + \omega_{jt} + v_{jt} \quad (4)$$

The main problem for estimation of specifications such as in Equation (4) arises from the plausible suggestion that firms decide on their choice of inputs (l, k) based on the realized firm-specific productivity shock (ω_{jt}) , *which only they observe*: for instance a favorable productivity shock to firm j might induce higher levels of investment whereas a negative shock might induce firms to reduce their productive worker headcount — the exact timing of these effects and definition of a firm’s ability to ‘adjust’ productive inputs will play an important part in the estimation strategies reviewed in this paper.¹⁰ If the effect of inputs k and l on output y cannot be separated from ω by the econometrician, β_k and β_l are *not identified*, because even knowing the (‘true’) population distribution of the data would not allow the econometrician to obtain unbiased and consistent estimates of the input coefficients. Since ω_{jt} is suggested to ‘transmit to’ the input choices, this particular problem is known as the ‘transmission bias’.

In the present setup, employing standard least squares estimation by regressing observable output on observable inputs yields biased and inconsistent coefficient estimates $\hat{\beta}_l$ and $\hat{\beta}_k$ if

⁷Generally, capital stock is more likely than labour input to be measured with error (Levinsohn and Petrin, 2003), which using our notation above can be expressed as $k_{jt} = k_{jt}^* + v_{jt}$ where * indicates the observed capital stock.

⁸This interpretation holds if panels are unbalanced, although in this case β_o no longer refers to the base period of the sample, but the average productivity in the base period of each firm.

⁹Contrary to the time-series dimension of a dataset (from $t = 1, \dots, T$), there is no natural ordering for the data in the cross-section dimension. In practical terms, however, it does not matter for estimation which of the N firms represents the reference point.

¹⁰Given that the econometrician also does not observe the common productivity γ_t and that fixed productivity effects may have a systematic relationship with the size of idiosyncratic shocks we could argue that $\beta_o + \eta_j + \gamma_t + \omega_{jt}$ are the productivity elements influencing input choice.

inputs are not ‘mean-independent’ from the omitted productivity effect (Marshak and Andrews, 1944). The adoption of time- and firm-fixed effects cannot solve this problem, given the continued existence of a contemporaneous correlation between ω_{jt} and the firm’s input choices.¹¹ This correlation is more likely to be present if inputs can adjust quickly. Commonly, labour is assumed to adjust more rapidly than physical capital such as machinery, which is accumulated over time and more difficult to be rid of, thus pointing to a (stronger) bias for the labour coefficient estimate. In this context, note however the general nature of firm-level panel datasets which commonly investigate/interview firms at annual intervals: ‘contemporaneous’ correlation thus should not be assumed to imply ‘instantaneous’ changes, but adjustment within the accounting period. Furthermore, adjustment of labour and capital should not be defined narrowly as hiring/firing of workers and acquisition/sale of machinery, but may include more flexible strategies of temporary reduction of working hours and physical capacity levels. In this light the notion of relatively ‘slow’ or ‘rapid’ input adjustment for labour versus capital is arguably less obvious which also suggests that the (relative) speed of adjustment may vary among different datasets. In any case, in order to understand the direction of the bias induced in factor input coefficients (we stick with the labour case), consider a firm’s optimisation problem in each time period under the assumption of perfectly competitive input and output markets

$$\max \pi_j = pA_j K_j^{\beta_k} L_j^{\beta_l} - wL_j - rK_j \quad (5)$$

where w denotes the wage paid by firm j , r the user cost of capital, p the output price (all of which are industry-wide equilibrium prices) and π_j is firm profit. Since all terms are contemporaneous we dropped the time subscript for readability. The corresponding first order condition with respect to labour input is then

$$L_j = \left(\frac{\beta_l p A_j}{w} \right)^{\frac{1}{1-\beta_l}} K_j^{\frac{\beta_k}{1-\beta_l}} \quad (6)$$

Rewritten in logarithms it can be seen that l_j is a function of the productivity shock ς_j

$$l_j = \frac{1}{1-\beta_l} (\ln \beta_l + \ln p + \beta_o + \varsigma_j - \ln w + \beta_k k_j) \Leftrightarrow l_j = f(\varsigma_j) \quad (7)$$

Thus when using OLS to estimate Equation (4) *without* accounting for the presence of productivity ω_{jt} , the bias induced by endogeneity between labour and the productivity shock is positive, such that the labour coefficient estimate will be biased upward ($\hat{\beta}_l > \beta_l$). Similarly for endogenous capital stock, although the nature of the firm’s problem is then transformed into a dynamic optimization problem.

Additional problems for estimation arise from the fact that productivity shocks to a given firm are likely to be serially correlated, that is the impact of a positive productivity shock is likely not confined to a single period. Most generally

$$\omega_{jt} = g(\omega_{jt-s}) + \xi_{jt} \quad (8)$$

where $g(\cdot)$ is some function of past productivity and ξ_{jt} is the idiosyncratic productivity shock

¹¹In fact the firm fixed effects may introduce additional problems for estimation if we suggest that there may be a systematic relationship between a firm’s productivity *level* and its idiosyncratic technology shock in period t , e.g. firms with higher (lower) levels of productivity experience higher (lower) rates of technical progress.

in period t . This consideration gives rise to dynamic empirical specifications which feature lagged terms of the dependent and independent variables and where long-run coefficients need to be computed following consideration for ‘common factor restrictions’. In addition, differences in productivity across firms may be substantial and also serially correlated, i.e., there is considerable unobserved heterogeneity which is persistent over time (Mundlak, 1961). Thus the difference between, say ω_{1t} (firm 1) and ω_{2t} (firm 2) may be large at any one point in time, but there is an *additional* problem if this productivity differential is persistent over time, i.e., the serial correlation in the unobserved differential ($\omega_{1t}-\omega_{2t}$); more on the serial correlation issue and how this is addressed follows further below.

These sources of ‘transmission bias’ and related issues aside, we already noted the potential for measurement error in the input variables which leads to the well-known ‘attenuation bias’. This acts as a downward bias on the factor input coefficients ($\hat{\beta}_l < \beta_l$, $\hat{\beta}_k < \beta_k$), which is more commonly suspected for physical capital. Note that this bias is present even though the measurement error itself is assumed random: assuming capital k_{jt}^* is the true capital stock, which is, however, measured with error (v_{jt}), i.e., $k_{jt} = k_{jt}^* + v_{jt}$, then a simple production function $y_{jt} = \beta_o + \beta_l l_{jt} + \beta_k k_{jt}^* + \varsigma_{jt}$ is thus estimated as $y_{jt} = \beta_o + \beta_l l_{jt} + \beta_k k_{jt} + (\varsigma_{jt} - \beta_k v_{jt})$, where observed mismeasured capital stock k_{jt} is negatively correlated with the error term in parentheses. As pointed out by Griliches and Mairesse (1995), measurement error can arise from differences in quality of labour input and utilisation of capital across firms. The severity of the attenuation bias is driven by the variance of the measurement error.

Moreover, all of the estimators discussed below assume homogeneity of the production technology (with regard to β_l , β_k) across all firms. If the sample contains heterogeneous firms, for example from different sectors, unaccounted heterogeneity may lead to bias in the estimated input coefficients as it will be captured by the error component and be correlated with the input measures (van Biesebroeck, 2007). For this and other reasons (related to firm fixed effects in the structural estimators detailed below) it has become common practice to estimate production functions at a more narrowly-defined sectoral level, rather than for ‘total manufacturing’, using ‘thinner and thinner slices’ of data (Griliches and Mairesse, 1995). Not all researchers agree that this solves the problems arising from heterogeneity and fixed effects, with Griliches and Mairesse (1995) concluding that “the observed variability-heterogeneity does not really decline as we cut out data finer and finer. There is a sense in which different bakeries are just as much different from each other, as the steel industry is from the machinery industry” (p. 22).

Finally, firm-level panel data commonly exhibits a non-negligible share of firms entering and exiting the market in any one period, introducing a ‘selection problem’ in the estimation equation. More specifically, assume that the probability of firm exit is a function of both unobserved productivity and observed capital stock, meaning firms that do not survive are likely to be those with low levels of capital stock and low productivity. For instance, a large negative productivity shock may induce exit, but not before the firm, in its struggle for survival, has cut down on investment, leading to (relatively) lower capital stock. In reverse, those firms that survived and stayed in the market and thus in the sample are likely made up of two types: those with high capital stocks, which are able to survive regardless of their productivity levels, as well as those who have small capital stocks but pull through on the back of high productivity levels.

Due to the latter the sample of survivors will be characterised by a *negative* relationship between productivity ς_j (high) and capital stock (low). This ‘attrition’ effect leads to a downward bias in the capital coefficient ($\hat{\beta}_k < \beta_k$). In the following, we largely abstract from heterogeneity and the selection problem and focus on the transmission bias, although we do not imply that the former can be neglected in applied work.

2.2 Tackling the transmission bias

Several solutions for the endogeneity of input choices with regard to unobserved productivity have been proposed in the literature — note that all of the methods described below assume panel data at the firm level (large N , small T).¹² We begin by introducing very simple transformations of the empirical equation which dominated the applied literature until relatively recently. Section 2.2.2 then briefly discusses instrumentation using price data, which commonly has not been deemed very satisfactory in addressing the problem. Section 2.2.3 introduces the work by Arellano and Bond (1991) and Blundell and Bond (1998) on ‘dynamic panel estimators’, which is followed by a set of three ‘structural’ estimators in sections 2.2.4 to 2.2.6. Note that we use this terminology only as convenient labels for the two groups of estimators although we do *not* imply that the dynamic panel estimators are ‘less structural’ than the ‘control function’ approaches (to use another label at times attached to this group of estimators). In fact, Bond and Söderbom (2006) derive similarly ‘structural’ foundations to the dynamic panel data estimators — see section 2.2.6. More generally, one could suggest that a truly structural regression approach derives first order conditions from a behavioural model and estimates them. The ‘structural’ estimators the authors merely appeal to theory to derive general assumptions under which their econometric estimators deliver consistent estimates, which is conceptually the same as identifying valid and informative instruments, be it in a Generalised Method of Moments (GMM) framework or indeed any standard IV regression.¹³ Three recent related contributions by Doraszelski and Jaumandreu (2008), Greenstreet (2007) and Bai (2009) complete our theoretical discussion in section 2.2.7.

Our discussion largely abstracts from the impact of measurement error (v_{jt}). The empirical specification discussed throughout the remainder of this article is thus

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_o + \eta_j + \gamma_t + \omega_{jt} \quad (9)$$

2.2.1 Transformations of the empirical specification

Given the panel structure of the data, the earliest solutions to the unobserved heterogeneity and endogeneity problems involved employing firm fixed effects (via dummy variables or the ‘within’-transformation) or an empirical specification with variables in first differences. In the former case this is because the time-series average of time-invariant η_j is simply η_j itself, so that the transformation $\tilde{z}_{jt} = z_{jt} - T^{-1} \sum_t z_{jt}$ does away with the firm-level effect. In the latter case taking first differences of the time-invariant η_j has the same effect. For endogeneity to be

¹²In particular the methods developed by Arellano and Bond (1991) and Blundell and Bond (1998) have been widely used in cross-country or cross-region studies of productivity (moderate N , moderate T), at times using time-averaged data. We do not discuss these macro panel empirics in this survey but refer to a detailed discussion in Eberhardt and Teal (2010).

¹³We are grateful to Måns Söderbom for this insight.

brought under control as a result of these transformation, it is necessary to assume *no* further firm-specific productivity shocks, $(\omega_{jt} = 0 \forall j, t)$. Thus the unobserved productivity *differences* across firms are assumed to be constant over time, with any increase in TFP affecting all firms in equal terms. Therefore any productivity increase over time, captured in the model by γ_t and empirically implemented via year dummies, must disseminate to all firms equally and within the same time period t . A perhaps more sensible view of the assumptions imposed is that there *do* exist firm-specific productivity shocks (ω_{jt}) but that the factor inputs are strictly exogenous with respect to these, implying $\mathbb{E}[\omega_{jt}k_{js}] = 0$ and $\mathbb{E}[\omega_{jt}l_{js}] = 0 \forall s, t$ (Wooldridge, 2009). This strict exogeneity assumption provides the motivation for the dynamic panel data models described below, since it rules out dynamics in inputs and productivity.

Under these assumptions the transformed production function can be estimated using OLS.¹⁴ From a theoretical standpoint, the assumption of constant productivity differences across firms is quite unrealistic, while from practical experience the fixed effects or first difference estimators commonly perform poorly, often leading to the collapse of the capital coefficient vis-à-vis the OLS results ($\hat{\beta}_k^{FE} \ll \hat{\beta}_k^{OLS}$) and indicating severely decreasing returns to scale ($\hat{\beta}_l + \hat{\beta}_k \ll 1$). Note that large decreasing returns to scale imply that firms “consistently price output below marginal cost, which obviously makes no economic sense” (Basu and Fernald, 1997: 251). Aside from transmission bias, this pattern is likely induced by measurement error in the input variables, whose impact is exacerbated by the variable transformation.¹⁵ Using fixed effects, within-transformation or first-differences has the additional drawback of removing substantial information from the data, since only variation over time remains to identify the parameters.

In other empirical contexts researchers are often comparing results for the standard fixed effects estimator and the alternative ‘random effects’ or ‘generalised least squares’ (GLS) estimator.¹⁶ The latter provides a consistent and efficient estimator under the strong assumption that the firm-specific effects (η_j) are random, meaning that — in a sample containing firms from different sectors — productivity levels for, say, garment manufacture are just as likely to be higher than for computer chip manufacture as the reverse. This assumption entails the further implication that firm-specific effects are not correlated with input choice, i.e., strictly exogenous: $\mathbb{E}[\eta_j k_{jt}] = 0$ and $\mathbb{E}[\eta_j l_{jt}] = 0 \forall t$. As a result the GLS estimator is seldom applied in the study of firm-level productivity.

2.2.2 Standard IV regression

One potential solution to the transmission bias problem is to use input prices (r, w) as instruments for input quantities (k, l) . Endogeneity concerns are commonly more focused on the capital stock variable, s.t. instrumentation applies predominantly to rental rates. From a firm’s first-order conditions (see the example for labour above), it can be argued that (lagged) input prices are *informative* instruments for input quantities, i.e., there exists at least a partial correlation between r and k , or w and l if we run of a regression of k (l) on all other variables

¹⁴Ideally clustering standard errors at the firm-level since the differenced or within-transformed errors are serially correlated.

¹⁵Assuming that serial correlation in the measurement errors is close to zero, while serial correlation of true unobserved inputs is positive.

¹⁶This estimator also fits our section headline since it is commonly implemented via ‘ θ -differencing’ where variables are transformed accounting for the variance of the firm-level effects and the remainder of the residuals respectively (obtained in a first stage regression).

included in the production function as well as r (w). Furthermore, for input price series to be *valid* instruments, it is required that they are uncorrelated with unobserved productivity, i.e., they are mean-independent, thus $\text{Cov}[r_{jt}\omega_{jt}] = 0$ and $\text{Cov}[w_{jt}\omega_{jt}] = 0$, where Cov represents the covariance.¹⁷ If input markets are perfectly competitive, then a single firm's actions do not affect market prices and hence the validity restriction should hold. Empirically, as well as conceptually, this assumption is however hard to defend: firstly, input prices are required to be firm-specific and need to display considerable variation across j in order to identify heterogeneous input choices in the first stage of the IV regression, which is unlikely, at least in developed market economies. Secondly, the perfect competition assumption is likely to be violated in that firms with higher unobserved productivity are likely to wield considerable market power and thus more command over input prices. While this solves the homogeneous price problem of instrumentation, it also induces correlation between productivity and prices, thus invalidating the latter as viable instruments. Thirdly, related to wages as instrument for labour input, a higher 'input price' w at the firm-level may be the result of higher unobserved worker quality (e.g. efficiency wage argument). This unobserved worker quality also enters the unobserved productivity term ω_j , thus leading to invalid instruments. Similar concerns about unobserved quality of inputs apply for rental rates and capital stock.

There may, of course, exist other candidate variables than input prices for instrumentation. While convincing external 'z'-variables are common in other empirical applications, no clear contenders have emerged in the production function literature.

2.2.3 Dynamic panel estimators with efficient 'own'-instrumentation

Due to the practical difficulties of finding reliable instruments in firm-level panels, the econometric literature turned toward the single *guaranteed* supply of informative instruments: past values of the regressors themselves. The GMM estimator developed by Arellano and Bond (1991 [AB]) has its origin in the study of dynamic equilibrium relationships more generally, but firm-level production functions are among the most common applications for this approach. The basic idea of the estimator revolves around the time-series properties of unobserved productivity ω_{jt} , which is assumed to display persistence over time, leading to serial correlation in the unobservables. Rather than to somehow adjust for this problem the empirical equation is transformed to explicitly model the persistence: for instance, assuming a first order autoregressive process, AR(1), for ω_{jt} ¹⁸

$$\omega_{jt} = \rho\omega_{jt-1} + \xi_{jt} \quad |\rho| < 1 \quad \xi_{jt} \sim MA(0) \quad (10)$$

will lead to the dynamic autoregressive distributed lag (ARDL) regression model, which adds lagged levels of the dependent and independent variables to the right-hand side of equation (9)

$$y_{jt} = \rho y_{jt-1} + \beta_l l_{jt} - \rho\beta_l l_{jt-1} + \beta_k k_{jt} - \rho\beta_k k_{jt-1} + (1 - \rho)(\beta_0 + \eta_j) + (\gamma_t - \rho\gamma_{t-1}) + \xi_{jt} + \{v_{jt} - \rho v_{jt-1}\} \quad (11)$$

¹⁷This is referred to as the 'exclusion restriction', with the instruments z (here r_{jt} and w_{jt}) being the variables 'excluded' from the 'structural equation' (here: the production function in Equation 9).

¹⁸In micro data stationarity ($|\rho| < 1$) is commonly assumed, although there is a small literature on unit root testing in short panels, see Bond et al. (2005).

This representation¹⁹ clearly indicates so-called ‘common factor restrictions’, whereby the coefficients on lagged regressors are nonlinear combinations of the coefficients on y_{jt-1} (ρ) and on the respective contemporaneous factor input (β_l, β_k). The unobserved productivity term ω_{jt} is no longer part of the regression equation due to the transformation carried out, although the productivity shock ξ_{jt} is still present. The term in $\{ \}$ disappears if there is no measurement error. In macroeconomic models, the coefficient on the lagged dependent variable can give insights into the speed of adjustment of the system ($1 - \rho$) and is therefore often a focus of attention. In the microeconomics of firm production the dynamic specification may be a crucial *requirement* for the identification of the parameters of interest, in our case β_l and β_k , whereas the dynamics themselves are not of particular interest. A more general, unrestricted dynamic representation of (11) is

$$y_{jt} = \pi_1 y_{jt-1} + \pi_2 l_{jt} + \pi_3 l_{jt-1} + \pi_4 k_{jt} + \pi_5 k_{jt-1} + \alpha_j^* + \gamma_t^* + e_{jt} \quad (12)$$

where $\alpha_j^* = (1 - \rho)(\beta_0 + \eta_j)$, $\gamma_t^* = (\gamma_t - \rho\gamma_{t-1})$, $e_{jt} = \xi_{jt} + \{v_{jt} - \rho v_{jt-1}\}$ and the common factor restrictions can be defined as $\pi_3 = -\pi_1\pi_2$, $\pi_5 = -\pi_1\pi_4$. Equation (12) can be tested for the implied restrictions and if not rejected these can be imposed employing a ‘minimum distance’ or nonlinear least squares estimator. If common factor restrictions are rejected, the long-run solution of the model can be backed out as nonlinear combinations of the estimated coefficients

$$y = \frac{\hat{\alpha}_j^*}{1 - \hat{\pi}_1} + \left(\frac{\hat{\pi}_2 + \hat{\pi}_3}{1 - \hat{\pi}_1} \right) l + \left(\frac{\hat{\pi}_4 + \hat{\pi}_5}{1 - \hat{\pi}_1} \right) k + \frac{\hat{\gamma}^*}{1 - \hat{\pi}_1} \quad (13)$$

using the Delta method to compute corresponding standard errors.²⁰

Merely transforming the production function equation from a static to a dynamic one does not solve the transmission bias issue: using OLS, the factor input variables are contemporaneously correlated with the productivity shock ξ_{jt} (contained in e_{jt}) and a possible fixed effect in productivity (contained in α_j^*); furthermore the lagged dependent variable is also correlated with the omitted fixed effect. The Fixed Effects/Within estimator suffers from ‘Nickell-bias’ (Nickell, 1981), which can be shown quite intuitively to lead to a downward bias in the coefficients if T is small: focusing for simplicity on y_{jt-1} and ξ_{jt} we consider the within-transformation applied to equation (12). We can now see that there is systematic correlation between the transformed terms \tilde{y}_{jt-1} and $\tilde{\xi}_{jt}$ (highlighted in bold and underlined):²¹

$$\tilde{y}_{jt-1} = \mathbf{y}_{jt-1} - \frac{1}{(T-1)} \left(y_{j1} + y_{j2} + \dots + \underline{y_{jt}} + \dots + y_{jT-1} \right) \quad (14)$$

$$\tilde{\xi}_{jt} = \underline{\xi_{jt}} - \frac{1}{(T-1)} (\xi_{j2} + \xi_{j3} + \dots + \mathbf{\xi_{jt-1}} + \dots + \xi_{jT}) \quad (15)$$

The negative correlation between components y_{jt-1} in (14) and $(T-1)^{-1}\xi_{jt-1}$ in (15), highlighted in bold, as well as ξ_{jt} in (15) and $(T-1)^{-1}y_{jt}$ in (14), underlined, dominate the other, positive

¹⁹Equation (11) is easily derived by solving the standard production function in (9) for ω_{jt} and then plugging the righthand side into equation (10) (lagged one period for ω_{jt-1}).

²⁰This is easily implemented in **Stata** using the **nlcom** command. The validity of the confidence intervals and t -statistics computed is, however, far from clear: the distribution of the computed long-run coefficients is heavily skewed to the right because estimates for $\hat{\pi}_1$ close to unity produce very large long-run values. In practice this issue is commonly ignored.

²¹The overall time-series dimension is reduced to $T - 1$ since in a model with a lagged dependent variable like in (12) we only have $T - 1$ equations available for estimation.

correlations between contemporaneous y and ξ components.²² Thus the fixed effects estimator is biased in the dynamic regression equation as long as T remains small.²³ Despite their bias in dynamic empirical equations of short T , it is nevertheless informative to investigate results for both the OLS and Fixed effects estimators, since the anticipated pattern across the estimates creates upper and lower bounds for the ‘true’ parameter value, such that $\hat{\rho}_{WG} < \rho < \hat{\rho}_{OLS}$. A number of bias-adjusted fixed effects estimators for dynamic panels are available in the literature, e.g. Kiviet’s (2005) bias-corrected LSDV,²⁴ although bias corrections to the short-run coefficients may worsen the bias of the long-run coefficient.

As our section headline suggests, this class of estimators seeks to use past values of the regressors (now including the lagged dependent variable) as instruments to address the transmission bias, building on explicit assumptions about the unobserved components of equation (12): α^* and e_{jt} — like in the static OLS and FE models we can account for the presence of common shocks γ_t^* by including year dummies in the dynamic estimation equation. The information about valid instruments can be expressed in so-called ‘moment restrictions’,²⁵ which represent the basis for the GMM estimation approach developed by Hansen (1992) — more details with reference to the AB GMM below. The availability of these moment conditions depends on the assumptions made about the unobserved error components and the righthand-side variables y_{jt-1} , l_{jt} and k_{jt} . The following assumptions are commonly employed:

- (i) the productivity level α_j^* is correlated with the input choices l_{jt} and $k_{jt} \forall t$, such that firms with higher productivity levels are assumed to have higher levels of labour and capital inputs; lagged output y_{jt-1} is correlated with productivity levels α_j^* by construction;
- (ii) the productivity shock ξ_{jt} is serially uncorrelated and furthermore uncorrelated with the productivity level α_j^* ;²⁶
- (iii) the productivity shock ξ_{jt} is uncorrelated with input choices made *before* time t . Given the timing assumption about capital stock (input *choice* of capital ‘investment’ is made in period $t - 1$ but capital stock does not increase until period t) the latter implies that the productivity shock ξ_{jt} is uncorrelated with actual capital stock k_{jt} in period t ;
- (iv) the initial conditions for output y_{j1} are uncorrelated with later ξ_{jt} .

Hence, l_{jt} and k_{jt} are assumed to be endogenous with respect to firm productivity levels, l_{jt} to be predetermined and k_{jt} to be endogenous with respect to ξ_{jt} , and the productivity shock ξ_{jt} to be strictly exogenous with respect to the productivity level. The first observation of output y_{j1} is predetermined with respect to the productivity shock ξ_{jt} . These assumptions translate into moment restrictions which act as summary explanations of how the parameters of interest $\beta^* = \{\beta_l, \beta_k\}$ ²⁷ are identified.²⁸ The latter are formulated with the underlying DGP in mind

²²The reason for this is that in the latter case *both* terms are always deflated by $(T - 1)$.

²³A frequently quoted paper by Judson and Owen (1999) employs Monte Carlo simulations to indicate that this bias can still be quite substantial even for moderately long T — this paper however investigates macro panel data. For balanced micro panel data, Bun and Kiviet (2003) discuss the size of the bias in more detail, while Bruno (2005) investigates the same in unbalanced panels.

²⁴The estimator is coded in `Stata` as `-xtlsdvc-`.

²⁵These are also referred to as ‘orthogonality conditions’ or ‘moment conditions’.

²⁶The latter assumption is redundant in the AB approach.

²⁷Strictly speaking, we have π_1 to π_5 as parameters of interest.

²⁸To recapitulate the discussion in Section 2.1: a parameter is unidentified if we cannot get to the ‘true’

and are thus ‘population moment restrictions’. The GMM estimator is then obtained by “finding the element of the parameter space that sets linear combinations of the sample cross products as close to zero as possible” (Hansen, 1982: p.1029), i.e., by means of an optimisation involving the sample equivalents of the moment restrictions which in turn are functions of the parameter estimates $\hat{\beta}^*$.

With reference to equation (11) a generic sample equivalent for the set of r moment conditions can be written as

$$b_N(\beta^*) = \frac{1}{N} \left(\sum_j Z_j' \xi_j(\beta^*) \right) \quad (16)$$

where we stacked variables over time such that Z_j is $T \times k$ for k ‘instruments’ (we detail below what constitutes these). For notational ease we assume no measurement error in this exposition. When formally written as a minimisation problem equation (16) is also referred to as the ‘criterion function’. If our number of moment restrictions r is equal to the number of endogenous variables q we obtain a ‘just identified’ solution (estimate) for β^* by minimising $b_N(\beta^*)$. In the ‘overidentified’ case where more moment restrictions are available for the same number of parameters ($r > q$) the solution is obtained not by minimising $b_N(\beta^*)$ but a weighted version of its square (‘quadratic distance’)

$$\hat{\beta}_{GMM}^* = \arg \min_{\beta^*} J_N(\beta^*) \quad \text{where } J_N(\beta^*) = b_N(\beta^*)' W_N b_N(\beta^*) \quad (17)$$

Now a ‘family of GMM estimators’ using the same moment conditions but different ‘weight matrices’ W_N are available, all of which lead to consistent estimates for β^* , but with different efficiency. This gives rise to a ‘2-step’ version as well as a ‘1-step’ version to determine the efficient estimator: in the former we proceed by first estimating a consistent $\hat{\beta}^*$ using some weight matrix, which yields consistent residuals $\hat{\xi}_j$. These are then employed to construct the *optimal* weight matrix

$$W_N = \left(\frac{1}{N} \sum_j Z_j \hat{\xi}_j \hat{\xi}_j' Z_j' \right)^{-1} \quad (18)$$

Note that in finite samples we introduce variation in the first step by having to estimate W_N which makes inference in the second step unreliable (Pagan’s (1984) generated regressor problem). Windmeijer (2005) provides a finite sample correction for this which is commonly used in practice. The ‘1-step’ version assumes homoskedastic error terms and therefore can apply a ‘known’ weight matrix,²⁹ thus dispensing with the first step where W_N is estimated. Note further that alternative transformations to those applied in the following, as well as quadratic moment restrictions and further assumption about variable and error properties create a host of additional moment conditions which can be applied in estimation.

We introduced $J_N(\beta^*)$ not merely for a more detailed discussion of the derivation, but also because J_N forms the basis for the tests of overidentifying restrictions referred to as Hansen, Sargan or J tests as well as the Difference Sargan test,³⁰ which can be employed to test for the

parameter value due to for instance endogeneity issues. In the present case, if k_{jt} is some function of productivity ω_{jt} , then $\mathbb{E}[k_{jt}\omega_{jt}] \neq 0$ and the $\hat{\beta}_k$ estimate from a standard OLS regression picks up a combination of the ‘true’ effect of capital stock *and* the effect of ω_{jt} .

²⁹See Hsiao (2003, p.88) for details.

³⁰Subtle differences apply: at least within **Stata** ‘Sargan’ refers to the test applied to the minimised J_N of the one-step version, whereas Hansen or Hansen- J refers to the minimised J_N of the two-step version, with the latter

validity of subsets of moment conditions.

We now discuss the two most popular GMM production function estimators in some more detail. By allowing unobserved firm-specific effects to be correlated with inputs, it is possible to take account of time-invariant unobserved heterogeneity. With reference to the AB ‘Difference GMM’, our previous section already indicated that the time-invariant productivity levels η_j (in the dynamic equation contained in $\tilde{\alpha}_j$) can be dropped from the model if we specify a production function equation in first differences

$$\Delta y_{jt} = \pi_1 \Delta y_{jt-1} + \pi_2 \Delta l_{jt} + \pi_3 \Delta l_{jt-1} + \pi_4 \Delta k_{jt} + \pi_5 \Delta k_{jt-1} + \Delta \gamma_t^* + \Delta e_{jt} \quad (19)$$

thus making assumption (i) above redundant since α_j^* is no longer contained in the estimation equation. Similarly for the second part of assumption (ii). In the AB estimator lagged *levels* of the covariates (say from time $t - 2$) act as instruments for the regressors in *differences* at time t . The first-difference moment conditions for endogenous labour and the lagged dependent variable are then

$$\mathbb{E}[y_{jt-s} \Delta e_{jt}] = 0 \quad \text{for } t = 3, \dots, T \text{ and } s \geq 2 \quad (20)$$

$$\mathbb{E}[l_{jt-s} \Delta e_{jt}] = 0 \quad \text{for } t = 2, \dots, T \text{ and } s \geq 2 \quad (21)$$

Assuming that capital is predetermined yields

$$\mathbb{E}[k_{jt-s} \Delta e_{jt}] = 0 \quad \text{for } t = 2, \dots, T \text{ and } s \geq 1 \quad (22)$$

This suggests that y_{jt-3}, l_{jt-2} and k_{jt-1} and *earlier* realisations can be used as instruments for $\Delta y_{jt-1}, \Delta l_{jt}$ and Δk_{jt} in Equation (19) — in the present setup we thus would require a minimum of $T_j = 4$. Note that the ‘quasi-fixed’ nature of capital is not a *requirement* here, but an assumption: instead of predetermined, capital stock can be assumed to be endogenous, such that productivity shocks in period t affect capital stock in the same period. The moment restrictions for capital can then be written as above but for $s \geq 2$. Recall that the timing of the firm’s choice of capital stock levels (investment, disinvestment) *after* the observation of the firm-specific productivity shock ω_{jt} created the endogeneity problem leading to transmission bias. Now, assuming that stock variables such as output, labour or capital stock are sufficiently persistent over time to satisfy the assumption of informativeness, the econometrician can argue that the firm’s labour and capital stock values at time $t - 2$ were chosen *prior* to the observation of the productivity shock at time $t - 1$, thus making them valid instruments. In order to increase the precision of the estimator and allow for the application of testing procedures informing about instrument validity it is common practice to employ not just one lagged value for each time-series observation of each variable, but a whole set of lagged values. In practice the informativeness of these lagged instruments deteriorates and the estimation algorithm therefore employs optimal weights to each of the moment restrictions employed. The ‘overfitting bias’ arising from use of too many moment restrictions is typically not an issue in short T panels.³¹

robust to serial correlation and heteroskedasticity in the residuals.

³¹Since moment conditions rise quadratically in number with T some datasets may need to be examined for this phenomenon. A tell-tale sign of ‘overfitting’ is that the p -value for the test for joint instrument validity tends toward unity (Bowsher, 2002). Commonly a restriction of the instrument set to more recent lags, as well as ‘collapsing’ the instrument matrix (Roodman, 2009) are common strategies to limit the impact of weak

If the output, labour and capital stock variables are highly persistent, the ‘Difference GMM’ instrumentation approach suffers from ‘weak instruments’ and yields substantial finite-sample bias (Griliches and Mairesse, 1995): the lagged levels are poor determinants of the differences which are dominated by the random element.³² Blundell and Bond (1998) [henceforth BB] develop the ‘System GMM’ estimator which is based on the same principles as the Difference GMM but avoids the latter’s bias when variables are highly persistent. Based on the assumption that variables in first-differences are uncorrelated with unobserved firm-specific effects, there are additional moment conditions

$$\mathbb{E}[\Delta k_{jt-s}(\alpha_j^* + e_{jt})] = 0 \quad (23)$$

$$\mathbb{E}[\Delta l_{jt-s}(\alpha_j^* + e_{jt})] = 0 \quad (24)$$

for $s \geq 1$. This implies that appropriately lagged variables in first difference can act as instruments in the original, untransformed production function in levels, equation (9). The validity of the additional moment conditions depends on the assumptions about the initial conditions of the variables y_{j1}, l_{j1}, k_{j1} as well as the absence of serial correlation in the errors. In general, it is well-known that more moment conditions cannot harm efficiency, although in many circumstances additional restrictions can be shown to be ‘redundant’, in the sense that they do not increase the asymptotic efficiency of estimation for the parameter(s) of interest (Breusch et al., 1999).

The ‘System GMM’ estimator requires stationary initial conditions:

$$\mathbb{E}[\Delta y_{j2} \alpha_j^*] = 0 \quad (25)$$

for $j = 1, \dots, N$, i.e., whatever drives the initial value y_{j1} , it is unaffected by the fixed effect α_j^* . This also implies that the economy is in long-run equilibrium. Similarly for l_{jt} and k_{jt} . The joint use of difference and level moment conditions results in a ‘system’ of equations and gives rise to the System GMM estimator, which Blundell and Bond (2000) have shown in Monte Carlo simulations to yield consistent estimates of β_l and β_k . However, Bun and Windmeijer (2009) question whether the System GMM estimator indeed solves the weak instrument problem of the AB estimator: in circumstances in which the variance of firm-specific unobserved heterogeneity is large relative to the variance of the idiosyncratic productivity shocks, instruments for the levels equation are weak, resulting in a worse performance of the BB relative to the AB estimator for given persistence in the data (ρ). Moreover, the problem of losing valuable variation in the data due to first-differencing equally applies to the AB and BB estimators (Wooldridge, 2009). Finally, the instruments added in the System GMM estimator, lagged differences instrumenting for levels, may not be uncorrelated with the fixed effect component of the error term, depending on the assumption made regarding the structure of the endogeneity problem (Greenstreet, 2007).

One practical advantage of the above GMM estimators over the ‘structural’ estimators discussed below is unfortunately often missed in empirical applications, namely that all key spec-

instruments reducing the power of the Sargan test.

³²Take y_t as highly persistent, i.e. $y_t = \varrho y_{t-1} + \zeta_t$ with ϱ close to unity. The first difference of this variable $y_t - y_{t-1} = \Delta y_t = (\varrho - 1)y_{t-1} + \zeta_t$ is then dominated by the ζ_t component, whereas the y_{t-1} element is discounted heavily by $(\varrho - 1)$. Rewriting this as a function of y_{t-1} yields $\Delta y_t = (\varrho - 1)(\varrho y_{t-2} + \zeta_{t-1}) + \zeta_t$: if for instance $\varrho = 0.95$ then the coefficient on y_{t-2} would be $(\varrho - 1)\varrho = -0.0475$ which would mean this variable is drowned out by the random ‘noise’ of the two ζ_t components.

ification assumptions made can be tested in the data. Different sets of assumptions lead to different sets of moment restrictions, which can be individually tested for validity, with the sample size being the only constraint since all tests are valid asymptotically.³³ Thus major assumptions relating to the correlation between factor inputs and productivity fixed effects η_j , as well as pertaining to the strictly exogenous, predetermined or endogenous nature of factor inputs with respect to the productivity shock ξ_{jt} and the presence of measurement error v_{jt} *can and should* be investigated by the researcher.

The use of past differences and levels of input as instruments for current input choices can be justified from a theoretical point of view as shown in Bond and Söderbom (2006) [henceforth BS], providing a ‘structural’ motivation for the dynamic panel estimators. BS show that input parameters are identified when gross output is used as a measure for firm output based on the assumption that all inputs are subject to adjustment costs which differ across firms. The underlying intuition is that individual firms’ optimal change in their input choice is different for inputs with different adjustment costs. This differentiated response generates independent variation across firms in the levels of chosen inputs and therefore allows identification. BS point out that this variation can be predicted from lagged information if inputs are assumed to be dynamic, allowing estimation based on IVs and in a panel setting the use of Difference and System GMM estimators. Past levels contain information on current levels because adjustment costs introduce dependence of current input levels on past realisations of productivity shocks. Importantly, this is a purely structural assumption and does not require any specific functional form for adjustment costs. To see this more clearly, consider a firm’s intertemporal maximization problem described by the following Bellman Equation

$$V_{jt}(k_{jt}, l_{jt}) \equiv \max\{\pi(k_{jt}, l_{jt}) - c_k(k_{jt}, i_{jt}) - c_l(l_{jt}, h_{jt}) + \theta \mathbb{E}[V_{t+1}(k_{jt+1}, l_{jt+1}) \mid k_{jt}, l_{jt}, i_{jt}, h_{jt}]\} \quad (26)$$

where $c_k(\cdot)$ and $c_l(\cdot)$ are cost functions for capital and labour respectively, h_{jt} gross hiring, i_{jt} gross fixed capital formation (investment) and θ is a (common) discount factor. Crucially, BS assume both capital and labour to be dynamic inputs for all firms

$$k_t = (1 - \delta)k_{t-1} + i_t \quad (27)$$

$$l_t = (1 - q)l_{t-1} + h_t \quad (28)$$

where δ and q denote depreciation and labour drop-out respectively, which are determined exogenously. All firms face exogenous prices which are the same across firms. The equations of motion (27) and (28) can be used to replace the respective inputs in (26). Differentiating the resulting expression with respect to control and state variables gives four first order equations. Denoting shadow values of capital and labour as $\lambda_t^k = \frac{1}{1-\delta} \frac{\partial V_t(\cdot)}{\partial k_{t-1}}$ and $\lambda_t^l = \frac{1}{1-q} \frac{\partial V_t(\cdot)}{\partial l_{t-1}}$ respectively, the first order conditions can be used to obtain

$$\frac{\partial \pi_t}{\partial k_t} - r_t \frac{\partial c_k}{\partial k_t} = \lambda_t^k \left[1 - (1 - \delta)\theta \mathbb{E}_t \left(\frac{\partial \lambda_{t+1}^k}{\partial \lambda_t^k} \right) \right] \quad (29)$$

³³In **Stata** this is enabled in the `xtabond2` command by allowing for sets of `gmm(\cdot)` entries to refer to different moment conditions.

$$\frac{\partial \pi_t}{\partial l_t} - w_t \frac{\partial c_l}{\partial l_t} = w_t + \lambda_t^l \left[1 - (1 - q)\theta \mathbb{E}_t \left(\frac{\partial \lambda_{t+1}^l}{\partial \lambda_t^l} \right) \right] \quad (30)$$

Equations (29) and (30) show that in the absence of adjustment costs, the shadow price of capital would equal its price r_t and the shadow price for labour would be zero. The dynamic component becomes evident through the term $\left(\frac{\partial \lambda_{t+1}^i}{\partial \lambda_t^i}\right)$ with $i \in [k, l]$, as this represents current and future realisations of the shadow value. In addition (29) and (30) imply that the shadow values depend on the unobserved productivity shock through the profit function $\pi(\cdot)$. Hence, adjustment costs introduce variation in the observed input levels across firms due to the dynamic characteristic of inputs and their dependence on the contemporaneous and past unobserved productivity shocks *without* imposing any requirement of serial correlation on the productivity shock itself. Obviously, the fundamental simultaneity problem is still present and current input levels still depend on the current unobserved productivity shock. What is gained from the presence of adjustment costs, however, is variation within and across firms in their input levels for capital and labour. BS argue that the System GMM estimator discussed above is able to identify the parameters because of the dynamic implications of adjustment costs as evident from (29) and (30). Hence, the crucial assumption to achieve identification through the System GMM estimator is the dynamic nature of the input variables. This implies that past input levels are informative instruments for current levels. These instruments should also be valid as past levels should be uncorrelated with current realizations of the productivity shock.

2.2.4 ‘Structural’ estimators (i): Olley and Pakes (1996)

Before these recent theoretical foundations in the case of production functions justified the use of lagged values as instruments, the GMM estimators were *essentially* based on assumptions about the evolution of observable and unobservable processes over time and not explicitly derived from a structural model of firm behaviour. Olley and Pakes (1996) were the first in a line of authors adopting an explicit model for the firm’s optimization problem to derive their production function estimator. Put simply, the trick employed by OP to side-step the endogeneity problem in the production function equation is to use information about *observed* investment i_{jt} to proxy for *unobserved* productivity ω_{jt} and to apply a control function estimator.

OP assume that k_{jt} and ω_{jt} are firm-specific state variables in the firm’s dynamic programming problem. The Bellman equation corresponding to the firm’s dynamic programming problem is written as

$$V_{jt}(k_{jt}, \omega_{jt}) \equiv \max\{\pi_{jt}(k_{jt}, \omega_{jt}) - c_j(i_{jt}) + \theta \mathbb{E}[V_{t+1}(k_{jt+1}, \omega_{jt+1})|k_{jt}, \omega_{jt}, i_{jt}]\} \quad (31)$$

where $k_{jt+1} = (1 - \delta)k_{jt} + i_{jt}$ is the law of motion for capital accumulation, δ is the depreciation rate and i_{jt} denotes firm j ’s investment in physical capital at time t . Investment is chosen at time t and adds to the capital stock at time $t + 1$.³⁴ The solution to the Bellman equation in (31) gives an investment policy function that depends on (unobserved) productivity as well as physical capital, $i_{jt}(k_{jt}, \omega_{jt})$. Labour is left out of the investment equation because it is a ‘non-dynamic’ input *by assumption*, implying that the levels of labour input chosen today do not affect the cost of labour (and therefore profits) tomorrow: labour is a fully flexible factor input, which can be adjusted ‘instantaneously’ (within period t) after observing ω_{jt} . Crucially, firm investment

³⁴This assumption, i.e., the dynamic nature of capital input, is used to identify the capital coefficient β_k .

is assumed to be strictly increasing in both capital stock and productivity (referred to as the ‘monotonicity assumption’), implying that *ceteris paribus* firms with higher k_{jt} or ω_{jt} *always* invest more. Furthermore, it is assumed that ω_{jt} is *the only unobservable* driving the investment decision, a condition referred to as ‘scalar unobservable’. Akerberg et al. (2006) note that the scalar unobservable assumption rules out unobserved differences in investment *prices* across firms, other unobserved state variables, and any other unobserved factors influencing investment but not production. The scalar unobservable assumption also rules out differential adjustment costs across firms, which is somewhat questionable given the suggestions by BS. Finally, when making investment decisions in period $t + 1$ any realizations of ω_{jt} prior to time t are not included in the investment function due to the OP assumption that productivity evolves according to an ‘exogenous first-order Markov process’, meaning that a firm builds expectations about its productivity at time $t + 1$ solely based on its productivity levels observed at time t .³⁵ Thus we can assume most generally that productivity evolves following $\omega_{jt} = g(\omega_{jt-1}) + \xi_{jt}$, where ξ_{jt} is the random ‘productivity shock’. Given some of the earlier assumptions made, Akerberg et al. (2006) remark that a first-order Markov process is indeed the *only* possible evolution process for productivity in the OP framework.³⁶ Conditional on functional form restrictions (the investment function is continuous in k_{jt} and ω_{jt}) and provided investment is positive ($i_{jt} > 0$) the investment equation can be inverted to yield an expression for productivity ω_{jt} as a function of the state variable, physical capital, and the control variable, investment: $\omega_{jt} = f_t(i_{jt}, k_{jt})$. Thus conditional on a (considerable) host of theoretically motivated (‘structural’) assumptions and provided the implementation is sufficiently flexible (specify $f_t(\cdot)$ as nonlinear function) this affords us a representation for unobserved productivity based on observable variables.

In practice the OP production function estimator is implemented in two steps: first, by regressing output y_{jt} on labour input l_{jt} and a nonparametric function ϕ of (k_{jt}, i_{jt}) proxying for firm-specific productivity

$$y_{jt} = \beta_l l_{jt} + \phi_{jt}(i_{jt}, k_{jt}) + \epsilon_{jt} \quad (32)$$

where

$$\phi_{jt}(i_{jt}, k_{jt}) = \beta_o + \beta_k k_{jt} + f_t(i_{jt}, k_{jt}) \quad (33)$$

represents the average productivity level, capital input and the inverted investment function proxying for ω_{jt} . $f_t(\cdot)$ is indexed by time to take account of changes in state variables i_{jt}, k_{jt} due to changing factor prices or market structure *common to all firms* in each time period. This can account for the common shocks γ_t in our general framework in equation (9). Labour l_{jt} is assumed to be exogenous with respect to ϵ_{jt} and is unbiased since the unobserved productivity, with which it is contemporaneously correlated, is represented via $\phi_{jt}(\cdot)$. Equation (32) is a ‘partially-linear’ equation which can be estimated using semiparametric methods (Robinson, 1988). OP suggest estimation based on a third-order polynomial series expansion, but any other flexible estimation approach would be equally valid. In order to allow for the time-*variation* in ϕ_{jt} , in practice, the polynomial terms need to be interacted with time. The first stage regression

³⁵The ‘exogenous’ label simply means that productivity is treated as determined by factors outside the model, i.e., like manna from heaven. More formally, $F_\omega \equiv F(\omega_{jt}|\omega_{jt-1}), \omega_j \in \Omega$, where $F(\omega_{jt}|\omega_{jt-1})$ is stochastically increasing.

³⁶See their paper for a discussion of how to relax this assumption.

solves the simultaneity problem between k and ω_{jt} and yields unbiased and consistent estimates $\hat{\beta}_l$ for the labour coefficient and $\hat{\phi}_{jt}(i_{jt}, k_{jt})$ for the nonlinear function accounting for capital stock and productivity.

In the second step, OP use these estimates to run a regression of $y_{jt} - \hat{\beta}_l l_{jt}$ on $\hat{\phi}_{jt}(\cdot)$ and k_{jt} , which yields an unbiased estimate $\hat{\beta}_k$ for the capital coefficient. The rationale for this can be developed as follows: at time t , from the econometrician's perspective, capital stock k is predetermined and thus deterministic, since the investment decision was made *in the previous period*. Similarly for average productivity across firms. Expected output at time t can therefore be written as

$$\mathbb{E}[y_{jt} - \beta_l l_{jt} | k_{jt}] = \beta_k k_{jt} + \mathbb{E}[\omega_{jt} | \omega_{jt-1}] \quad (34)$$

From the assumption of a Markov process for productivity and the inverted investment function in (33) above, we can deduce that

$$\mathbb{E}[\omega_{jt} | \omega_{jt-1}] = g(\omega_{jt-1}) + \xi_{jt} = g(\phi_{jt-1}(i_{jt-1}, k_{jt-1}) - \beta_o - \beta_k k_{jt-1}) + \xi_{jt} \quad (35)$$

meaning expected productivity at time t is some function $g(\cdot)$ of observed productivity at time $t-1$. Furthermore, from our first stage we know that $\mathbb{E}[\beta_l] = \hat{\beta}_l$. All of these insights allow us to rewrite (34) as

$$y_{jt} - \hat{\beta}_l l_{jt} = \beta_k k_{jt} + g(\hat{\phi}_{jt-1}(i_{jt-1}, k_{jt-1}) - \beta_o - \beta_k k_{jt-1}) + \xi_{jt} + \epsilon_{jt} \quad (36)$$

recalling that ξ_{jt} is the random shock to productivity ω_{jt} . Due to the fact that ξ_{jt} and l_{jt} are contemporaneously correlated, OP cannot include l_{jt} on the righthand side of the production function to be estimated. Instead they subtract $\hat{\beta}_l l_{jt}$ from their measure of output on the lefthand side using the coefficient estimate obtained in the first stage regression. In summary, capital stock in equation (36) is exogenous with respect to the error term ($\mathbb{E}[k_{jt}(\xi_{jt} + \epsilon_{jt})] = 0$) since ϵ_{jt} is i.i.d. and the level of k_{jt} was determined in the previous period and can therefore not have been affected by the change in productivity in the present period (ξ_{jt}); productivity ω_{jt} is proxied by $g(\cdot) + \xi_{jt}$, where $\mathbb{E}[\xi_{jt}] = 0$; labour input is accounted for and does not cause any endogeneity problems by the output-adjustment on the lefthand side. Thus in the second step equation (36) can be estimated using non-linear least squares (NLLS)³⁷ to obtain an unbiased estimate for β_k — this estimation approach is required due to β_k entering the equation twice and in combination with other parameters. The function $g(\cdot)$ (again implemented via higher order polynomials) is merely acting as a control and its estimate in the second stage is not of interest.³⁸

The OP model can be extended to incorporate firm exit, in which case an additional stage is entered between the two described above, where a probit regression (exit, non-exit) is fitted on a nonlinear function of i_{jt}, k_{jt} using the same argument of proxied productivity as in the OP first stage. The predictions from this intermediate stage are then entered in the $g(\cdot)$ function in the above second stage. This aside, the researcher may choose to investigate gross output

³⁷In **Stata** this is done using the `nl` command.

³⁸Note that this does not represent $\omega_{jt} = f_t(i_{jt}, k_{jt})$, but $\phi_{jt}(i_{jt}, k_{jt}) = \beta_o + \beta_k k_{jt} + f_t(i_{jt}, k_{jt})$, thus is *not* an estimate for firm-level productivity.

instead of value-added, or may add further covariates such as firm age in the empirical equation. All of these can be accommodated in the OP framework, considering in each case whether the variable is non-dynamic (like labour in the present case) or flexible (like capital stock).

Note that one element of the general production function model in equation (9) is notable for its absence in the OP approach: the presence of firm-level fixed effects in productivity η_j . As Griliches and Mairesse (1998, notation adjusted) point out, “net investment i_{jt} would depend primarily on the ‘news’ in ω_{jt} and not on its level”. Thus the OP estimator (and the other structural approaches building on an inverse productivity function) cannot account for these time-invariant firm-level productivity effects, which leads to potential bias in the production function estimates. In practice researchers seek to reduce the impact of this bias by running sector-specific regressions, i.e., by ‘slicing the data’, thus avoiding the impact of the most glaring productivity differences across industries.

Wooldridge (2009) points out that the OP model can be estimated in a single step, which may bring about efficiency gains. Wooldridge suggests to estimate the following equations simultaneously

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + f_1(i_{jt}, k_{jt}) + \epsilon_{jt} \quad (37)$$

and

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + f_2(i_{jt-1}, k_{jt-1}) + (\xi_{jt} + \epsilon_{jt}) \quad (38)$$

where $f_i(\cdot)$ ($i \in (1, 2)$) are unknown (non-linear) functions of the instruments: $f_1(i_{jt}, k_{jt})$ has been defined above and $f_2(\cdot)$ is defined as $f_2(i_{jt-1}, k_{jt-1}) = \mathbb{E}[\omega_{jt} | \omega_{jt-1}] = \omega_{jt} + \xi_{jt}$. The two corresponding population moment conditions are defined as

$$\mathbb{E}[(\xi_{jt} + \epsilon_{jt}) | f_1(k_{jt}, l_{jt}, i_{jt}, k_{jt-1}, l_{jt-1}, i_{jt-1}, \dots, k_{j1}, l_{j1}, i_{j1})] = 0 \quad (39)$$

and

$$\mathbb{E}[(\xi_{jt} + \epsilon_{jt}) | f_2(k_{jt}, k_{jt-1}, l_{jt-1}, i_{jt-1}, \dots, k_{j1}, l_{j1}, i_{j1})] = 0 \quad (40)$$

for $t = 1, 2, \dots, T$. Hence, the two moment restrictions are written in two equations with the same dependent variable but different instrument sets. In Equation (37), capital, labour, and investment, their lagged values and any (non-linear) functions of these variables enter the instrument set, whereas in Equation (38) contemporaneous labour and investment are excluded from the instrument set.

The main advantage of this one-step approach is that standard errors can be obtained relying on the standard GMM framework. This advantage may be important because in the original OP framework, inference is based on bootstrap standard errors which may be unreasonably large in the case of strongly unbalanced datasets (since bootstrapping requires sampling with replacement) — as we will also discuss in our empirical application in Section 3. It also improves

efficiency of the estimates because contrary to the two-step estimator, it uses information from correlation *across* the two equations and furthermore an optimal weighting matrix accounts for serial correlation and heteroscedasticity. Moreover, it allows for straightforward testing of overidentification restrictions (validity of the identification assumptions made). Note that this estimation procedure carries over in a straightforward manner to the LP and ACF estimation procedures outlined below.

2.2.5 ‘Structural’ estimators (ii): Levinsohn and Petrin (2003)

Building on OP, Levinsohn and Petrin (2003) proposed the use of intermediate input demand instead of investment demand as a proxy for productivity ω_{jt} . This implies that intermediate inputs are chosen at time t once ω_{jt} is known to the firm. The same applies to labour input choices, which in turn implies that labour and intermediate inputs are chosen simultaneously and labour maintains its assumed non-dynamic/flexible character. The simultaneity in the timing of a firm’s choice of intermediate inputs and labour, however, avoids the need to allow labour to affect the optimal choice of intermediate inputs. Intermediate inputs (e.g. materials, fuels, electricity) are readily available in most firm-level datasets since value-added is commonly constructed from gross output and intermediate inputs (at least as an aggregate) by the researcher.

The use of intermediate inputs as proxy addresses some of the concerns raised against the OP approach, which hinges on the assumptions of scalar unobservable, monotonicity and non-zero investment series, all of which are required for the inversion of the investment equation. In practice, investment data is frequently either missing for a large number of firms or firms report zero investment, which eliminates these observations from the estimation equation. Since we cannot assume that these data are missing at random or that firms randomly opt for zero investment in certain years, the required sample reduction when applying the OP procedure is of at least similar concern to the non-random exit of firms from the market. In the LP procedure, intermediate inputs (electricity, material inputs) are modelled as a function of the two state variables ω_{jt} and k_{jt} similar to the use of investment in the OP approach. Since intermediate inputs are only zero if the firm stops operating, this choice avoids the substantial cull of observations in the OP approach. Furthermore, the monotonicity assumption that firms with larger capital stocks or productivity employ higher levels of intermediate inputs is somewhat more defensible. Note also that the presence of adjustment costs to investment implies that a firm’s optimal investment path is not characterised by a smooth correspondence to productivity shocks since investment is now more lumpy.³⁹ The LP approach continues to rely on the scalar unobservable and monotonicity assumptions. In particular, the monotonicity assumption, i.e., that *ceteris paribus* a firm that experiences a higher ω_{jt} *always* uses more intermediate inputs, may be violated if a more productive firm chooses to cut waste leading it to economise on inputs.

The production function to be estimated by LP is

$$o_{jt} = \beta_o + \beta_l l_{jt} + \beta_k k_{jt} + \beta_m m_{jt} + \omega_{jt} + \epsilon_{jt} \quad (41)$$

where m_{jt} denotes intermediate inputs. Note that LP use (log) gross output o_{jt} as dependent variable, although they also provide a value-added approach in an earlier application of the

³⁹The ‘non-convexity’ of adjustment costs violates the smoothness (monotonicity) assumption imposed by OP.

idea.⁴⁰ LP specify demand for intermediate inputs as

$$m_{jt} = m_{jt}(k_{jt}, \omega_{jt}) \quad (42)$$

where demand is required to be monotonically increasing in ω_{jt} , similar to the assumption by OP regarding investment. This assumption allows to invert the function to obtain a proxy for unobserved productivity

$$\omega_{jt} = f_t(k_{jt}, m_{jt}) \quad (43)$$

The first stage of the production function is then rewritten as

$$o_{jt} = \beta_l l_{jt} + \zeta_{jt}(k_{jt}, m_{jt}) + \epsilon_{jt} \quad (44)$$

where

$$\zeta_{jt}(k_{jt}, m_{jt}) = \beta_o + \beta_k k_{jt} + \beta_m m_{jt} + f_t(k_{jt}, m_{jt}) \quad (45)$$

This equation is empirically implemented in the same way as the first stage regression in OP. Given consistent estimates $\hat{\beta}_l$ and $\hat{\zeta}_{jt}$, LP proceed in the same fashion as OP to identify β_k in the second stage

$$o_{jt} - \hat{\beta}_l l_{jt} = \beta_k k_{jt} + \beta_m m_{jt} + g(\hat{\zeta}_{jt-1} - \beta_o - \beta_k k_{jt-1} - \beta_m m_{jt-1}) + \xi_{jt} + \epsilon_{jt} \quad (46)$$

In addition to β_k the LP procedure also needs to estimate the β_m coefficient in the second stage. Since m_{jt} is not orthogonal with respect to ξ_{jt} , LP instrument current intermediate input levels m_{jt} through one-period lagged levels m_{jt-1} . All of this is formulated within a GMM framework, such that LP come up with two independent population moment restrictions with respect to k and m

$$\mathbb{E}[k_{jt}(\xi_{jt} + \epsilon_{jt})] = \mathbb{E}[k_{jt}\xi_{jt}] = 0 \quad (47)$$

$$\mathbb{E}[m_{jt-1}(\xi_{jt} + \epsilon_{jt})] = \mathbb{E}[m_{jt-1}\xi_{jt}] = 0 \quad (48)$$

where the second moment restriction spells out the instrumentation strategy for m_{jt} in the second stage. In their empirical implementation LP furthermore specify a set of overidentifying restrictions for the second stage regression, using lags of all input variables (in case of intermediate inputs this implies m_{jt-2}) — this is done to increase efficiency of the estimator and to allow for the testing of overidentifying restrictions. Yet, it is important to bear in mind that adding further lags reduces the sample size by limiting the available time series.

The underlying assumption is that lagged levels of a firm's intermediate inputs are informative instruments for current levels. This is likely to be the case despite the 'perfectly variable' nature of intermediate inputs, for example, if there is some persistence in m , i.e., $m_{jt} = \rho m_{jt-1} + v_{jt}$.

In summary, OP and LP achieve identification of β_l through specific *structural* assumptions. OP assume that labour input is non-dynamic, i.e., a firm's choice of l at time $t - 1$ does not affect the cost of input use in time t . This assumption is required as otherwise a firm's choice of

⁴⁰Levinsohn and Petrin (1999) specify a value-added model where intermediate inputs are only used to proxy for unobserved productivity, thus only enter the inverted material demand equation $\omega_{jt} = f_t(k_{jt}, m_{jt})$. No instrumentation of m_{jt} is necessary in the second stage since this variable does not enter the regression.

labour in $t - 1$ would affect its choice of investment and hence identification of l_{jt} would not be achieved from using the inverse investment function as a proxy for ω_{jt} . In the case of LP, labour input is also non-dynamic, but more crucially, l_{jt} and m_{jt} are assumed to be ‘perfectly variable’ inputs, i.e., they are chosen by the firm once ω_{jt} is observed (but within the same time-period). If intermediate inputs were chosen *before* observing ω_{jt} the inversion of the intermediate goods demand function would not be possible. Identification of k_{jt} is also achieved by using a timing assumption, whereby in both estimators the moment condition $\mathbb{E}[k_{jt}(\xi_{jt} + \epsilon_{jt})] = 0$ holds if capital is ‘quasi-fixed’, i.e., if its level is determined *prior* to observing the productivity shock. In the LP case with gross output as the dependent variable, intermediate inputs m_{jt} are identified using once-lagged level values as instruments, due to contemporaneous correlation between m_{jt} and the productivity innovation ξ_{jt} arising from the ‘perfectly variable’ characteristic of intermediate inputs.

2.2.6 ‘Structural’ estimators (iii): Akerberg, Caves and Frazer (2006)

ACF point out that the first stage in the OP and LP procedures fails to identify the labour coefficient β_l except under *very* special assumptions.⁴¹ The reason for this is that labour demand l_{jt} is a function of the same state variable ω_{jt} as investment or intermediate inputs and therefore does not vary *independently* from the inverted investment or intermediate input functions used to proxy for unobserved productivity.⁴² In the respective first stage regressions

$$y_{jt} = \beta_l l_{jt} + \{\beta_o + \beta_k k_{jt} + f_t(i_{jt}, k_{jt})\} + \epsilon_{jt} \quad (49)$$

$$o_{jt} = \beta_l l_{jt} + \{\beta_o + \beta_k k_{jt} + \beta_m m_{jt} + f_t(k_{jt}, m_{jt})\} + \epsilon_{jt} \quad (50)$$

labour input l_{jt} and the nonparametric functions proxying unobserved productivity, $f_t(i_{jt}, k_{jt})$ for OP and $f_t(m_{jt}, k_{jt})$ for LP, are perfectly collinear.⁴³ Collinearity in this context prevents identification because even the population distributions of labour input and the productivity proxy do not vary independently from each other. Hence, increasing sample size does not mitigate the problem. In the LP approach this implies that l_{jt} is a time-varying function in intermediate inputs and capital. Obviously, this in turn implies that l_{jt} does *not* vary independently from $f_t(m_{jt}, k_{jt})$, hence β_l is not identified. ACF discuss a number of alternative DGPs for labour demand l_{jt} which may potentially ‘save’ LP’s identification strategy, but conclude that there is no credible DGP available given the restriction imposed by the assumption of ‘perfect variability’ of labour and intermediate inputs. Similarly for the OP approach: β_l is not identified since l_{jt} is collinear with the nonparametric function $f_t(i_{jt}, k_{jt})$. If labour input demand is indeed a

⁴¹For the LP procedure, there are two alternative sets of assumptions that allow identification: first, intermediate inputs are chosen *before* labour input but while productivity levels remain constant. At the same time, firms experience an i.i.d. price or demand shock which is serially uncorrelated, has sufficient variation across firms and shifts labour input choices. Second, there must be serially correlated optimisation error in the measure for labour while no measurement error is allowed in the measure of intermediate inputs. For OP, there is an additional — somewhat more realistic — scenario that allows identification: firms choose labour based on an imperfect guess of their productivity levels before they fully learn about the realisation of the productivity shock.

⁴²BS make a similar, albeit more sweeping point with regard to production functions that use gross output instead of value-added. They show that in Cobb-Douglas production functions parameters of flexible non-dynamic inputs (labour, intermediate inputs) are not identified through cross-section variation if all firms face common input prices and inputs are chosen optimally.

⁴³More generally in the BS critique: if firms optimally choose ‘perfectly variable’ inputs in a perfectly competitive setting, the input parameters are not identified due to perfect collinearity in all inputs. The ACF critique applies to a special case of the BS critique, i.e., to a situation in which one input (k) is dynamic and therefore not ‘perfectly variable’, but the others are not.

function of both state variables $l_{jt} = g_t(\omega_{jt}, k_{jt})$, then perfect collinearity with $f_t(i_{jt}, k_{jt})$ prevents identification of β_l .

ACF's rescue strategy for the structural approach is based on acknowledging the 'perfectly variable' characteristic of labour, but turning it into an '*almost* perfectly variable' characteristic: the trick applied for identification is to assume firms chose l_{jt} at time $t-b$ ($0 < b < 1$). This means the firm chooses labour input *after* capital stock k_{jt} was determined at $t-1$ (via investment choice i_{jt-1}) but *before* intermediate inputs m_{jt} are chosen at time t . The first-order Markov process describing productivity evolution is then defined *between the three subperiods*, $t-1$, $t-b$ and t , namely

$$p(\omega_{jt}|I_{jt-b}) = p(\omega_{jt}|\omega_{jt-b}) \quad (51)$$

$$p(\omega_{jt}|I_{jt-1}) = p(\omega_{jt}|\omega_{jt-1}) \quad (52)$$

where I_{jt-b} and I_{jt-1} denote the 'information sets' available to firm j at time $t-b$ and $t-1$ respectively and $p(\cdot)$ is some general functional form. This implies that l_{jt} now enters the demand function for intermediate inputs as a state variable:

$$m_{jt} = h_t(\omega_{jt}, k_{jt}, l_{jt}) \quad (53)$$

This function can be inverted if m_{jt} is strictly increasing in ω_{jt} and there are no other unobserved processes driving the demand for intermediate inputs (scalar unobservable). The first stage uses the inverse function of (53) to control for unobserved productivity in the gross-output production function

$$o_{jt} = \beta_o + \beta_k k_{jt} + \beta_l l_{jt} + h_t^{-1}(m_{jt}, k_{jt}, l_{jt}) + \epsilon_{jt} \quad (54)$$

Since the function for intermediate inputs $h_t(\cdot)$ and therefore also its inverse for ω_{jt} contains l_{jt} , identification of β_l is not possible in the first stage. Note the subtle difference in the identification problem here compared with that in the LP first stage: identification in the latter was not possible because labour was determined by the same state variable as intermediate inputs; in the former, identification is not possible because intermediate inputs are also a function of labour, while labour is no longer a function of the same state variables as intermediate inputs due to the explicit timing assumption described above. The purpose of the ACF first stage estimation is to eliminate the portion of output determined by *unanticipated* shocks at time t , measurement error or any other random noise: ϵ_{jt} . Isolating this quantity is useful because initially the econometrician cannot separately identify the productivity shock observed by the firm from the random component of the unobservables. In practice the first stage is implemented by regressing output on a polynomial function of labour, capital and intermediate inputs, like in the two previous structural estimators. The estimated output net of ϵ_{jt} is then simply the residual from the first stage regression

$$o_{jt} - \hat{\epsilon}_{jt} \equiv \hat{\Psi}_t(m_{jt}, k_{jt}, l_{jt}) = \hat{\beta}_k k_{jt} + \hat{\beta}_l l_{jt} + \hat{h}_t^{-1}(m_{jt}, k_{jt}, l_{jt}) \quad (55)$$

which as indicated are simply the predicted values from the first stage regression.

In order to implement stage two and identify the input coefficients, ACF derive two moment conditions. With the moment conditions in mind, unobserved productivity can be specified to

evolve following

$$\omega_{jt} = \mathbb{E}(\omega_{jt}|\omega_{jt-1}) + \xi_{jt} \quad (56)$$

where ξ_{jt} is mean-independent of all ω_{t-1} . Given that k_{jt} and l_{jt-1} are uncorrelated with ω_{jt} by ‘structural’ assumption (capital is predetermined since investment decisions are made in the previous period; labour in the previous period is determined prior to the productivity shock in the current period), two independent population moment conditions can be defined as

$$\mathbb{E}(\xi_{jt}k_{jt}) = 0 \quad (57)$$

$$\mathbb{E}(\xi_{jt}l_{jt-1}) = 0 \quad (58)$$

These moment restrictions summarise the identification strategy for β_l and β_k in the second stage. In order to employ their sample analogues in the second stage regression, ACF recover the distribution of ξ_{jt} , the innovation in firm level productivity, which enters the optimisation problem

$$\min_{\beta^*} \left[\sum_h \sum_t \sum_j \hat{\xi}_{jt}(\beta^*) Z_{jht} \right]' \cdot C \cdot \left[\sum_h \sum_t \sum_j \hat{\xi}_{jt}(\beta^*) Z_{jht} \right] \quad (59)$$

where $\hat{\beta}^* = (\hat{\beta}_k, \hat{\beta}_l)$, $Z_{jht} = [k_{jt} \ l_{jt-1}]'$ and C is an arbitrary 2×2 weight matrix. The function $[\cdot]' \cdot C \cdot [\cdot]$ is once again the ‘criterion function’ of the optimisation problem.

In practice, the optimal values for $\hat{\beta}^*$ and thus $\hat{\xi}_{jt}(\beta^*)$ are obtained by repeated iteration over a sequence of steps: first, assuming some starting values for β_l and β_k ,⁴⁴ say $\beta_l^{\#1}$, $\beta_k^{\#1}$, productivity $\omega_{jt}^{\#1}(\cdot)$ is obtained by computing

$$\omega_{jt}^{\#1}(\beta_k^{\#1}, \beta_l^{\#1}) = \hat{\Psi}_{jt} - \beta_k^{\#1} k_{jt} - \beta_l^{\#1} l_{jt} \quad (60)$$

where it is notable that first stage result $\hat{\Psi}_{jt}$ stays fixed throughout the entire iteration process. In a second step, a suitably general representation of the Markov process in equation (56) is employed to estimate

$$\hat{\omega}_{jt}^{\#1}(\beta_k^{\#1}, \beta_l^{\#1}) = \varphi(\hat{\omega}_{jt-1}^{\#1}(\beta_k^{\#1}, \beta_l^{\#1})) + \xi_{jt}^{\#1} \quad (61)$$

perhaps using nonparametric techniques or a polynomial function for $\varphi(\cdot)$.⁴⁵ The residuals from this regression will be an estimate of the innovation/productivity shocks series $\hat{\xi}_{jt}^{\#1}$. In step 3 this innovation estimate is entered in the criterion equation and the result is evaluated: if it constitutes a global minimum, the iteration process stops and $\hat{\beta}_k^{\#i}$, $\hat{\beta}_l^{\#i}$ are the ACF estimates for the labour and capital coefficients — i here simply indicates the iteration round, but is of no significance. If they do not, we return to the first step, select some different values for β_k , β_l and the sequence of steps continues. This iteration process is carried out using a minimisation routine which iterates until a cutoff point for the global minimum is reached.

Standard errors for the resulting $\hat{\beta}_k$, $\hat{\beta}_l$ can be obtained by bootstrapping, which due to the three-stage process is somewhat more elaborate and time-intensive than usual. As in the case of OP and LP, such multiple-stage estimators can result in large bootstrap standard errors in the case of strongly unbalanced data (see Section 3).

⁴⁴In practice one can simply pick the OLS estimates for the capital and labour coefficients.

⁴⁵This can be implemented in *Stata* by using for example the `locpoly` command.

An important feature of the method suggested by ACF is that it does not rule out labour being a dynamic input. This is possible as a firm’s choice of l_{jt} and k_{jt} would depend on l_{jt-1} while that of intermediate inputs m_{jt} does not, since it is by assumption a ‘perfectly variable’ and non-dynamic input. The persistence generated by allowing labour to be dynamic can also be exploited for identification (e.g. employing further lags of l_{jt} in additional moment conditions). Identification is also aided by the presence of unobservables affecting m_{jt} , but only through the channel of l_{jt} and k_{jt} . For the same reason as why labour can be a dynamic input, these unobservables can also be correlated over time.

The estimation approach proposed by ACF is in fact quite *similar* to that proposed by Arellano and Bond (1991) and Blundell and Bond (1998). The moment conditions used in the dynamic panel framework of the latter rely on setting the orthogonality conditions with respect to the composite error term $\omega_{jt} + \epsilon_{jt}$ equal to zero. In the structural approach by ACF, the first step serves to net out ϵ_{jt} which allows forming moment conditions defined in ω_{jt} . The main *difference* consists in the assumptions about the unobserved productivity shock ω_{jt} . In the dynamic panel framework, ω is treated parametrically and assumed to have a linear functional form. This is its main drawback vis-à-vis the structural approach, in which ω_{jt} is allowed to assume any functional form. At the same time, the dynamic panel approach produces consistent estimates incorporating firm-level fixed effects in ω_{jt} , while this is not possible in the structural framework. A further advantage is that the dynamic panel estimator does not rely on the scalar unobservable assumption in the input demand functions employed in the structural estimators.

2.2.7 Further Alternatives

A number of alternative approaches to the above are available in the literature, but at present have not penetrated the mainstream of applied empirics. We briefly review these and discuss the barriers to wider application.

Yet another structural approach was proposed by Doraszelski and Jaumandreu (2008) [henceforth DJ]. Their goal is to estimate a production function taking into account a firm’s R&D input, thus the estimator is constructed with a more narrowly-defined application than those reviewed above. These authors make the important point that the literature so far has assumed that changes in productivity are exogenous to a firm’s decisions due to the assumption that productivity follows a first-order Markov process. DJ’s approach assumes that R&D is endogenous with respect to a firm’s observed productivity realisations. In order to allow for endogenous R&D in a production function, DJ specify a controlled first-order Markov process according to which productivity at time t is determined by productivity at time $t - 1$ and R&D expenditure at time $t - 1$. Because of the controlled Markov process assumption, the standard OP method cannot be applied as investment in physical capital cannot be inverted to back out unobserved productivity as described above (Buettner, 2005). Instead, DJ invert the labour demand equation (see Equation (7)) to back out productivity. For estimation DJ rely on the sieve estimator devised by Ai and Chen (2003). DJ’s procedure has the convenient property that the functional form of the proxy is entirely known since it is derived from the first order condition of a firm’s profit maximization problem, which makes estimation feasible despite the presence of R&D as an additional endogenous input. Obviously, the estimator is constrained in terms of wider ap-

plication by its focus on R&D in a Griliches (1979)-type ‘knowledge production function’ and the need for data on R&D, a variable frequently not available at the firm-level.⁴⁶

Greenstreet (2007) proposes an alternative way of estimating production functions, formulating a firm’s optimal control problem recursively as a sequential process and applying Kalman filtering in order to obtain a firm’s production function parameters. The structural assumption needed to apply the Kalman filter is that firms themselves hold only imperfect beliefs about their own contemporaneous productivity realisations. Therefore, they take decisions regarding input choice based on imprecise beliefs about present and future productivity. Based on their beliefs and input choices, firms forecast output. At the end of each period, firms learn their actual output as well as productivity realisation which can be simply computed from observed inputs and output. The difference between actual and observed output is the forecast error, which contains information used by the firm to form expectations for the following period’s productivity realisation.

A crucial assumption in this model is that firms and the econometrician use the same observable information in order to predict productivity realisations. Nevertheless, firms are assumed to know the true parameters of the model (β_l, β_k) while the econometrician does not. Also, similarly to ACF, Greenstreet makes an important timing assumption by requiring that input choices are predetermined with respect to the prediction error.⁴⁷ The model relies on normality assumptions of all error terms in the state-space form which is a convenient assumption as the Kalman filter then produces minimum mean squared error estimates.

Greenstreet’s estimator does not rely on proxies obtained from inverting input demand equations and therefore imposes less restrictive structure on firm behavior. One of the advantages of this approach is that no implicit assumption regarding the industry equilibrium is necessary.⁴⁸ Probably the most significant advantage of this estimator over the dynamic panel estimators and the structural estimators of OP/LP/ACF is that the process defining productivity evolution over time can be formulated in an almost arbitrarily flexible way. This avoids the scalar unobservable problem of the structural estimators and offers the possibility to allow other variables such as R&D to influence the evolution of firm-level productivity. Due to the recursive structure of the model, the major drawback of Greenstreet’s approach is that it requires an assumption about initial productivity beliefs, which are assumed to have the same distribution as the DGP producing observed initial productivity for an entrant’s cohort. This implies that the estimation includes only firms which enter the market during the period of observation; furthermore, if firm data exhibits gaps, the estimator is able to estimate the parameters only up to the gap, but none of the later data can be used as the sequential estimator cannot bridge the data gap — both these points have strong implications for data requirements which are unlikely to be fulfilled in standard practical cases.

⁴⁶Moreover, when the production function includes R&D as an additional input, one might also worry about cross-sectional dependence among firms due to the presence of R&D spillovers (Eberhardt et al., 2010). It is not clear how such spillovers can be accommodated in the DJ framework and what the implications of neglecting their potential presence are.

⁴⁷This is a testable assumption as input choices should be independent from information about productivity learnt by firms *after* input decisions have been made. This information can be computed as the difference between firms’ estimated productivity beliefs and the Kalman filter’s smoothed productivity estimate using an establishment’s entire history in the panel.

⁴⁸The estimators discussed previously assumed a specific market equilibrium, i.e., in the case of OP, LP, and ACF, firms find themselves in a Markov Perfect Nash Equilibrium.

Finally, an entirely different approach to the transmission bias problem can be derived from Bai (2009), whose estimator relies on a common factor framework studied primarily in macro-economic datasets (see Coakley et al. (2006) and Eberhardt and Teal (2010) for detailed discussions). In our present production function case, this would imply productivity is modelled as a set of unobserved common factors f_t with heterogeneous factor loadings λ_i : $\omega_{jt} = \lambda_i' f_t$. Crucially, the same unobservable (latent) factors driving productivity are also assumed to drive inputs, i.e., $k_{jt} = f(\mathbf{f}_t)$ and $l_{jt} = f(\mathbf{f}_t)$. Given this commonality the dynamic panel version of the framework implies parameter constraints between regressors and the error structure in the system of T simultaneous equations, which are explored adopting a Full Information Maximum Likelihood (FIML) approach. Further assumptions about the initial conditions of the reduced form equation are necessary and the maximisation process is carried out using the Expectation Maximisation algorithm and its more recent variants. The multifactor error framework is clearly relevant for the study of production function estimation with unobserved productivity shocks and variable endogeneity; however, the complexity of the estimator and its data requirement make implementation beyond the reach of the average practitioner and the scope of this paper.

3 Empirics

This section provides an application of the estimators discussed above with the objective of comparing their relative performance and linking this to the findings in the wider empirical literature on production functions. For this purpose, we use a sample of UK high-tech firms for the period 2000-2007, taken from a database which is more readily accessible to researchers than firm-level data collected by national statistical offices and commonly made available only in secure data environments. Furthermore, the data is representative of similar firm-level data sets available for a large range of countries from the same source.⁴⁹

The UK FAME data set and similar data provided by Bureau van Dijk, most notably AMADEUS, have been used extensively in empirical work involving productivity analysis (e.g. Benfratello and Sembenelli, 2002; Faggio, Salvanes, and van Reenen, 2007; Li and Harris, 2008; Javorcik and Li, 2008). It is therefore of interest to investigate to what degree *plausible* production function estimates can be obtained from this type of firm-level data employing the estimation procedures presented above. It is important to stress that the data employed is in our eyes *typical* for firm-level datasets used for productivity analysis, but that they are characterised by serious limitations, which makes them *far from ideal* for the straightforward application of the estimators discussed above. We feel the latter point is too often neglected in applied work and therefore go to great lengths to discuss data source, construction and properties in Sections 3.1, 3.2 and 3.3 below.

3.1 Data

The data used for the analysis come from the ‘Financial Analysis Made Easy’ (FAME) database.⁵⁰ They cover the entire population of registered UK firms (FAME downloads data from Compa-

⁴⁹See website of Bureau van Dijk: <http://www.bvdep.com/en/companyInformationHome.html>.

⁵⁰We use the March 2009 edition of FAME.

nies House records).⁵¹ The FAME database is a commercial database provided by Bureau van Dijk.⁵² The version used here covers around 2.78 million active firms. For all of these firms, information such as name, registered address, firm type, and industry code are available. Availability of financial information varies substantially across firms. The smallest firms are legally required to submit only basic balance sheet information such as shareholders' funds and total assets. The largest firms provide a range of profit and loss information as well as detailed balance sheet data. As a result, given the data required to compute value added, the sample used for our productivity analysis covers mostly larger firms.⁵³ The March 2009 version of the FAME database also lists around 1 million so-called 'inactive' firms. These inactive firms were found to have exited the market and belong to one of the following categories: dissolved, liquidated, entered receivership or declared non-trading. The fact that FAME tracks *inactive* firms allows us to identify all firms exiting the market throughout the eight-year period of study, an issue relevant to gauge the importance of selection bias in the productivity estimates.⁵⁴

We limit the analysis to high-tech firms in the SIC (2003) 32 industry as shown in Table 1.⁵⁵ We focus on this narrowly defined industrial sector to allow for easier comparison across the various regression models, since the structural estimators do not account for industry fixed effects. Table 1 shows the number of registered firms in the UK in the industry as well as the number of firms in the sample used. The fact that the sample contains only slightly more than 10% of the firms listed in FAME highlights the limited availability of the data needed to estimate Cobb-Douglas production functions. This also suggests that the sample available in FAME for the estimation of production function is most likely *not* representative of the underlying population of firms because item non-response is a function of firm size. This is an important caveat for work using this data source, in that the 'external validity' of the empirical results obtained is at best tenuous and in many cases not given. The details involved in the construction of the sample are discussed in the following.

3.2 Variable Construction

All monetary variables obtained from FAME are expressed in thousands of British Pound Sterling and are deflated to 1999 price levels using SIC 2-digit level deflators from UK National Statistics.⁵⁶

⁵¹In the remainder of this work we use firms to mean registered firms. Hence firm refers to the legal entity that organizes production, in contrast to census-type data that uses the plant or production unit.

⁵²<http://www.bvdep.com/en/FAME.html>

⁵³Harris and Li (2008) weight the FAME data using weights obtained from the Annual Respondents Database (ARD) collected by the Office for National Statistics (ONS), claiming that the unweighted FAME database is not representative of the population of firms in the UK due to under-representing small and medium-sized firms. Yet, the bias does not arise from the lack of SMEs in the database but from the fact that no financial information is available on these firms as they are legally not obliged to report the data to Companies House. Weighting assumes that the fact that small firms do not report financial information is random, i.e., that the data is missing at random. This is unlikely to be the case, hence, weighting is unlikely to solve the missing data problem. In the present case the median (mean) firm size is 107 (353) employees in our Sample 1 (see discussion below) and 123 (443) employees in our Sample 2.

⁵⁴It avoids the common problem in stratified samples that missing observations are erroneously interpreted as having exited.

⁵⁵FAME provides primary as well as secondary SIC codes for nearly all firms in the database. To select firms belonging to the SIC 32 sector, we only use the primary SIC code.

⁵⁶<http://www.statistics.gov.uk/statbase/TSDSeries1.asp>

Output — Value added

We use value added as a measure of firm output, which is computed as total sales (turnover) less material costs.⁵⁷ The decision to use value-added as a measure for firm output is motivated mainly by the findings of BS who suggest that identification of perfectly variable inputs is not possible without input price variation across firms in a gross-revenue specification of the Cobb-Douglas production function.

As pointed out by Greenstreet (2007), single-deflated value-added, i.e., deflated nominal value-added, is inappropriate as a measure for output since changes in relative prices of inputs and output would be erroneously interpreted as changes in productivity. Instead, real value-added is computed from subtracting real input from real output.⁵⁸ Since FAME does not report firm-level output prices, output is deflated to 1999 price levels using SIC 3-digit industry-level producer price indices obtained from UK National Statistics.⁵⁹ Intermediate inputs are measured as cost of sales which is directly available in FAME. The measure for intermediate input is deflated using two-digit industry level input deflators from UK National Statistics.⁶⁰

It is important to stress that using deflated sales as a measure of output makes the underlying assumption that firms face a perfectly competitive market environment. If firms are subject to imperfect competition or offer differentiated products then even within a narrowly defined industrial sector of analysis, output prices are significantly dispersed across firms as well as correlated with a firm's inputs. Klette and Griliches (1996) provide theoretical and empirical evidence suggesting that using deflated sales as a measure of (unobserved) real output creates a downward bias in the production function parameters. The authors illustrate how a firm's omitted output price can be proxied by a firm's output growth relative to industry growth and therefore suggest to include growth in industry output in the firm-level production function to correct for the omitted output price variable.

On the other hand, Mairesse and Jaumandreu (2005) provide evidence arguing that the availability of firm-specific output prices and the econometrician's ability to use them to compute real firm output has little effect on the estimated coefficients. They point out that if changes in real output are only weakly related to changes in output prices over time *relative to* the link between demand shocks and the other possible determinants of demand, the correlation between output volumes and prices is weak and hence also the correlation between prices and production inputs, thus mitigating the downward bias from using industry deflators. At the same time, Mairesse and Jaumandreu point to another potential source of bias arising from unobserved capacity utilization of firms. However, we do not have any direct way of accounting for this given a lack of the necessary data.

Labour input

As labour input, FAME provides the number of full-time equivalent employees, recorded annually. Ideally, a measure of hours worked should be used to measure labour input more accurately. Since no such measure is available the number of full-time equivalent employees will have to serve

⁵⁷We alternatively also adjusted sales for changes in inventories, but this further reduces the number of firms for which there is data available.

⁵⁸As noted above, Bruno (1984) pointed out that if productivity growth is the object of interest, the use of double-deflated value added produces a negative bias in productivity growth (see also Basu and Fernald, 1995, 1997). However, Baily (1986) suggests that the bias may be negligible in practice.

⁵⁹<http://www.statistics.gov.uk/statbase/TSDSeries1.asp>

⁶⁰ibid.

as an imperfect proxy.

Capital and Investment

Ideally, we would like to have a measure of current capital services instead of capital stocks, i.e., a flow measure instead of a stock measure (Jorgenson and Griliches, 1967).⁶¹ But no such measure for capital services is readily available in FAME.⁶² The implicit price of such service is the ‘user cost of capital’ (or rental prices) which are usually unobserved. Therefore, we use capital stocks as a proxy. This appears to be an acceptable choice under the assumption that the quantity of an asset held by a firm is proportional to the quantity of the corresponding service obtained from that asset. For this to be the case, the aggregate of a firm’s capital holdings should represent an average over various different vintages and age groups of the capital employed. That this assumption may approximately hold in practice is supported by empirical work by Shashua et al. (1974), who find that the bias resulting from using capital stocks instead of flows is relatively minor for multiproduct firms. Hence, we assume that the flow of capital services in time t is proportional to the capital stock in t .⁶³

Capital input is measured as total tangible assets by book value, recorded annually. Tangible assets include land and buildings, fixtures and fittings, plants and vehicles, and other tangible assets.⁶⁴

⁶¹The flow of productive services coming from the cumulative stock of past investments is referred to as ‘capital services’ of an asset. Hence, the appropriate measure of capital input would be for example machine hours instead of the number of machines.

⁶²The fundamental difficulty in measuring capital services lies in the fact that a firm usually owns the capital from which it obtains the service (Jorgenson and Griliches, 1967). Therefore, no market transactions occur when the capital good is used to ‘deliver a service’.

⁶³There is evidence at the aggregate level for the UK for this to be a realistic assumption (Wallis and Turvey, 2009).

⁶⁴Ideally, we would compute the implicit rental rates for each type of tangible capital which are then used to aggregate the different types of assets into a single measure of capital services. The problem in FAME is that coverage for the overall figure of tangible assets is already relatively low as only 74% of registered high-tech firms in SIC 32 report tangible asset data. Given that a firm reports tangible assets, the share of firms reporting a breakdown of fixed assets into the different asset types is prohibitively low. Nevertheless, for demonstration purposes, we also computed tangible capital in the following way: we aggregate the different types of assets into a single measure for capital services by computing implicit capital rental rates in real terms as suggested by Jorgenson and Griliches (1967)

$$R_{jt} = \left(\frac{r_{jt} - \pi_{jt}}{1 + \pi_{jt}} + \delta \right) \left(\frac{P_{jt}}{P_{ct}} \right) \quad (62)$$

As pointed out by Greenstreet (2007), variation in the measure of capital services must be due solely to variation in levels of capital stocks. Hence, in order to avoid variation in the aggregate capital services measure to also come from variation in R_{jt} , simple averages of R_{jt} for the entire period observed, 2000-2007, are computed. As a measure for r_{jt} , as suggested by Greenstreet (2007), we use bank lending rates as reported by the IMF International Financial Statistics (IFS) Series 60P.ZF. The end of year capital services measure is computed as

$$K_t = \sum_{k=1}^4 R_j A_{jkt} \quad (63)$$

The problem is that FAME does not report the entire investment history of firms. Therefore, in order to compute the stock of capital for the first period for which data on a firm is observed, we follow Greenstreet (2007) and compute the stock of assets for the initial time period t as

$$A_t = (1 - \delta) \left[\left(\frac{D_t}{\delta} \right) - \left(\frac{I_t}{2} \right) \right] / P_{bt} + \left[I_t - \left(\frac{\delta}{2} \right) I_t \right] / P_t \quad (64)$$

where A_t denotes the value of a fixed asset at the end of period t . I_t denotes investment during period t , δ is the depreciation coefficient and D_t denotes the share in assets that has depreciated between $t - 1$ and t . P_t denotes the average price deflator during period t and P_{bt} is the price deflator at the beginning of period t . Depreciation rates are assumed to be six, two, and 20 percent for plant and machinery, buildings, and vehicles respectively as suggested by the UK Office for National Statistics. As deflators, sector-level PPIs are used. In order to construct P_{bt} , geometric averages of deflators in t and $t - 1$ are computed.

Tangible assets are deflated by a producer price deflator at the 2-digit sector level. In order to compute investment, rather unconventionally, we use the perpetual inventory method, i.e. $I_t = K_{t+1} - (1 + \delta)K_t$, where K is fixed assets deflated using a producer price deflator at the sector level. In order to determine the rate of replacement of old capital goods, we have to assume a value for the depreciation parameter, $\delta = 0.093$, using an average of the figures suggested by the UK Office for National Statistics for the components of fixed assets. This rather unconventional method of computing investment is preferable in the case of FAME as very few firms report investment (16%) while information on fixed assets is available for 51% of firms in the SIC 32 industry. Nevertheless, the lack of investment data reported by firms is a serious shortcoming of the FAME data.

Data gaps and cleaning

We restrict the sample to firms for which there are at least two consecutive years of data available. Moreover, we compute capital-labour and capital-output ratios and drop firms in the top and bottom 5 percentiles of the respective distributions to avoid the impact of outliers on our results. We discuss sample attributes in greater detail in the following section.

3.3 Descriptives and Pre-Estimation Testing

We use two different samples for estimation. In the first sample, we include all firms that report sufficient data to construct value-added, capital and labour inputs, which yields $n = 318$ firms with $n = 1,742$ observations ('Sample 1'). The second sample contains only firms for which we *also* have investment data, which is necessary for the estimation of the OP method. This 'Sample 2' is made up of $N = 214$ firms with $n = 958$ observations, thus over 30% smaller. The main differences between the full and reduced samples are higher average input and output levels for the restricted data set. Table 2 provides summary statistics for both samples. Looking at the minimum values of value added and inputs in both samples suggests that higher means in the reduced sample result from dropping firms at the lower end of the respective distributions from the sample. Figure 1 shows that the distribution for the reduced sample (in dashed blue lines) lies to the right of that for the full sample in all variable series considered. This confirms the hypothesis that data coverage is correlated with firm size, thus we would expect generally different empirical results between the two samples.

In Table 3 we furthermore provide information on the in-sample data coverage: for each of the samples we provide information on the time-series length of the data series employed in the regressions T_i . For Sample 1 we can compute an average of $T = 5.5$ observations per firm, for Sample 2 the equivalent number is $T = 4.5$. Note how data coverage is much more varied in the latter, where only 18% of firms have 7 or 8 time-series observations (30% of observations); for Sample 1 45% of firms fulfil the same criterion (60% of observations). The standard Fixed Effects estimator relies on within-firm variation over time — we would thus expect systematic differences in its results between the two samples studied. Furthermore, 60 out of 318 firms (19%) in Sample 1 have at least one gap in their time-series (of the 60 firms 3 have two gaps). For Sample 2, this rises to 111 out of 214 firms (52%; of the 111 firms 20 have two gaps). Curiously, the graphs in Figure 2 suggest that in the case of Sample 2 it is the firms with missing observations that are systematically larger. Both of these findings have bearings on the empirical results, in that we would expect to see differences between the static and dynamic

regression results, given the latter’s requirement of a continuous data series without gaps.

The scatter plots in Figure 3 display the relation between the output measure, value added, and inputs in the two samples. The bivariate plots show a monotonic positive relationship between input and output levels for all inputs, as should be expected. Moreover, there do not appear to exist any influential observations.⁶⁵

In order to highlight the time-series properties of the data we follow the suggestion of Blundell and Bond (2000) and estimate simple AR(1) regressions for output and inputs. In all regressions we add a set of $T - 2$ year dummies to take out the effect of common shocks. Our focus here is on the estimated level of persistence, say $\hat{\rho}$, in the data, as well as the pattern of these estimates across a number of empirical estimators, namely OLS, Fixed Effects, AB Difference GMM and BB System GMM. Recall that high persistence in the data raises the weak instrument problem in the AB estimator, whereby lagged levels of variables are not sufficiently informative instruments for differences at time t . Table 4 reports the results for Sample 1 and Table 5 for the restricted Sample 2. Across all variables and samples the OLS AR(1) estimates show signs of very high persistence, with all estimates of $\hat{\rho}$ above .96. As expected these estimates collapse in the Fixed Effects case due to Nickell bias. Weak instruments are in evidence in the Difference GMM estimates, which are much closer to (and even below) the FE estimates regardless of variable, sample or selected instrument set. The System GMM estimates confirm the high levels of persistence in the data, even suggesting estimates for $\hat{\rho}$ in excess of unity. In a longer panel the latter finding would be indicative evidence for an explosive process in all variable series, but given the nature of this panel (short T , gaps in the time series) we interpret this finding merely as strong evidence for high levels of persistence in the data.

Finally, we also briefly return to the issue of firm exit touched upon in Section 2.1. In the full sample of firms used, only 6 out of 318 firms (around 2%) included in the analysis exited the market during the entire period 2000-2007. While this extremely low share underlines the concerns one should have about using FAME data to learn about UK firm-level productivity,⁶⁶ it mitigates a possible bias from sample selection through firm exit as suggested in OP. Also comparing the mean of the log of capital for survivors, which is 7.15, to that for firms that exited, which is 8.49, suggests that firms that exited do *not* have systematically lower capital levels. Similarly, the median of the log of capital for survivors is 7.11 and for firms that failed 8.08.

3.4 Empirical Implementation

We begin our analysis by estimating input coefficients using simple OLS, pooling across all observations and running separate regressions with and without year dummies. We then add firm fixed effects (FE) and as an alternative to account for the impact of time-invariant heterogeneity estimate the production function in first differences (FD). The FE estimator requires strict exogeneity in the factor inputs, i.e. $\mathbb{E}(\epsilon_{jt}|x_0, x_{j1}, \dots, x_{jt}, \dots, x_{jT}) = 0$ where $x \in k, l$, since the within-transformation introduces variable dependence on the time-series mean in each observation. In contrast the FD estimator only requires weak exogeneity, i.e. $E(\epsilon_{jt}|x_0, x_{j1}, \dots, x_{jt}) = 0$.

⁶⁵We ran robustness checks for the static and dynamic regressions in Sample 1 omitting the 5 largest firms which appear as possible outliers in the first scatter plot for output-labour. Results are virtually identical to those we report below.

⁶⁶Actual exit rates are considerably higher. Looking at the entire cohort of UK firms incorporated in 2001, Helmers and Rogers (2010) find that approximately 30% of firms exited by 2005.

The latter allows *future* input realisations to be correlated with the contemporaneous error term and therefore imposes less restrictive assumptions on the structure of the correlation between inputs and the error term. On the other hand, it is well-known that first-differencing exacerbates measurement bias in the regression. For all of the above estimates standard errors are clustered at the firm-level.

The first stage of the OP and LP procedures is implemented through a third-degree polynomial expansion. In the OP estimation, we also include interaction terms with year dummies in the first stage specification. The second stage in both OP and LP models is estimated using nonlinear least squares where $\hat{\phi}$ obtained in stage one enters as a quadratic and cubic term. We rely on bootstrapping for inference given that this is a two-step procedure. Note that the bootstrap involves *both* stages of the estimation procedure.⁶⁷

The ACF procedure is implemented as follows. The first stage is similar to the OP and LP procedures described above. Like in the LP we use intermediate inputs in the first stage as a proxy for unobserved productivity. The second stage is implemented through numerical methods, namely a Newton-Raphson optimisation procedure. Initial values for input coefficients are taken from an OLS regression and used to compute an initial value for ω_{jt} using $\hat{\Psi}_{jt}$ obtained in the first stage. Then ω_{jt} is regressed nonparametrically on ω_{it-1} to obtain ξ_{jt} .⁶⁸ The residuals ξ_{jt} are then used to minimize the ‘criterion function’ in equation (59) above. This optimisation procedure is applied iteratively until the algorithm has determined the minimum and the corresponding coefficient estimates $\hat{\beta}_l, \hat{\beta}_k$. Following ACF, we use k_{jt} and l_{jt-1} as second-stage instruments. Inference is based on bootstrapping over the three stages of the estimation procedure.

We implement both the AB Difference GMM and BB System GMM estimators. For each of these dynamic models we begin with an empirical specification which follows the assumptions of the OP, LP and ACF estimators: labour and the lagged dependent variable are treated as endogenous, while capital stock is treated as predetermined. We then test all of these assumptions using the suitable overidentification tests (Hansen, Sargan, Difference Sargan — depending on whether the two-step or one-step method is implemented) and adopt the lag-length which is consistent with these test results. Note that measurement error in the factor inputs and the lagged dependent variable can cause the breakdown of the exogeneity assumptions as just laid out and necessitate the adoption of alternative lag structures for instrumentation. Furthermore, in the System GMM case we need to worry about exogeneity with respect to the firm fixed effect η_j , which is also a testable assumption. Initial condition assumptions about whether the economy is in long-run equilibrium cannot be tested formally and will need to be assessed intuitively. Since our time-series dimension is relatively short and we do not see any signs of overfitting bias (Sargan p -values close to unity) we always employ all possible moment restrictions implied by the adopted lag-structure. In all GMM models we specify year dummies to account for common shocks γ_t . We estimate both the unrestricted and the restricted models shown in Equation (11) and Equation (12) above. The restricted coefficients are obtained by first consistently estimating the unrestricted coefficients in Equation (11) and then using these coefficients

⁶⁷For the LP procedure, we use the `levpet` command in `Stata` (Levinsohn, Petrin and Poi, 2004).

⁶⁸We use `STATA`’s `locpoly` to estimate a local polynomial regression using a 4-th order polynomial.

in a Chamberlain-type minimum distance estimation (Chamberlain, 1984; Wooldridge 2002). Instead of the one-step GMM estimator used by BB, we use the asymptotically more efficient two-step procedure which uses the optimal weight matrix, correcting for possible small sample downward bias as suggested by Windmeijer (2005).⁶⁹

3.5 Results

In discussing our results we are particularly interested in assessing whether factor parameter estimates differ across the different estimators if estimators rely on different timing assumptions regarding firms' input decisions and/or the assumptions made with regard to the unobserved productivity term, i.e., scalar unobservability and strict monotonicity.

The results are shown in tables 6 to 9 for the static and dynamic results, in each case we present the regressions for Sample 1 and 2 respectively. Due to the data limitations the larger Sample 1 does not allow for estimation of the OP. We also do not estimate either AB or BB in their static form — the latter practice is uncommon although econometrically viable (see Baptist, 2008).

3.5.1 Static Specification

The results from the pooled OLS regression for the full Sample 1 in Table 6 yield a labour input coefficient of .69 and a capital coefficient of .21 if no time dummies are included and of .71 and .19 with time dummies. These technology parameters are remarkably close to the 'widely accepted' benchmark data on labour share in output .7 and (under CRS assumption) the implied capital share .3. The fixed effects estimates in Columns (3) and (4) of Table 6 drop as expected vis-à-vis their OLS counterparts for both labour and capital coefficients. This finding would agree with the idea of an *upward* bias in OLS estimates due to the transmission problem. At the same time, however, it needs to be noted that capital stock may be more likely measured with error when firms are large, given the complexity of accounting for all book values and their different vintages. Further assuming a systematic relationship between large firms and their unobserved productivity this measurement error may come to bear more significantly in the within transformation where a large share of variability is taken out of the data. An alternative estimation using first-differences instead of FE in Columns (5) and (6) shows an even more pronounced drop in the levels of both input coefficients. From our preliminary data analysis we know that all variables are highly persistent, such that near-unit root behaviour renders variables in first differences akin to random walks. Our findings thus far are a perfect reflection of the common experience in the applied literature, whereby attempts to account for firm-level fixed effects while not considering any further impact of heterogeneous productivity shocks lead to dramatically lower capital coefficients and large decreasing returns to scale.

The estimates for the LP approach in Column (7) are somewhat disappointing: the labour coefficient drops to .20, which is unrealistically low. In contrast, the capital coefficient increases to .33 which may be reasonable if one assumes that the OLS estimate is biased downward. Recall that capital is deemed predetermined in the LP approach, while labour is assumed to be

⁶⁹We use Roodman's (2006) `xtabond2` command to implement the AB and BB estimators. Måns Söderbom provided code to implement the minimum distance procedure. Söderbom's code is available on his personal website: <http://www.soderbom.net/Resources.htm>

perfectly variable and determined by the firm *after* the productivity shock has been observed. While the identification strategy for the capital coefficient may thus be valid, the collapse of the labour coefficient points to the ACF criticism that labour is unidentified in the first stage due to perfect collinearity with the productivity term proxied by the polynomial function in capital stock and intermediate inputs. In the final column, we report the estimates for the ACF approach. The labour coefficient is .61, which is lower than any of the OLS estimates in Columns (1) and (2) and higher than the FE estimate in Column (3), but lower than the FE estimate in Column (4). The ACF capital coefficient is .32, which is higher than for the OLS, FE or FD specifications. This suggests that capital is less weakly correlated with the productivity shock than labour, which results in a downward bias of the capital coefficient when using OLS. These results for the ACF procedure appear to be in line with these authors' findings using Chilean firm-level data, which also found generally lower coefficients for labour input relative to OLS, but a capital coefficient that is either larger or smaller than the OLS estimate (depending on the industrial sector). Yet, in our sample the bootstrap standard errors obtained for the ACF are much larger than for the other estimators. The main reason for this may be that the ACF estimation approach is more demanding in its data requirements. Given the many holes in our panel, bootstrap samples may differ substantially in terms of the number of available observations to estimate the ACF procedure; as a result, estimates for the input coefficients differ considerably across bootstrap samples. This illustrates an important drawback of employing the ACF procedure when only 'imperfect' firm-level data is at hand — which is arguably the standard case.

All models in this larger sample reject constant returns to scale (in favour of decreasing returns) at standard significance levels. Note that our sample is defined very narrowly in terms of sector of operation, and it may be that the various attempts to control for endogeneity in the standard OLS regression in column (1) in effect throw out the baby with the bath water: introduction of fixed effects or their accommodation via first differencing may introduce data-dependencies which lead to more severe bias than the endogeneity of factor inputs in the standard OLS model. The stringent assumptions of the LP estimator also seem certain to be violated given the large decreasing returns to scale ($\hat{\beta}_l + \hat{\beta}_k \approx .55$). The fact that the ACF model brings the labour coefficient back up to the level where we expect it to be suggests that labour is indeed unidentified in the first stage LP regression, which nevertheless does not seem to prevent a consensus to emerge for the capital coefficient across LP and ACF.

Table 7 reports the results for Sample 2, which required investment data and was shown to be made up of larger firms and characterised by a high number of gaps in the individual firm's time series. Comparing results across samples, the estimates for OLS in this restricted sample are lower in case of the labour coefficient and higher for capital compared to those in the larger sample in Table 6. The labour input coefficient falls from .69 to .64 with no year dummies and from .71 to .65 including year dummies. For the capital coefficient, the estimates increase from .21 and .19 to .31 and .30 respectively. If these changes were deemed to arise from 'structural' differences between the two samples, rather than any differential manifestation of transmission bias, then we may argue that the more capital-intensive result in the restricted sample is in line with our evidence of firm characteristics (refer again to Figure 2, in particular the bottom left density plot for capital stock). Overall these estimated factor coefficients are surprisingly

close to the ‘known’ labour and capital shares. As should be expected, the estimates for the FE specification in Column (3) falls relative to the OLS estimates in Column (1), although not as dramatically as in the Sample 1 case. The conceptually preferable FE estimate for the specification including year dummies (thus accounting for common shocks), counterintuitively, yields a *higher* labour coefficient than the corresponding OLS specification whereas the capital coefficient is substantially reduced. The same pattern is repeated in the FD estimates. We could argue that the seemingly irregular results for the models with year dummies relate to the large number of gaps in the time-series. For all of the estimators discussed so far in Sample 2 we find no evidence against constant returns to scale, which is in stark contrast to the largely decreasing returns in Sample 1.

The OP estimate for the labour input yields a coefficient lower than OLS as predicted by theory in the presence of transmission bias. The coefficient is larger than the FE estimate excluding year dummies. The capital coefficient obtained by using the OP procedure of approximately .07 is much lower than in any other specification, casting some doubt on the estimation procedure and the investment data on which it crucially depends. We have noted above the lack of investment data in the FAME data set, which motivated our construction of the investment measure; however, the capital coefficient obtained by using the OP procedure suggests that there may exist a non-negligible problem with the way we have constructed investment which supports a preference for an estimator that relies on intermediate inputs instead. Figure 4 shows the distribution of investment, capital stock and TFP derived from the OP estimator.⁷⁰ This is a visual test of the monotonicity assumption, where investment is assumed to be strictly increasing in both capital stock and productivity. The figure indicates some non-monotonicity of investment, in particular with regard to the capital stock variable, possibly reflecting the presence of adjustment costs to investment. The violation of the monotonicity assumption may partly explain the poor performance of the OP estimator with regard to the capital coefficient.

Similarly to the results in Table 6, the results for the LP procedure are hard to reconcile with the theoretical discussion in Section 2.1. To test whether the monotonicity assumption holds in the case of intermediate inputs, Figure 5 shows the smoothed plot of intermediate inputs, productivity, and physical capital. This suggests that the monotonicity assumption for intermediate inputs appears to hold as higher productivity realisations appear to be associated with higher intermediate input levels for given physical capital levels.⁷¹ The monotonic increasing relationship is certainly more pronounced in the case of intermediate inputs than for investment. This finding is also reassuring with regard to the ACF estimator as we use intermediate inputs to back out unobserved productivity. The estimated LP labour coefficient is far too low while the capital coefficient increases dramatically relative to the OLS, FE, FD and OP estimates. In their regressions of Chilean firm data ACF compare the estimates from their own empirical approach against estimates obtained using the LP method in various industrial sectors: while the differences do not appear to follow any discernable pattern, labour coefficients from LP are found to be lower in a number of sectors. ACF interpret this finding as an overall downward bias in the LP estimates. Our estimates also suggest a strong downward bias in the labour

⁷⁰We normalized TFP (ω) in order to simplify the comparison with the corresponding plot for the LP estimator shown in Figure 5. We use the *lattice* package in *R* to draw these plots.

⁷¹Note that we also normalized TFP (ω) for this plot.

input estimates. This aside, the LP capital coefficient appears to be too large, indicating a possible upward bias. An interesting point raised by ACF is the greater sensitivity of the LP procedure (relative to their own estimator) to the type of intermediate input used in the first stage regression. Unfortunately, FAME provides only a single measure for intermediate inputs and we therefore cannot verify this claim in the present application.

Finally, the ACF estimates are similar to the results for the full sample. The estimate for labour input of .61 and for the capital coefficient of .31 are reasonable and both move into the directions consistent with theory. In their estimates using the Chilean data, ACF find consistently lower returns to scale than when using OLS. They argue that this is indeed the expected direction of change when using their estimator as OLS estimates are expected to be upwardly biased. In our ACF estimates, the reduction in the estimate of returns to scale comes mainly from a lower labour input coefficient. While the ACF method yields lower labour input coefficients vis-à-vis OLS in both samples, the capital coefficient in Table 6 increases considerably relative to the OLS result but remains virtually the same in Table 7. Due to the strong increase of the capital coefficient in Table 6, returns to scale estimated using ACF are larger than those implied by OLS. ACF found in their own application that the capital coefficient may be larger or smaller than the OLS counterpart, depending on the specification and sector for which the production function was estimated. ACF argue that this pattern emerges from the fact that l_{jt} is more correlated with ω_{jt} than k_{jt} . Overall, our own estimates using ACF appear to be in line with the estimates in these authors' application.

Two issues are worth noting in conclusion to these static production function and control function regressions: firstly, the undisputable collapse of the OP and LP estimates in the face of fairly standard, developed-country firm-level data. For the OP case, where we obtain sensible labour coefficients but rather small capital coefficients, this is likely to be closely related to the lack of observed investment series in our data. Having said that, the breakdown of the monotonicity assumption in the presence of zero investment is a sizeable hurdle for the application of this empirical strategy. The LP case, in contrast, while somewhat affected by the inadequacy of our investment series, is primarily characterised by grossly underestimated labour coefficients. As suggested by ACF, the estimator seems unable to identify labour in the first stage regression under the present and many similar circumstances. Secondly, we need to highlight the relative performance of the standard OLS estimator, which in both samples can be seen to be largely replicated by the vastly more complex ACF estimator: the latter's point estimates are comfortably contained within the OLS 95% confidence intervals.

3.5.2 Dynamic Specification

Tables 8 and 9 report the corresponding results for an ARDL(1,1) specification implementing the dynamic panel estimators discussed above. Using the full sample in Table 8, the OLS estimates in column (1) marginally reject the common factor restriction at the 10% level, thus without imposing common factor restrictions the long-run coefficients obtained are $\hat{\beta}_l = .790 (.182)$ and $\hat{\beta}_k = .161 (.131)$ respectively (reported in the table footnote). Hence, the estimate for the labour coefficient is larger, while the capital coefficient is lower than in the static model in Table 6. Due to the imprecision of the capital coefficient these long-run parameters cannot reject CRS

($p = .638$). The Arellano and Bond (1991) test indicates serial correlation in the residuals. The coefficient for the lagged dependent variable is relatively close to unity and thus may be upward-biased, which would suggest the necessity for firm-specific fixed effects.

As expected given the relatively short time-series compared to the number of firms (Nickell, 1981), adding firm fixed effects to the dynamic specification (their statistical significance confirmed by an F -test, $p = .000$) significantly reduces the coefficient on the lagged dependent variable to around .3. Since this and the following models in Sample 1 cannot reject common factor restrictions we focus on these results henceforth. While the labour coefficients are virtually unchanged, the capital coefficient is again much lower for the dynamic specification than for the static model and is now insignificantly different from zero. Our serial correlation tests furthermore indicate that the residuals are subject to AR(2) correlation over time. We know that the downward shift in the estimates is due to the ‘Nickell-bias’ induced by contemporaneous correlation between regressors and residuals as a result of the within-transformation.

Applying the AB Difference GMM estimator we opt for two alternative specifications: the first, identified as AB_0 , adopts the same timing assumptions as the structural estimators by OP, LP and ACF, namely that capital stock is predetermined whereas labour is endogenous (as is, by construction, the lagged dependent variable). The second model, identified as AB_* , represents the preferred specification based on diagnostic tests — note that ‘preferred’ here does not imply that the specification satisfies all diagnostic criteria (more on this below). Similarly for the System GMM models. Since all of these do not seem to suffer from serial correlation (given that AR(1) is introduced by construction) we need not discuss the AB serial correlation test results any further. Note that when applying the Difference GMM estimator in this sample we lose almost exactly a quarter of the observations due to the gaps in the firm-level data series. The AB_0 model in column (3) provides a coefficient on the lagged dependent variable of around .5, which consistent with theory is thus within the bounds set by the OLS (upward-bias) and the FE (downward-bias) estimates. The long-run labour coefficient at .65 is actually below that of the FE model, while the capital coefficient has recovered somewhat from the lows of the Nickell-bias induced FE, but is still statistically insignificant. A Sargan test of overall instrument/moment restriction validity for this model indicates that the null of exogeneity is rejected, thus rendering our instrumentation strategy invalid. The empirical implementation in *Stata* (`xtabond2`) allows us to test the validity of each moment restriction separately, from which analysis we can conclude that in the present case the instrumentation of the lagged dependent variable is invalid.⁷² In column (4) we therefore adjust the instrument matrix to obtain AB_* , which yields sensible capital and labour coefficients very much in line with our preferred static model and no evidence for invalid instruments (Sargan $p = .82$, various Difference-in-Sargan tests for the levels equation instruments). This finding is rather surprising, given the high degree of persistence in the variable series, in which case we would expect the past levels employed by the AB estimator to represent weak instruments for the endogenous variables in first-differences (i.e. they are only weakly correlated and thus not terribly ‘informative’). Recall the AR(1) regression results in Table 4: all Difference GMM estimates are lower than the System GMM estimates and closer to the Nickell-biased FE estimates, which can be interpreted as evidence for weak instruments.

⁷²Employing a Difference-in-Sargan test for the lag $t - 3$ as instrument for Δy_{it-1} we find the null of exogeneity rejected at $p = .001$.

Moving on to the System GMM estimates in Columns (5) and (6) of Table 8 (focus is on the restricted model results), we can see that the BB_0 model yields a substantially higher lagged dependent variable coefficient but a substantially lower capital coefficient, which is statistically insignificant. Crucially, the Sargan test implies that the overall instrumentation strategy is invalid (exogeneity rejected, $p = .018$), while the Difference-in-Sargan test suggests that the instruments/moment conditions for the additional levels equation may be to blame (exogeneity rejected, $p = .018$). Further investigation confirms that the instruments for the lagged dependent variable in the difference equation violate the exogeneity assumption ($p = .002$) — this is in line with our finding in AB_0 . Next, in the BB_* model we attempt to remedy this problem by restricting the lags for the lagged dependent variable instrumentation to those from $t - 4$ and onward for Δy_{it-1} . The long-run labour coefficient is again relatively similar to the AB and BB_0 estimates, while the long-run capital estimate changes only marginally in the direction of the former. While the Sargan test indicates overall instrument validity ($p = .159$) the Difference-in-Sargan tests again indicate that the additional moment restrictions afforded by the levels equation of the System GMM are invalid. Further analysis indicates that the moment restrictions for the labour and lagged value-added variables are invalid; attempts to remedy this problem by selectively dropping moment restrictions does not lead to a qualitative change in the coefficient estimates but increases imprecision, suggesting low informativeness of the remaining instrument sets. The use of lagged differences as instruments for levels thus did not prove successful in the present dataset while the correctly specified AB approach seemed to yield meaningful and seemingly reliable results. The latter finding is rather curious, given that we indicated the high persistence in the data and the previous findings using simulations (Blundell and Bond, 2000).

All of the GMM estimates for the lagged dependent variable lie within the bounds set by the OLS and FE estimates, however given a lack of obvious priors about the ‘correct’ level of persistence in the economic system, it is difficult to distinguish further between the various approaches based on the $\hat{\rho}$ -coefficient alone. Our diagnostic testing suggested that the Difference GMM estimates in column (4) provides our best bet for consistency and efficiency in Sample 1.

Turning to the results for Sample 2 in Table 9, we note that the estimates obtained for the restricted sample yield larger returns to scale for all specifications, excessively so in models (2) and (5). Beginning again with the OLS results we find the common factor restrictions accepted and can thus focus on the restricted results. In comparison with the larger sample, the OLS results here yield a markedly higher labour coefficient, around .9, while capital has dropped to .09 (statistically insignificant). Residuals are seemingly free from serial correlation.

Once fixed effects are included (F -test indicates significant differences across firm intercepts, $p = .000$) the already moderate capital coefficient plunges further, while the coefficient on the lagged dependent variable drops from .91 in OLS to .35 here. Serial correlation is found in the residuals and the model bears all the hallmarks of the Nickell bias.

When we move on to the Difference GMM estimator we drop over 40% of the observations (and equivalently 78 out of 214 firms), due to the large number of gaps in our Sample 2 data. In the basic identification setup AB_0 we reject the common factor restrictions, thus leading to long-

run coefficients of $\hat{\beta}_l = .588 (.413)$ and $\hat{\beta}_k = -.183 (.260)$. While the instrumentation strategy seems valid, the loss of precision in the empirical estimates as well as the negative coefficient on capital stock may be attributed to the weak instrument problem in this substantially reduced dataset. Attempts to remedy this problem by adopting alternative moment restrictions does not prove terribly successful: in column (4) we present our preferred specification AB_\star which nevertheless displays an excessively large labour coefficient and barely improves on the OLS capital estimate. These results suggest that relying entirely on the estimation equation in first differences may prove too challenging for the data if gaps in the data are too numerous. Recall further that the weak instrument problem may be driven by the high level of persistence we found in our AR(1) regressions reported in Table 5.

Using the System GMM estimator in the basic identification strategy, BB_0 in column (5) produces similarly extreme coefficients on labour and capital as in the AB estimator, arguably even more so: $\hat{\beta}_l = 2.071 (1.253)$ and $\hat{\beta}_k = -.024 (.722)$ (since the common factor restriction is rejected we report the long-run coefficients implied by the unrestricted ARDL model). In contrast the selective restriction of moment conditions can yield a more reasonable picture as is evidenced in column (6): in the BB_\star we restrict the instrumentation for capital stock to the difference equation and furthermore impose the identity matrix as the optimal weight matrix in the one-step estimation procedure (both in the difference and levels equations). In effect this turns the System GMM into a standard (inefficient) 2SLS estimator, following the original specification strategy in Blundell and Bond (1998). This approach yields sound overidentification diagnostics and statistically significant long-run factor coefficients of .77 for labour and .34 for capital. While the imposition of the identity matrix in this regression and the effective return to a more conventional IV estimator seems *ad hoc* and thus questionable, we may suggest in our defence that the challenges imposed by the data availability in the present case seem to be tackled successfully although the weak instrument problem still remains, at least for the labour and lagged dependent variables.

3.5.3 Summary of Empirical Results

In summary the empirical results from our two (sub)samples of UK high-tech firms yield a number of important insights into the application and relative performance of modern production function estimators.

- (i) Data availability, structure of the missing observations and time-series persistence are found to play an important role in the performance of all standard, IV and ‘structural’ production function estimators tested and account in part for the considerable differences in the estimates for the input coefficients.
- (ii) The LP strategy yielded rather implausible labour coefficients, most likely linked to the identification problem for labour in the first stage of their procedure as pointed out by ACF. In contrast the LP capital coefficients were very close to the preferred estimates in both samples. In case of the OP estimator, where the identification issue is said to be less pronounced in the first stage, the poor performance of the estimation to identify capital stock may in part be due to the specific problems introduced when we created the investment data series, which is also reflected in the suggested violation of the monotonicity assumption.

- (iii) ACF estimates were seemingly meaningful and surprisingly similar in the two samples, despite the considerable data problems in the smaller sample. It is, however, notable that in the case of Sample 2 these estimates did not differ substantially from those of a standard OLS regression, which to a lesser extent is also the case in Sample 1. Furthermore, data properties proved having a strong impact on the size of the bootstrapped standard errors.
- (iv) The dynamic panel data estimators by Arellano and Bond and Blundell and Bond have proven to be very flexible in their application, given that they allow us to test the validity of individual moment restrictions. The results for the AB estimator in Sample 1, which are virtually identical to those of the ACF in the static case, are however somewhat questionable given the finding of high persistence in the variable time series. As long as instrument informativeness has to be assumed, rather than tested and confirmed, the AB estimator will always be questioned as driven by the weak instrument problem (although in the preferred specification in Sample 1 the unrestricted and restricted estimates are very different from the OLS estimates, toward which they converge in this case). The BB estimator could not be made to work in Sample 1, highlighting the ability of the researcher to identify this short-coming, rather than to rely on the validity of all assumptions as in case of the structural estimators. In the more challenging Sample 2, a specification could be identified for BB to satisfy the instrument validity concerns while also yielding meaningful long-run parameter estimates.

3.6 Going beyond the sample regression results

Our empirical illustration thus far has emphasised the technology parameter estimates for capital and labour. In the absence of a substantial array of established residual diagnostic tests in panel econometrics (for long- T panels see Banerjee et al., 2010) we have adopted the ‘conventional wisdom’ of capital and labour income shares of .7 and .3, as well as the proximity to constant returns to scale as a means to gauge the viability of results from various regression models. In the present section we want to add to this analysis by conducting a number of robustness checks of the regression results and by further probing the implied TFP level estimates using correlation and ‘second-stage regression’.

3.6.1 Slope coefficient estimates

In the following we employ bootstrap samples to investigate the distribution of slope coefficient estimates across a pseudo-universe of firms consistent with the moments of the regression residuals. We do this in two distinct ways. First we construct 200 bootstrap samples based on the residuals of each estimator and estimate the slope coefficients from said estimator. This procedure arrives at estimates for OLS, 2FE, FD and LP which indicate the spread of estimates conditional on the validity and consistency of the estimation procedure. In Figure 6 we provide histograms for each of the 2FE, FD and LP estimates in deviation from the OLS results (for labour and capital coefficients respectively). Focusing first on the relative labour coefficients in the left column, we can see that OLS and 2FE results are quite similar, with 2FE on average slightly larger in magnitude as evidenced by the slightly larger left tail of the histogram. FD estimates for labour are on average somewhat smaller than OLS ones, although there is still some overlap between the two for 95% confidence intervals. LP labour estimates, on the

other hand, as was already evident from our results above, are substantially smaller than OLS based ones, on average we observe a location shift of $-.5$ for the LP labour coefficients. For the capital coefficients in the right column, 2FE and in particular FD represent coefficients considerably smaller than for OLS based regressions, whereas the LP estimates are somewhat larger.

As a second robustness check we follow Levinsohn and Petrin (2003) and estimate OLS, 2FE, FD and LP models with the bootstrapped samples based on the OLS residuals. So while our previous procedure constructed 200 bootstrap samples from each estimator’s residuals, the present one only creates samples based on the OLS estimator and then employs all 4 estimators to the same bootstrapped sample. In Figure 7 we plot the estimated distributions (using Epanechnikov kernel estimation) of the labour (left plot) and capital (right plot) slope estimates for the (i) OLS (blue solid line), (ii) 2FE (black dash-dots), (iii) FD (green dots) and (iv) LP (red dashes) estimators.⁷³ It can be seen that contrary to the example provided in Levinsohn and Petrin (2003), the estimates from various empirical models do not differ substantially in their central mass, which is around $.7$ and $.2$ for labour and capital respectively, although the spread of the estimates differs considerably across the different estimators. We obtain very similar distributions when we use LP residuals to construct the bootstrapped samples, with the exception of a location shift as the mode for the labour coefficients is now around $.2$.

3.6.2 TFP estimates

An important question is whether the different point estimates we obtained in the production function regressions for various empirical models actually have a bearing on the resulting TFP level estimates. Given the stark differences between, say, the OLS and ACF results on the one hand and 2FE and LP results on the other, it would be somewhat surprising if this were not the case. In the following we focus on the static specifications in our Sample 1, thus limiting the discussion to the OLS, FE-type, LP and ACF (in total: six) estimators in the production function regressions in Table 6. It is worth noting that we follow common practice and compute the TFP levels from the production function estimates whilst ignoring any concerns about inference in the latter: even though the computed or bootstrapped standard errors indicate considerable difference in the size of the 95% confidence intervals for the estimates across our empirical models,⁷⁴ we take the point estimates at (deterministic) face value.

We begin by computing Pearson pair-wise correlation coefficients for the TFP-level estimates⁷⁵ from these six estimators: the top-left panel of Table 10 indicates a uniformly high level of correlation in this case, with the ‘weakest’ correlation coefficient between FE and ACF TFP a sizeable $.63$. We calculate the same correlation matrix following a number of transformations: in the top-right panel of Table 10 we compute TFP-levels in deviation from the firm-specific means, in the bottom-left panel in deviation from the cross-section means and in the bottom-right panel

⁷³Note that the bandwidth applied in the kernel estimation is allowed to differ across these four data series: imposing the same bandwidth leads to essentially the same result of a common mass point but readability is severely affected.

⁷⁴As an illustration, for the labour coefficient we obtain the following 95% CI: $[.548, .838]$ for OLS, $[.530, .860]$ for 2FE, $[.400, .744]$ for FD with year FE, $[.051, .349]$ for LP and $[-.439, 1.663]$ for ACF.

⁷⁵Note that in all cases we estimated these TFP-levels ‘manually’ in *Stata*: $\hat{\omega}_{jt} = y_{jt} - \hat{\beta}_l l_{jt} - \hat{\beta}_k k_{jt}$ with observable variables in logs and $\hat{\beta}_l, \hat{\beta}_k$ the slope coefficient estimates from each regression model. To create ‘TFP levels’ we then need to take the exponential of the resulting residual.

a combination of the two transformations. Taking out this variation in many cases leads to a further *increase* in the correlation coefficients, which are now uniformly close to unity.

As a second step we follow van Beveren (2010) and investigate the commonality across the TFP estimates with a simple regression model: following the collapse of the dot-com bubble in 2000 the US NBER suggested a recession beginning in March 2001, which can be anticipated to have also affected the electronics industry in the UK. We regress TFP estimates on a set of year dummies to chart the evolution of TFP over time. The results in Table 11 indicate substantial levels differences in TFP by production function estimator applied, with FE and LP TFP levels considerably higher than those for OLS and ACF. Year dummy coefficients, however, do not differ substantially across models and follow a very distinctive pattern: after a substantial dip in 2001 of around 15%, productivity recovered to previous levels in the following year and then showed healthy growth rates of around 15% per annum until the end of our sample period. If we interpret the year 2001 dummy as the ‘crisis effect’ then with the exception of ACF all TFP estimates are negative and statistically significant at the 5% level (ACF: 10% level). One may draw the conclusion from this and the previous exercise that some location shifts aside, the differences across the TFP-level estimates must be fairly limited, given the uniformly high co-movement across all six models.

In Figure 8 we provide scatter plots for the various TFP estimates (in logs), with the standard least squares based estimates always on the y -axis and the 45° line (indicating equality of TFP level estimates across estimators) in red. As can be seen there are some visible deviations from the OLS-based estimates, in particular in case of the LP- and FE-based TFP estimates; however the kernels for the re-centred TFP estimates (in logs) in the bottom right of the figure (‘within’-transformation) are virtually indistinguishable.

The aim of our next exercise is very simple and mimics the standard practice in the applied literature: we investigate the ‘determinants’ of TFP by regressing the TFP-levels implied by our above production functions on some variable of interest. Our main focus here is the relationship/correlation between TFP and innovation, where we consider patent applications seeking patent protection in the UK filed either at the European Patent Office (EPO) or the UK Intellectual Property Office (UKIPO) as our measure for the latter.⁷⁶ The relationship between intellectual property in the form of patenting and firm performance is a long-standing issue in the literature, which has generated a considerable amount of empirical investigations (Griliches, 1987). The patent data is taken from EPO’s Patstat (April 2010 version) and was matched to FAME using applicant and firm names since the two datasets do not share a unique common identifier.⁷⁷ The data on patent applications is available from 1982 to 2008 and in the results presented here we construct a ‘patent stock’ from this flow variable, which is then employed for the years 2000 to 2007. The mean patent-stock for our sample of $n = 318$ firms is around 440, although the median value is zero: 50 firms have any patent stock, thus 268 firms have no evidence of applying for a patent in the one-and-a-half decades before as well as during our

⁷⁶This includes patent applications that have come through the Patent Cooperation Treaty (PCT) route designating the UK. The PCT route offers a way for an applicant to obtain patent protection in several countries worldwide through a single application filed with the World Intellectual Property Organization (WIPO) in the same way as if the applicant had filed separate applications in all countries.

⁷⁷For more information on the matched dataset see Helmers et al. (2010).

sample period. The share of patenting firms (16%) is quite sizeable, reflecting the high-tech nature of the firms in our sample and most likely also the fact that the sample of firms reporting sufficient data to estimate a production function contains mostly larger firms. The fifty firms with patent stock submitted 111 patent applications between 2000 and 2007 and no firm which submitted an application during this time did not already have some patent applications before.

In Table 12 we present least squares regression results for a number of ‘second stage’ models: we begin with the patent stock as the sole RHS variable (aside from an intercept) in Panel A, then add a set of common year dummies in Panel B, which we include for the remainder of models. In order to take out some additional variation without reducing the sample size we then add region dummies in Panel C. Since these are time-invariant they may further capture parts of other firm-level, time-invariant unobservables.⁷⁸ Conceptually, we feel it somewhat contradictory if fixed effects were added to this setup: the control function estimators are *invalid* in the presence of firm-level fixed effects and our production function estimation for LP and ACF was carried out *under the maintained assumption* of a common TFP level across all firms in this industrial sector. It is thus difficult to argue for such TFP-level differences in a second stage analysis of the ‘determinants of TFP’.

In our discussion of the regression results in Table 12 we focus on statistical significance (again a common practice in ‘second stage’ regressions) and *relative* magnitude of the regression coefficients *across* the different models. We are not interested in any economic interpretation of the estimated coefficients (although they point in the right direction) or in making causality statements about the findings — the objective here is to illustrate how we can obtain starkly different results when TFP was computed using different production function estimators. Due to the unit values of some of the covariates their coefficients can be very small but we do not attempt any calculation of marginal effects here.

The patent stock model in Panel A yields statistically insignificant results for the OLS and ACF estimators, yet the two FE and the LP estimator indicate a positive and significant relationship between TFP and patent stock, with *t*-statistics of around 4.5 for the FE and LP models. Results are virtually unchanged when we add year dummies to account for common shocks. Furthermore, the year dummy coefficients (not reported) still indicate a negative TFP shock in 2001 and subsequently rising TFP levels across all models, in line with our earlier findings of a short crisis following the collapse of the dot-com bubble. The patent stock correlation remains largely unaffected by the inclusion of regional dummies in Panel C,⁷⁹ which further indicate that firms in London, the East of England and Yorkshire have higher productivity levels than those in other regions of the UK (dummy coefficients not reported). The FE and LP models still indicate a positive significant relationship between patent stock and TFP levels, which is of similar magnitude for FE and LP, but only around one tenth in magnitude for the 2FE model (now only marginally significant). Note that the above patterns of significance/insignificance are unchanged when we use bootstrap standard errors (available on request).

⁷⁸We also ran regressions using more detailed geographical data for each firm in the sample which yielded very similar results to those presented here.

⁷⁹The regions and regional distribution of firms are as follows: East Midlands (14 firms), East of England (41), London (47), North-East (4), North-West (15), Scotland (21), South-East (101), South-West (22), Wales (19), West Midlands (22) and Yorkshire (12).

We repeated this regression exercise with variables in logarithms, which necessitates the inclusion of a dummy variable indicating all firms with zero patent stock (‘log patent stock’ is computed as $\log(\text{patent stock} + .001)$). Although not as dramatic as in the levels regressions, these result in substantial differences between the ACF and OLS estimates on the one hand and the FE and LP estimates on the other, mirroring the patterns described above. The coefficients on ‘log patent stock’, although statistically significant across all models are between twice and three times the magnitudes for the FE- and LP-based TFP level estimates than for those using ACF- and OLS-based estimates (results available on request). This reiterates our point of sensitivity of the result to the choice of production function estimator. It highlights our concerns that empirical findings may be highly dependent on the choice of the productivity estimator and therefore suggests that this choice is far from arbitrary when using finite samples.

4 Concluding remarks

In this survey we have provided a theoretical overview of the concerns over ‘transmission bias’ and the solutions suggested in the literature on Cobb-Douglas production function estimation at the firm-level. The discussion in Section 2 suggests that there are broadly conceived two approaches which may be used to address the bias arising from this problem. On the one hand, the dynamic panel estimators, notably the BB System GMM estimator, and on the other the ‘structural’ proxy estimators, most notably the ACF estimator. Both types of estimators can be theoretically motivated and the main difference between them lies in the specific assumptions made regarding the unobserved productivity shock. *A priori*, from a purely theoretical point of view, there is no reason to give preference to either of both estimators — and in fact they are conceptually very similar. This, however, does not automatically mean that choosing among these estimators is in practice down to personal preference.

In the second, empirical part of the paper the differences in the empirical results deriving from different empirical estimators is therefore the central focal point of the analysis. From this exercise we draw four major conclusions:

- (i) The UK firm-level data available in FAME appears to be far from ideal for the estimation of Cobb-Douglas production functions, given that only a small proportion of firms provide the necessary input and output data. Crucially, any production function regression sample constructed from FAME or similar datasets such as AMADEUS or ORBIS most likely cannot be thought of as a sound representation of the firm universe, and thus the external validity of the empirical findings is seriously undermined. This is an important observation in light of a considerable (and growing) number of empirical papers focusing on productivity estimates obtained from these data sources.
- (ii) Considering the substantial differences in the estimates of input coefficients for the two samples employed here, data availability, in particular with regard to investment data, should be an important concern when choosing an empirical estimator. More specifically, in data sources such as FAME the lack of reported investment data and the need to construct this variable from the book value of tangible assets undermines the application of the OP estimator. More generally, the data availability within-sample (missing observations,

unbalancedness) as well as the persistence of the data should be taken into account when considering the viability of various empirical estimators.

- (iii) In terms of economic and statistical significance of the empirical results, among the control function estimators only the ACF approach appears to produce reasonable input coefficients. The LP procedure results in dramatically underestimated labour coefficients, while the OP procedure yields too low a capital coefficient. Data properties (unbalanced panel with missing observations) hampered the bootstrapping procedure required for inference in the structural estimators, most significantly so in the ACF approach. Regarding the dynamic panel estimators, the results for AB and BB seem to be highly contingent on specifying the ‘correct’ instrument matrix, while the informativeness requirement could not be reliably established in all cases. As an important methodological difference between the two groups of estimators, the ability to test the assumptions imposed on the DGP (a feature which is well-known albeit underappreciated/seldomly applied in practice) presents an important advantage of the dynamic panel estimators over the structural approaches.
- (iv) We highlighted the very misleading conclusion about resulting TFP levels if the comparison is focused narrowly on the analysis of co-movement: correlation coefficients comparing TFP estimates computed from the various production function estimators are consistently high, while a basic analysis of the evidence of common shocks over time, in particular the collapse of the dot-com bubble in 2001, also results in seemingly very consistent results. These types of analysis might lead to the erroneous conclusion that despite the vast differences in factor input coefficients in the production function regressions, the resulting TFP estimates might still on balance tell a very similar story. However, when we conduct ‘TFP regressions’, analysing the relationship between productivity and innovation, we find substantial differences between models using TFP levels based on the OLS and ACF on the one hand, and those based on the FE and LP estimators.

From a theoretical point of view, any of the production function estimators presented here may appeal as equally suitable tools to estimate TFP. However, our review suggests that in practice, the untested assumption that different estimators produce statistically indistinguishable results is at best questionable and most likely outright wrong. We therefore call on practitioners to spend more time and effort on the investigation of data properties, to compare empirical results across different estimators making different assumptions about the DGP and thus to build up a general picture of the most likely processes driving the data and the most suitable empirical strategy to account for these difficulties.

References

- [1] Ai, Chunrong and Xiaohong Chen (2003): ‘Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions’, *Econometrica*, Vol. 71, No. 6, pp. 1795-1843.
- [2] Akerberg, Daniel, Lanier Benkard, Steven Berry, and Ariel Pakes: (2006): ‘Econometric Tools for Analyzing Market Outcomes’, Forthcoming chapter in *Handbook of Econometrics*, Volume 6.
- [3] Akerberg, Daniel, Kevin Caves and Garth Frazer (2006): ‘Structural Identification of Production Functions’, mimeo.
- [4] Angrist, Joshua and Jörn-Steffen Pischke (2009): *Mostly Harmless Econometrics — An Empiricist’s Companion*. Princeton University Press.
- [5] Arellano, Manuel and Stephen Bond (1991): ‘Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations’, *Review of Economic Studies*, Vol. 58, pp. 277-297.
- [6] Arnold, Jens (2005) ‘Productivity Estimation at the Plant Level: A practical guide’, unpublished mimeo, Summer 2005.
- [7] Bai, Jushan (2009): ‘Likelihood approach to small T dynamic panel models with interactive effects’, unpublished working paper, June 2009.
- [8] Baily, Martin Neil (1986): ‘Productivity Growth and Materials Use in U. S. Manufacturing’, *The Quarterly Journal of Economics* Vol. 101, No. 1, pp. 185-196.
- [9] Banerjee, Anindya, Markus Eberhardt and James Reade (2010): *Panel Econometrics for Worriers*, unpublished working paper, October 2010.
- [10] Baptist, Simon (2008) *Technology, Human Capital and Efficiency in Manufacturing Firms*, doctoral thesis, Department of Economics, University of Oxford, Trinity Term 2008.
- [11] Basu, Susanto and John Fernald (1995): ‘Are Apparent Productive Spillovers a Figment of Specification Error,’ *Journal of Monetary Economics*, Vol. 36 (1), pp. 165-188.
- [12] Basu, Susanto and John Fernald (1997): ‘Returns to Scale in U.S. Production: Estimates and Implications,’ *Journal of Political Economy*, Vol. 105 (2), pp. 249-283.
- [13] Benfratello, Luigi and Alessandro Sembenelli (2002): ‘Research joint ventures and firm level performance,’ *Research Policy*, Vol. 31, Issue 4, pp. 493-507.
- [14] Bernard, Andrew and Bradford Jensen (2004): ‘Exporting and Productivity in the USA’, *Oxford Review of Economic Policy*, Vol. 20, No. 3, pp. 343-357.
- [15] Blalock, Garrick and Paul Gertler (2004): ‘Learning from exporting revisited in a less developed setting ’, *Journal of Development Economics*, Vol. 75(2), pp. 397-416.
- [16] Blundell, Richard and Stephen Bond (2000): ‘GMM Estimation with persistent panel data: an application to production functions’, *Econometric Reviews*, Vol. 19, No. 3, 321-340.
- [17] Bond, Stephen, Céline Nauges and Frank Windmeijer (2005) *Unit roots: identification and testing in micro panels*. CeMMAP working papers CWP07/05, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- [18] Bond, Stephen and Måns Söderbom (2005) *Adjustment costs and the identification of Cobb Douglas production functions*. IFS Working Papers W05/04, Institute for Fiscal Studies.
- [19] Bruno, Michael (1984): ‘Raw Materials, Profits, and the Productivity Slowdown,’ *Quarterly Journal of Economics*, Vol. 99, No. 1, pp. 1-30.
- [20] Bun, Maurice and Jan Kiviet (2003) ‘On the diminishing returns of higher-order terms in asymptotic expansions of bias’, *Economics Letters* Vol. 79(2), pp. 145-152.
- [21] Bun, Maurice and Frank Windmeijer (2009): ‘The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models’, Tinbergen Institute Discussion Paper, BTI 2009-086/4.
- [22] Buettner, Thomas (2005): ‘R&D and the dynamics of productivity’, mimeo LSE.
- [23] Breusch, Trevor, Hailong Qian, Peter Schmidt and Donald Wyhowski (1999) ‘Redundancy of moment conditions’, *Journal of Econometrics*, Vol. 91(1), pp. 89-111.
- [24] Bruno, Giovanni (2005): ‘Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models’, *Economics Letters*, Vol. 87(3), pp. 361-366.

- [25] Coakley, Jerry, Ana-Maria Fuertes and Ron Smith (2006): ‘Unobserved heterogeneity in panel time series models’, *Computational Statistics & Data Analysis*, Vol.50(9), pp. 2361-2380.
- [26] de Mel, Suresh, David McKenzie and Christopher Woodruff (2010): ‘Wage Subsidies for Microenterprises’, *American Economic Review, Papers & Proceedings* Vol. 100(2), pp.614-18.
- [27] de Mel, Suresh, David McKenzie and Christopher Woodruff (2009): ‘Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns’, *American Economic Journal: Applied Economics* Vol. 1(3), pp.1-32.
- [28] de Mel, Suresh, David McKenzie and Christopher Woodruff (2008): ‘Returns to Capital in Microenterprises: Evidence from a Field Experiment’, *The Quarterly Journal of Economics*, Vol. 123(4), pp.1329-1372.
- [29] Del Gatto, Massimo, Adriana Di Liberto and Carmelo Petraglia (2010): ‘Measuring Productivity’, *Journal of Economic Surveys*, forthcoming.
- [30] Disney, Richard, Jonathan Haskel and Ylva Heden (2003): ‘Restructuring and Productivity Growth in UK Manufacturing’, *The Economic Journal*, Vol. 113, No. 489, pp. 666-694.
- [31] Doraszelski, Ulrich and Jordi Jaumandreu (2008): ‘R&D and Productivity: Estimating Production Functions when Productivity is Endogenous’, CEPR Discussion Papers 6636.
- [32] Eberhardt, Markus and Francis Teal (2010): ‘Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics’, *Journal of Economic Surveys* (forthcoming).
- [33] Eberhardt, Markus, Christian Helmers and Hubert Strauss (2009): ‘Do Spillovers Matter When Estimating Private Returns to R&D?’, EIB Economic and Financial Reports.
- [34] Faggio Giulia, Kjell Salvanes, and John Van Reenen (2007): ‘The Evolution of Inequality in Productivity and Wages: Panel Data Evidence’, NBER Working Paper No. 13351.
- [35] Greenstone, Michael, Richard Hornbeck and Enrico Moretti (2010): ‘Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings’, *Journal of Political Economy*, forthcoming.
- [36] Greenstreet, David (2007): ‘Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules’, University of Oxford Department of Economics Working Paper No. 345.
- [37] Griffith, Rachel, Rupert Harrison and John van Reenen (2006): ‘How Special Is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing’, *American Economic Review*, Vol. 96(5), pp. 1859-1875.
- [38] Griliches, Zvi (1987): ‘R&D, Patents, and Productivity’, NBER.
- [39] Griliches, Zvi and Jacques Mairesse (1995): ‘Production Functions: The Search for Identification’, NBER Working Papers No. 5067.
- [40] Harris, Richard and Qian Cher Li (2008): ‘Evaluating the Contribution of Exporting to UK Productivity Growth: Some Microeconomic Evidence’, *The World Economy*, pp. 212-235.
- [41] Haskel, Jonathan, Sonia Pereira and Matthew Slaughter (2007): ‘Does inward foreign direct investment boost the productivity of domestic firms?’, *Review of Economics and Statistics*, Vol. 89, Issue 3, pp. 482-496.
- [42] Helmers, Christian and Mark Rogers (2010): ‘Innovation and the Survival of New Firms in the UK’, *Review of Industrial Organization*, Vol. 36, pp. 227-248.
- [43] Helmers, Christian, Philipp Schautschick and Mark Rogers (2010): ‘Firm-Level Analysis of Intellectual Property in the UK: The Oxford Firm-level Intellectual Property Database’, mimeo University of Oxford.
- [44] Hsiao, Cheng (2003): *Analysis of Panel Data*, 2nd edition, Cambridge University Press.
- [45] Javorcik, Beata (2004): ‘Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages’, *The American Economic Review*, Vol. 94, No. 3, pp. 605-627.
- [46] Javorcik, Beata and Yue Li (2008): ‘Do the Biggest Aisles Serve a Brighter Future? Global Retail Chains and Their Implications for Romania’, World Bank Policy Research Working Paper No. 4650.
- [47] Jorgenson, Dale and Zvi Griliches (1967): ‘The Explanation of Productivity Change’, *Review of Economic Studies*, Vol. 34, No. 3, pp. 249-283.
- [48] Judson, Ruth and Ann Owen (1999) ‘Estimating dynamic panel data models: a guide for macroeconomists’, *Economics Letters*, Vol. 65(1), pp. 9-15.

- [49] Kiviet, Jan (1995) ‘On bias, inconsistency, and efficiency of various estimators in dynamic panel data models’, *Journal of Econometrics*, Vol. 68(1), pp. 53-78.
- [50] Klette, Tor Jakob and Zvi Griliches (1996): ‘The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous’, *Journal of Applied Econometrics*, Vol. 11, No. 4, pp. 343-361.
- [51] Levinsohn, James and Amil Petrin (1999): ‘When Industries Become More Productive, Do Firms? Investigating Productivity Dynamics’, NBER Working Paper No. W6893.
- [52] Levinsohn, James and Amil Petrin (2003): ‘Estimating Production Functions Using Inputs to Control for Unobservables’, *Review of Economic Studies*, Vol. 70, No. 2, pp. 317-341.
- [53] Levinsohn, James, Amil Petrin and Brian Poi (2004): ‘Production Function Estimation in Stata Using Inputs to Control for Unobservables’, *Stata Journal*, Vol. 4, Issue 2, pp. 113-123.
- [54] Mairesse, Jacques and Jordi Jaumandreu (2005): ‘Panel-data Estimates of the Production Function and the Revenue Function: What Difference Does It Make?’, *Scandinavian Journal of Economics*, Vol. 107(4), pp. 651-672.
- [55] Marschak J. and W.H. Andrews (1944): ‘Random Simultaneous Equations and the Theory of Production,’ *Econometrica*, Vol. 12, No. 3-4, pp. 143-205.
- [56] Nickell, Steven (1981): ‘Biases in Dynamic Models with Fixed Effects’, *Econometrica*, Vol. 49, pp. 1417-1426.
- [57] Olley, Steven and Ariel Pakes (1996): ‘The Dynamics Of Productivity In The Telecommunications Equipment Industry’, *Econometrica*, Vol. 64, pp. 1263-1297.
- [58] Pagan, Adrian (1984): ‘Econometric Issues in the Analysis of Regressions with Generated Regressors’, *International Economic Review*, Vol. 25(1), pp. 221-247.
- [59] Pavcnik, Nina (2002): ‘Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants’, *The Review of Economic Studies*, Vol. 69, No. 1, pp. 245-276.
- [60] Robinson, Peter (1988): ‘Root-n-Consistent Semiparametric Regression’, *Econometrica*, Vol. 56, No. 4, pp. 931-954.
- [61] Roodman, David (2006): ‘How to Do *xtabond2*: An Introduction to “Difference” and “System” GMM in Stata’, Working Papers 103, Center for Global Development.
- [62] Roodman, David (2009): ‘A Note on the Theme of Too Many Instruments’, *Oxford Bulletin of Economics and Statistics*, Vol. 71(1), pp. 135-158.
- [63] Shashua L., A. Melnik and Y. Goldschmidt (1974): ‘A note on stock and flow capital inputs in the production function of multiproduct firms’, *Applied Economics*, Vol. 6, pp. 229-233
- [64] Söderbom, Måns and Francis Teal (2003) *Openness and Human Capital as Sources of Productivity Growth: An Empirical Investigation*, CSAE Working Paper Series CSAE WPS/2003-06.
- [65] Solow, Robert (1956): ‘A Contribution to the Theory of Economic Growth,’ *Quarterly Journal of Economics*, Vol. 70, No. 1, pp. 65-94.
- [66] Syverson, Chad (2004): ‘Market Structure and Productivity: A Concrete Example’, *Journal of Political Economy*, Vol. 112, No. 6, pp. 1181-1222.
- [67] Syverson, Chad (2010): ‘What Determines Productivity’, *Journal of Economic Literature*, forthcoming.
- [68] Tybout, James and Daniel Westbrook (1995): ‘Trade liberalization and the dimensions of efficiency change in Mexican manufacturing industries’, *Journal of International Economics*, Vol. 39(1-2), pp. 53-78.
- [69] van Beveren, Ilke (2010): ‘Total Factor Productivity Estimation: A Practical Review’, *Journal of Economic Surveys*, forthcoming.
- [70] van Biesebroeck, Johannes (2007): ‘Robustness of Productivity Estimates’, *Journal of Industrial Economics*, Volume LV, No. 3, pp. 529-569.
- [71] Wallis, Gavin and Alex Turvey (2009): ‘Volume of Capital Services: estimates for 1950 to 2007’, *Economic & Labour Market Review*, Vol. 3, No. 1, pp. 53-61.
- [72] Wooldridge, Jeffrey (2009): ‘On Estimating Firm-Level Production Functions Using Proxy Variables To Control For Unobservables’, *Economics Letters*, Vol. 104(3), pp. 112-114.

Table 1: Industry descriptions

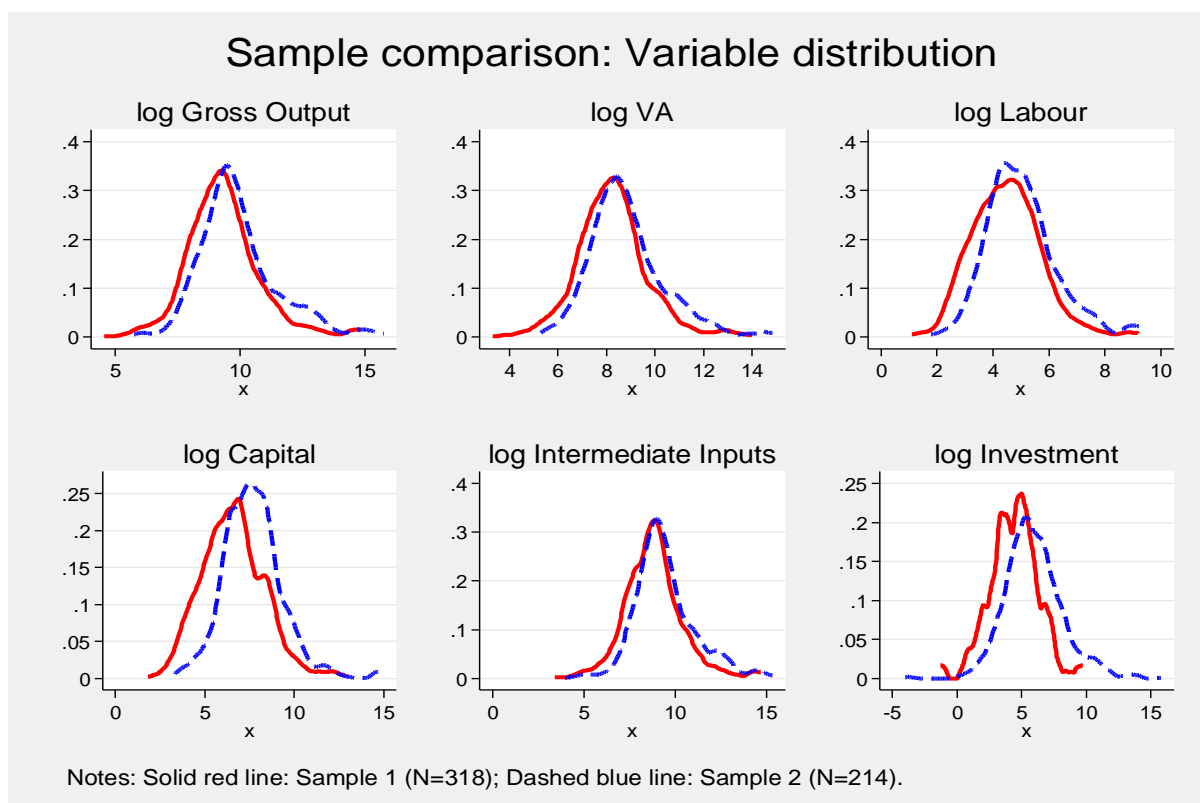
SIC	Description	No. firms in FAME	No. firms in sample	%
	Manufacture of			
32	Radio, television and communication equipment and apparatus	3,150	318	10.09%
321	Electronic valves and tubes and other electronic components	1,580	158	10.00%
322	Television and radio transmitters	604	92	15.23%
323	Television and radio receivers	966	68	7.04%

Table 2: Summary statistics

	Mean	Std.Dev.	Min	Max	Obs
SAMPLE 1					
$\ln Y_{it}$	8.568	1.529	3.307	14.828	1,742
$\ln L_{it}$	4.774	1.265	1.098	9.313	1,742
$\ln K_{it}$	7.184	1.799	1.828	14.881	1,742
$\ln M_{it}$	9.152	1.670	3.414	15.299	1,742
SAMPLE 2					
$\ln Y_{it}$	8.808	1.558	4.688	14.828	958
$\ln L_{it}$	5.003	1.249	1.792	9.313	958
$\ln K_{it}$	7.679	1.663	3.315	14.881	958
$\ln I_{it}$	5.926	2.232	-4.013	15.800	958
$\ln M_{it}$	9.414	1.682	3.959	15.299	958

Notes: Sample 1 represents the full $N = 318$ firm sample, Sample 2 is the selection restricted to firms with investment data.

Figure 1: Density Plots — Comparing samples



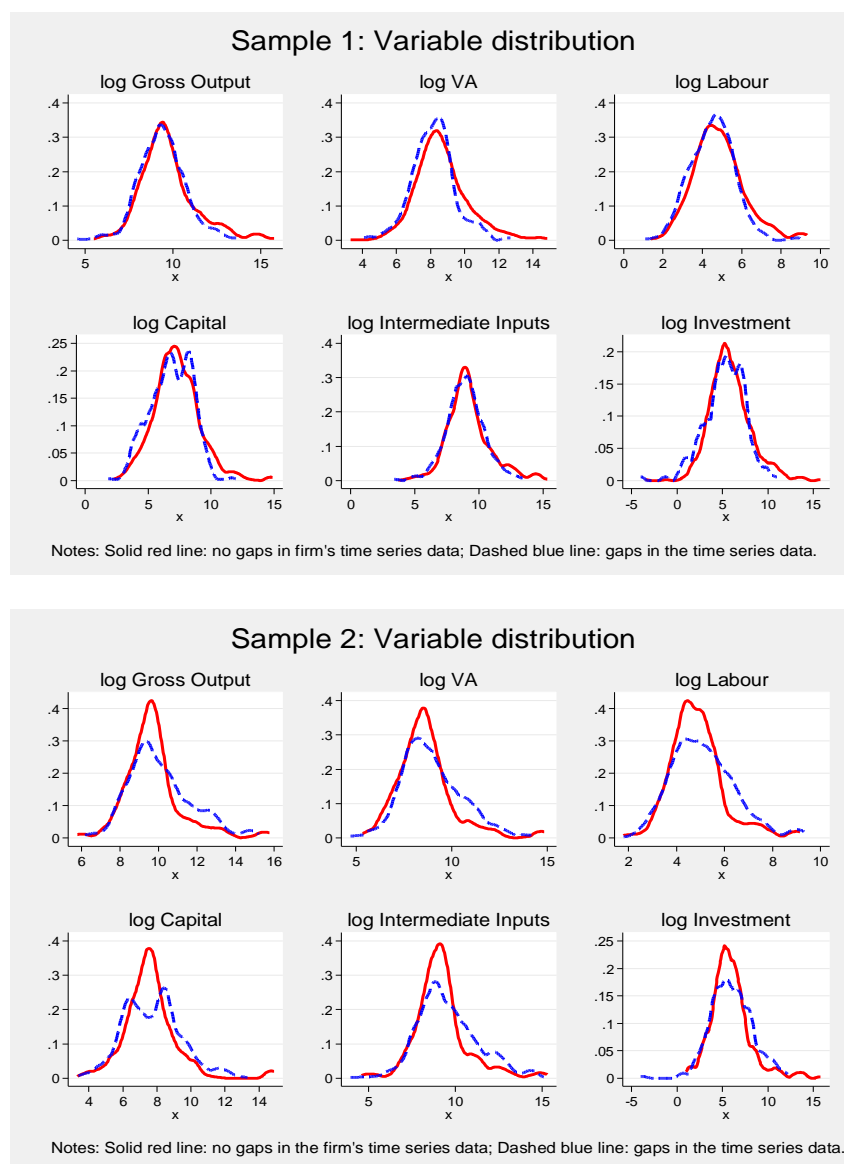
Notes: The density plots (bandwidth .29 imposed on all plots) compare the distribution of the production function variables between (a) firms in the larger Sample 1 (solid red lines), and (b) firms in the restricted Sample 2 (dashed blue lines).

Table 3: Data coverage

T_i	SAMPLE 1			SAMPLE 2		
	obs	share	firms	obs	share	firms
2	88	5.1%	44	78	8.1%	39
3	108	6.2%	36	96	10.0%	32
4	144	8.3%	36	204	21.3%	51
5	160	9.2%	32	125	13.1%	25
6	180	10.3%	30	168	17.5%	28
7	406	23.3%	58	175	18.3%	25
8	656	37.6%	82	112	11.7%	14
Total	1,742	100.0%	318	958	100.0%	214

Notes: Sample 1 represents the full $N = 318$ firm sample, Sample 2 is the selection restricted to firms with investment data.

Figure 2: Density Plots — Comparing samples



Notes: The density plots (bandwidth .29 imposed on all plots) compare the distribution of the production function variables between (a) firms for which we have data without gaps in their time-series data on the one hand and (b) firms which do have one or two gaps in their time-series data on the other.

Figure 3: Scatter plots — Value added and inputs

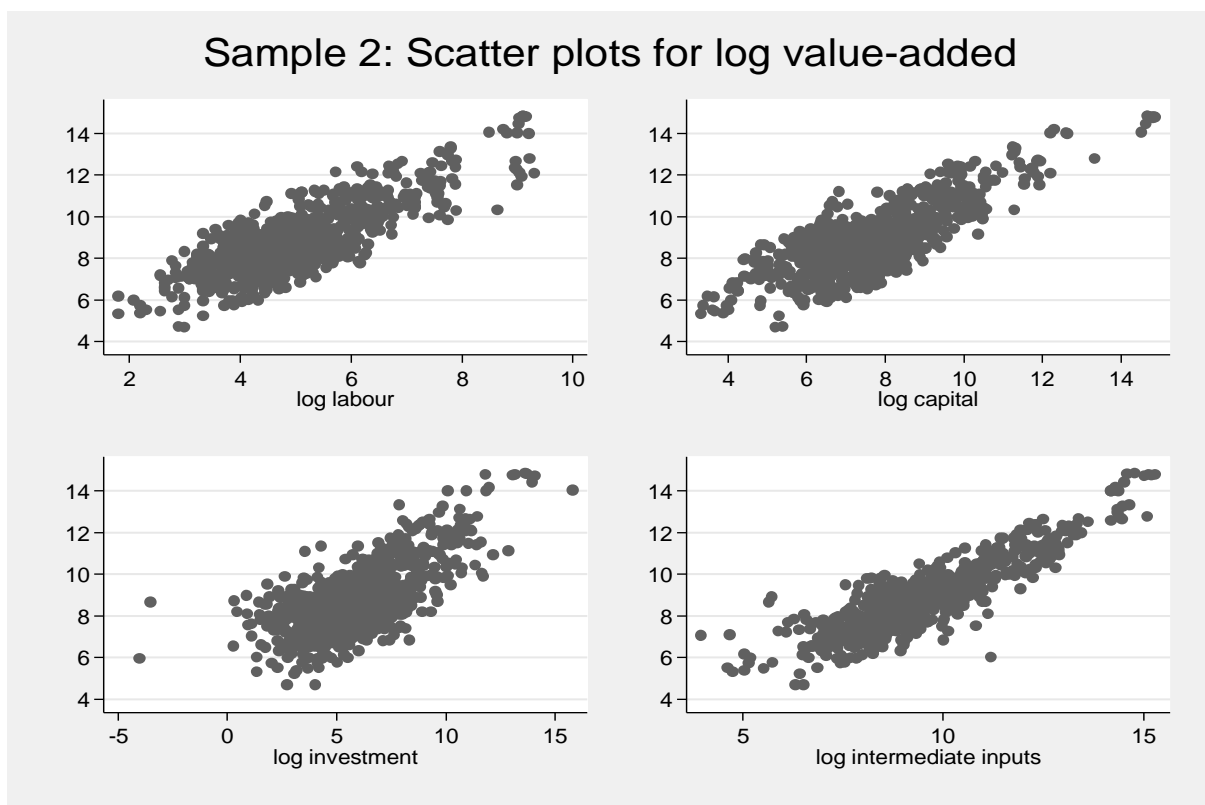
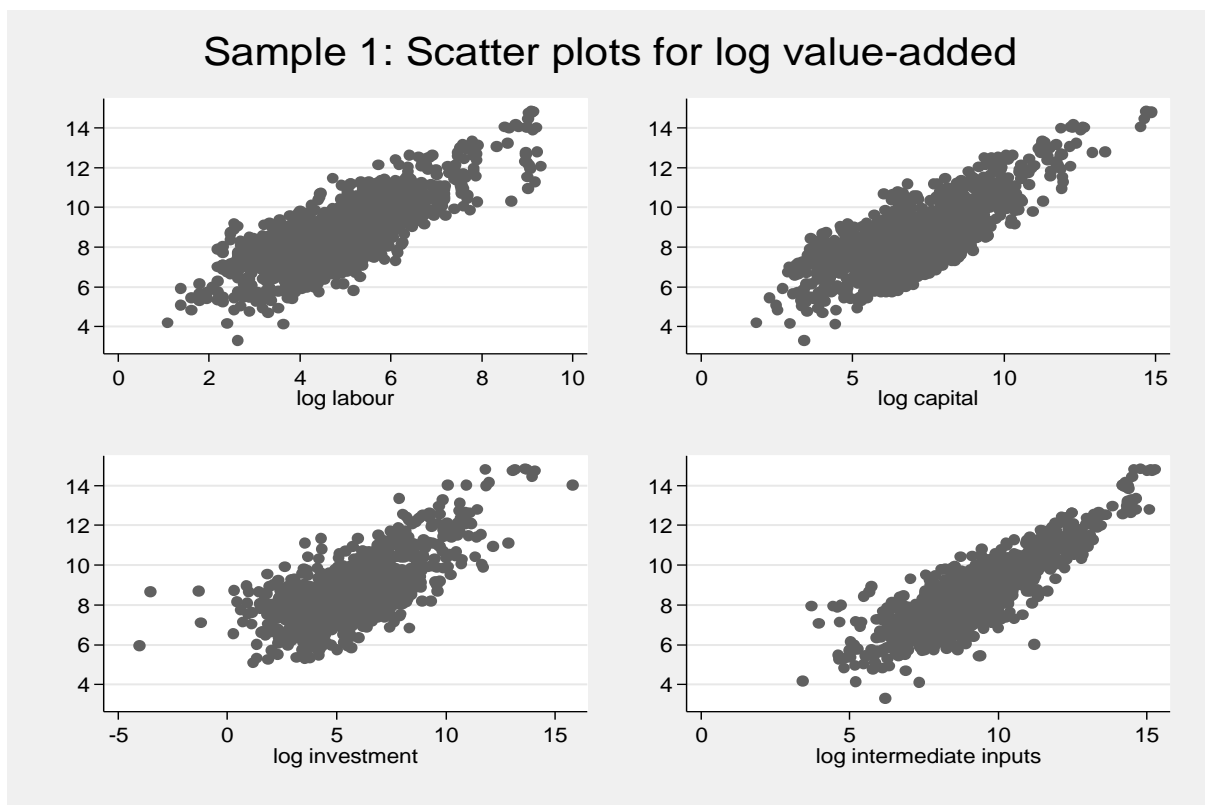


Table 4: AR(1) Estimates — Sample 1

	OLS	FE	AB(t-2)	AB(t-3)	BB(t-2)	BB(t-3)
	(1)	(2)	(3)	(4)	(5)	(6)
VALUE ADDED						
$\ln Y_{it-1}$	0.963** (0.008)	0.454** (0.062)	0.663** (0.155)	0.647** (0.156)	0.967** (0.067)	0.685** (0.109)
AB Test AR(1)	0.983	0.580	0.002	0.003	0.000	0.000
AB Test AR(2)	0.082	0.000	0.331	0.338	0.178	0.302
Sargan Test			0.222	0.417	0.103	0.014
Diff Sargan					0.051	0.015
LABOUR						
$\ln L_{it-1}$	0.983** (0.006)	0.685** (0.037)	0.527** (0.160)	0.663** (0.208)	1.175** (0.062)	1.170** (0.068)
AB Test AR(1)	0.000	0.440	0.383	0.296	0.000	0.007
AB Test AR(2)	0.593	0.000	0.207	0.230	0.108	0.117
Sargan Test			0.271	0.221	0.054	0.324
Diff Sargan					0.002	0.575
CAPITAL						
$\ln K_{it-1}$	0.992** (0.006)	0.699** (0.026)	0.487** (0.130)	0.645** (0.176)	1.131** (0.045)	1.148** (0.051)
AB Test AR(1)	0.000	0.116	0.044	0.031	0.000	0.000
AB Test AR(2)	0.293	0.009	0.955	0.902	0.733	0.730
Sargan Test			0.382	0.497	0.117	0.210
Diff Sargan					0.035	0.109
Observations	1,361	1,361	1,014	1,014	1,361	1,361

Notes: Estimators employed are OLS, FE — Firm fixed Effects (within), AB — Arellano and Bond (1992) Difference GMM, BB — Blundell and Bond (1998) System GMM. All models include year dummies (coefficients not reported). We report p -values for all test statistics (serial correlation, Sargan and Difference-in-Sargan tests). Standard error in parentheses, **, * indicate statistical significance at the 5% and 1% level respectively.

Table 5: AR(1) Estimates — Sample 2

	OLS	FE	AB(t-2)	AB(t-3)	BB(t-2)	BB(t-3)
	(1)	(2)	(3)	(4)	(5)	(6)
VALUE ADDED						
$\ln Y_{it-1}$	0.965** (0.013)	0.439** (0.083)	0.416* (0.171)	0.333* (0.164)	1.136** (0.102)	0.873** (0.125)
AB Test AR(1)	0.050	0.835	0.146	0.250	0.006	0.009
AB Test AR(2)	0.223	0.011	0.649	0.774	0.212	0.264
Sargan Test			0.041	0.050	0.013	0.008
Diff Sargan					0.018	0.021
LABOUR						
$\ln L_{it-1}$	0.968** (0.011)	0.496** (0.098)	0.360** (0.178)	0.397 (0.229)	0.871** (0.087)	0.662** (0.190)
AB Test AR(1)	0.002	0.617	0.963	0.939	0.007	0.335
AB Test AR(2)	0.512	0.083	0.525	0.494	0.142	0.179
Sargan Test			0.370	0.330	0.487	0.797
Diff Sargan					0.413	0.960
CAPITAL						
$\ln K_{it-1}$	0.967** (0.009)	0.574** (0.076)	0.402** (0.123)	0.265* (0.121)	0.993** (0.045)	0.730** (0.226)
AB Test AR(1)	0.061	0.479	0.328	0.605	0.103	0.236
AB Test AR(2)	0.422	0.080	0.101	0.108	0.166	0.157
Sargan Test			0.258	0.182	0.194	0.495
Diff Sargan					0.105	0.278
Observations	613	613	361	361	613	613

Notes: See Table 4 for details.

Table 6: Static production function estimates — Sample 1

	OLS		FE		FD		LP	ACF
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln L_{it}$	0.693** (0.074)	0.713** (0.076)	0.488** (0.103)	0.695** (0.084)			0.200** (0.076)	0.612 (0.536)
$\Delta \ln L_{it}$					0.500** (0.095)	0.572** (0.088)		
$\ln K_{it}$	0.209** (0.057)	0.194** (0.057)	0.119* (0.060)	0.140** (0.046)			0.333** (0.090)	0.324 (0.406)
$\Delta \ln K_{it}$					0.079 (0.047)	0.115** (0.042)		
Year dummies	Included		Included		Included			
\sum coefficients	0.902	0.907	0.607	0.835	0.579	0.689	0.533	0.936
Test CRS	0.020	0.027	0.000	0.036	0.000	0.002	0.000	
AB Test AR(1)	0.000	0.000	0.000	0.000	0.683	0.056		
AB Test AR(2)	0.000	0.000	0.825	0.000	0.113	0.002		
RMSE	0.909	0.864	0.437	0.353	0.439	0.405	n/a	n/a
Observations	1,742		1,742		1,361		1,742	1,742

Notes: Estimators employed are OLS, FE — Firm fixed Effects (within), FD — OLS with variables in first-differences, LP — Levinsohn and Petrin (2003), ACF — Akerberg, Caves and Frazer (2006). For OLS, FE, and FD, standard errors clustered by firm, for OP, LP, and ACF we employ bootstrapped standard errors (OP and LP: 400 replications; ACF: 100 replications). CRS — p -values for H_0 (Constant Returns to Scale). AB Test — Arellano and Bond (1992) serial correlation test), p -values for H_0 (no serial correlation of order indicated). RMSE — root mean squared error, not computed for the multi-step structural estimators.

Table 7: Static production function estimates — Sample 2

	OLS		FE		FD		OP	LP	ACF
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\ln L_{it}$	0.643** (0.086)	0.651** (0.089)	0.567** (0.128)	0.824** (0.105)			0.590** (0.089)	0.135 (0.082)	0.613* (0.262)
$\Delta \ln L_{it}$					0.684** (0.115)	0.783** (0.114)			
$\ln K_{it}$	0.310** (0.073)	0.302** (0.074)	0.258** (0.071)	0.162** (0.053)			0.067 (0.044)	0.428** (0.130)	0.309 (0.359)
$\Delta \ln K_{it}$					0.148* (0.064)	0.061 (0.062)			
Year dummies	Included		Included		Included		Included		
\sum coefficients	0.952	0.953	0.825	0.986	0.832	0.843	0.657	0.563	0.922
Test CRS	0.283	0.291	0.133	0.874	0.148	0.116		0.000	
AB Test AR(1)	0.000	0.000	0.000	0.022	0.044	0.580			
AB Test AR(2)	0.000	0.000	0.079	0.000	0.400	0.007			
RMSE	0.876	0.829	0.399	0.311	0.395	0.362	n/a	n/a	n/a
Observations	958		958		613		958	958	958

Notes: Estimators employed are OLS, FE — Firm fixed Effects (within), FD — OLS with variables in first-differences, OP — Olley and Pakes (1996), LP — Levinsohn and Petrin (2003), ACF — Akerberg, Caves and Frazer (2006). For OLS, FE, and FD, standard errors clustered by firm, for OP, LP, and ACF we employ bootstrapped standard errors (OP and LP: 400 replications; ACF: 100 replications). CRS — p -values for H_0 (Constant Returns to Scale). RMSE — root mean squared error, not computed for the multi-step structural estimators.

Figure 4: Productivity as a function of capital and investment – OP (Sample 2)

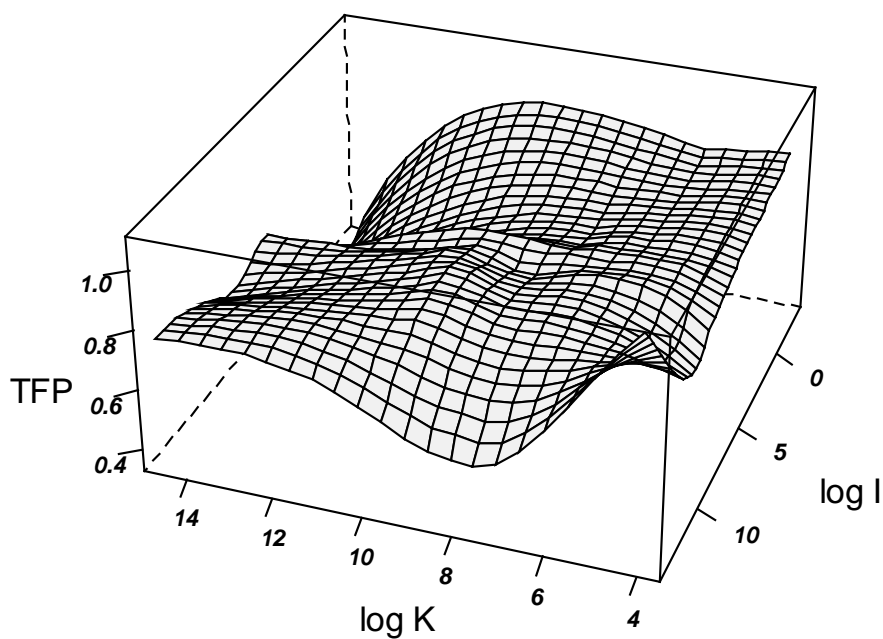


Figure 5: Productivity as a function of capital and intermediate inputs – LP (Sample 2)

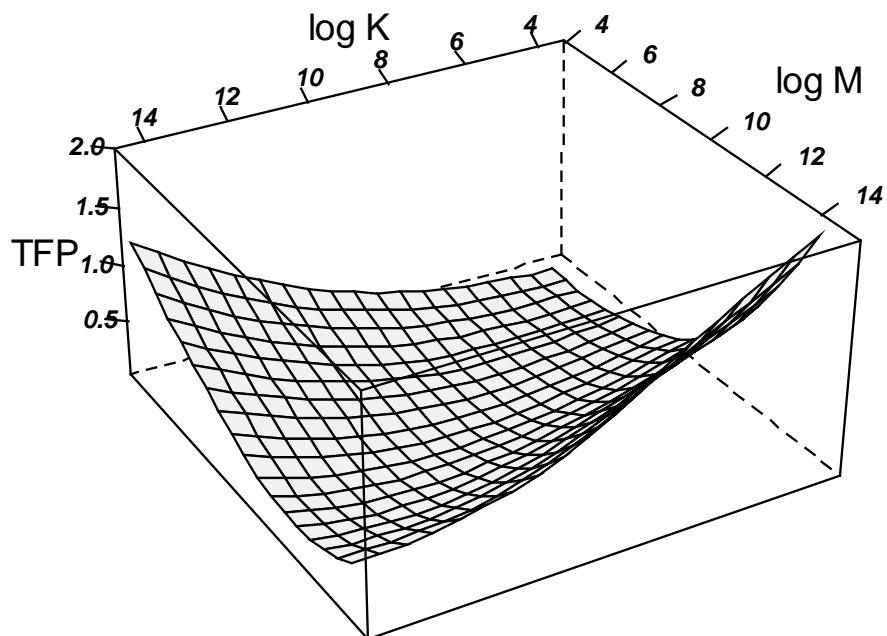


Table 8: Dynamic production function estimates — Sample 1

	OLS	FE	AB ₀	AB _*	BB ₀	BB _*
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln L_{it}$	0.616 (0.080)	0.721 (0.102)	0.710 (0.238)	0.575 (0.248)	0.551 (0.186)	0.500 (0.189)
$\ln L_{it-1}$	-0.527 (0.081)	-0.237 (0.081)	-0.265 (0.147)	-0.010 (0.157)	-0.406 (0.149)	-0.302 (0.154)
$\ln \bar{K}_{it}$	0.104 (0.041)	0.063 (0.048)	0.157 (0.155)	0.393 (0.175)	0.137 (0.149)	0.156 (0.075)
$\ln \bar{K}_{it-1}$	-0.085 (0.039)	-0.039 (0.036)	-0.113 (0.061)	-0.175 (0.071)	-0.052 (0.052)	-0.065 (0.054)
$\ln Y_{it-1}$	0.883 (0.014)	0.308 (0.059)	0.488 (0.100)	0.356 (0.111)	0.739 (0.059)	0.646 (0.066)
Year dummies	Included in all models					
$\hat{\beta}_l$	0.728** (0.066)	0.694** (0.095)	0.647** (0.197)	0.604** (0.221)	0.672** (0.154)	0.624** (0.155)
$\hat{\beta}_k$	0.133** (0.039)	0.066 (0.048)	0.177 (0.103)	0.304* (0.148)	0.106 (0.058)	0.129 (0.062)
$\hat{\rho}$	0.893** (0.013)	0.317** (0.056)	0.501** (0.095)	0.373** (0.089)	0.735** (0.055)	0.660** (0.062)
AB Test AR(1)	0.503	0.591	0.000	0.001	0.000	0.000
AB Test AR(2)	0.025	0.000	0.725	0.461	0.770	0.858
Sargan Test			0.045	0.822	0.018	0.159
Diff Sargan					0.013	0.000
COMFAC	0.053	0.746	0.825	0.188	0.282	0.393
# Instruments			75	69	95	88
RMSE	0.393	0.294	0.489	0.647	0.413	0.445
Observations	1,361	1,361	1,014	1,014	1,361	1,361

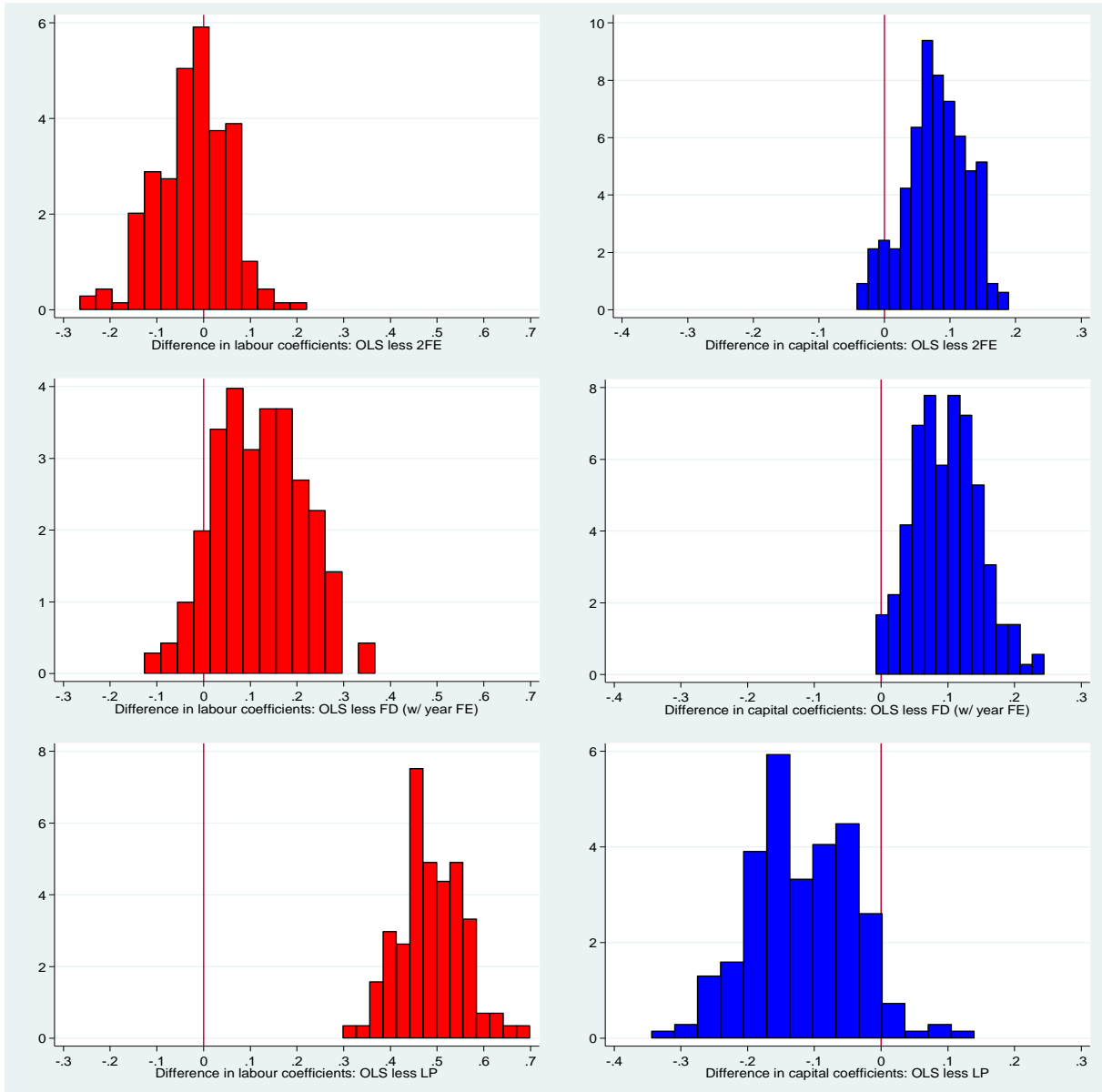
Notes: Estimators employed are OLS, FE — Firm fixed Effects (within), AB — Arellano and Bond (1992) Difference GMM, BB — Blundell and Bond (1998) System GMM. For OLS and FE standard errors clustered by firm, for AB and BB we use the 2-step estimator with the Windmeijer (2005) correction. COMFAC is a minimum distance test of the common factor restrictions. p -values reported for all test statistics. In (1) we obtain long-run coefficients (standard errors) $\hat{\beta}_l = .790 (.182)$ and $\hat{\beta}_k = .161 (.131)$ when the common factor restrictions are not imposed. AB₀ and BB₀ treat labour and lagged value-added as endogenous and capital as predetermined. AB_{*} and BB_{*} are the preferred specifications of the respective estimators, based on diagnostic testing (see discussion in the main text). For the BB estimator the Difference-in-Sargan statistics test the validity of all the additional instruments employed for the levels equation. All long-run coefficients are computed using the Minimum Distance estimator (code from Måns Söderbom) and employing the Delta method to obtain standard errors (absolute values reported). **, * indicate statistical significance at the 1% and 5% level. RMSE — root mean squared error, computed for unrestricted model in the upper panel of the table.

Table 9: Dynamic production function — Sample 2

	OLS	FE	AB ₀	AB _*	BB ₀	BB _*
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln L_{it}$	0.818 (0.100)	0.868 (0.122)	0.891 (0.233)	0.744 (0.239)	0.991 (0.133)	0.752 (0.193)
$\ln L_{it-1}$	-0.726 (0.104)	-0.323 (0.089)	-0.521 (0.167)	-0.397 (0.155)	-0.752 (0.136)	-0.517 (0.156)
$\ln \bar{K}_{it}$	0.069 (0.062)	0.032 (0.063)	-0.190 (0.183)	0.016 (0.167)	0.044 (0.081)	0.309 (0.130)
$\ln \bar{K}_{it-1}$	-0.050 (0.060)	0.009 (0.059)	0.075 (0.088)	0.064 (0.079)	-0.046 (0.083)	-0.036 (0.124)
$\ln Y_{it-1}$	0.897 (0.021)	0.344 (0.078)	0.371 (0.112)	0.265 (0.118)	0.885 (0.055)	0.538 (0.098)
Year dummies	Included in all models					
$\hat{\beta}_l$	0.899** (0.082)	0.885** (0.113)	1.072** (0.214)	0.857** (0.232)	1.040** (0.117)	0.767** (0.189)
$\hat{\beta}_k$	0.085 (0.061)	0.031 (0.064)	0.028 (0.147)	0.109 (0.147)	-0.005 (0.071)	0.341** (0.129)
$\hat{\rho}$	0.908** (0.019)	0.353** (0.067)	0.381** (0.111)	0.334** (0.113)	0.869** (0.054)	0.607** (0.086)
AB Test AR(1)	0.293	0.216	0.004	0.031	0.001	0.005
AB Test AR(2)	0.132	0.001	0.788	0.936	0.682	0.823
Sargan Test			0.244	0.414	0.096	0.379
Diff Sargan					0.440	0.205
COMFAC	0.336	0.935	0.056	0.113	0.071	0.305
# Instruments			75	69	95	68
RMSE	0.352	0.219	0.767	0.714	0.376	0.475
Observations	613	613	361	361	613	613

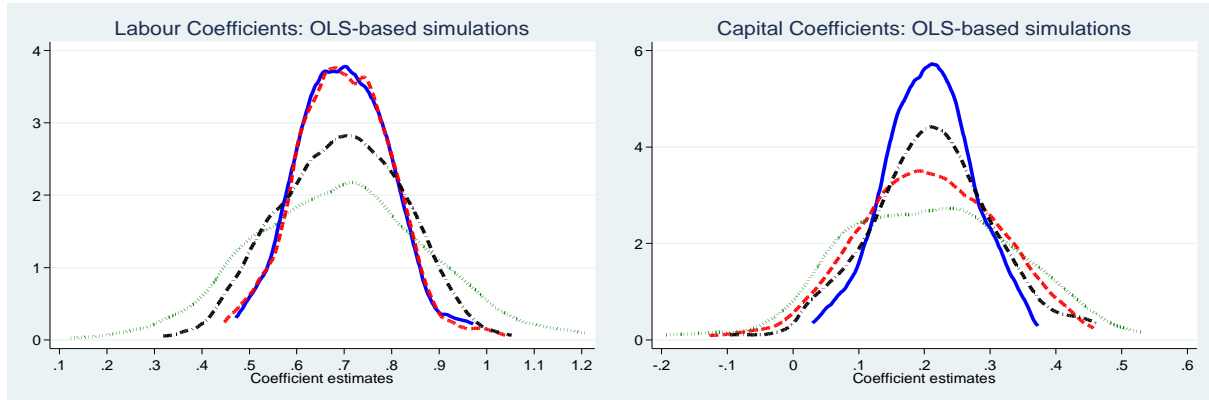
Notes: See Table 8 for details. In (3) we obtain long-run coefficients (standard errors) $\hat{\beta}_l = .588 (.413)$ and $\hat{\beta}_k = -.183 (.260)$ when the common factor restrictions are not imposed. In (5) we get $\hat{\beta}_l = 2.071 (1.253)$ and $\hat{\beta}_k = -.024 (.722)$.

Figure 6: Simulated Data — Difference in Slope Coefficients (Sample 1)



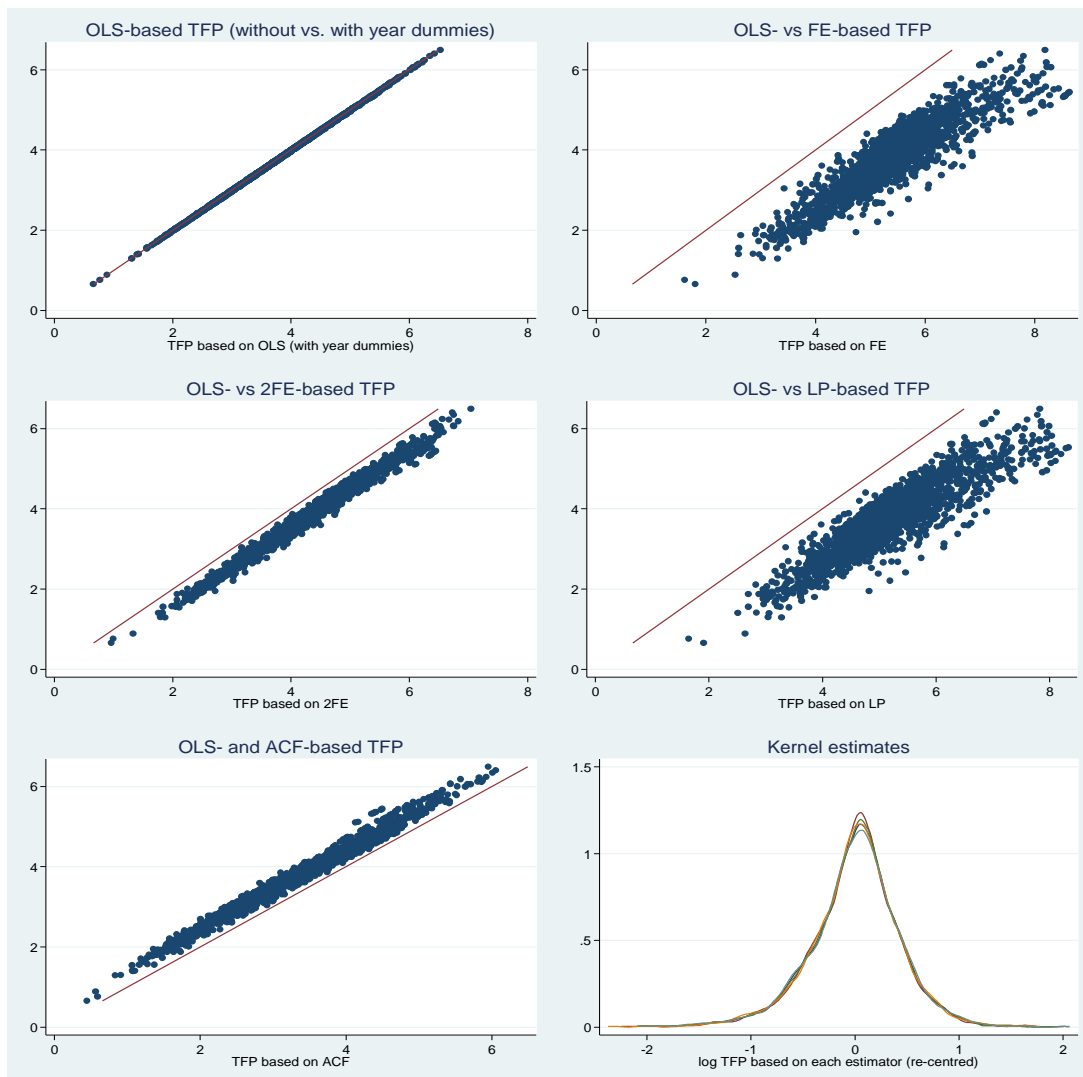
Notes: The histograms show (for labour and capital in the left and right column respectively) the distribution of the deviation from the OLS coefficient estimates for the two-way fixed effects (top), first difference with year fixed effects (middle) and LP estimator (bottom) for 200 iterations. For each estimator we created 200 bootstrap samples (independently across estimators) and collect the slope coefficient estimates $\hat{\beta}_l$ and $\hat{\beta}_k$. We then subtract the vector for each of the aforementioned estimators from that of the OLS estimator. For ease of comparison we fixed the range of the labour plots and the capital plots to be the same across estimators.

Figure 7: Simulated Data — Slope Coefficients (Sample 1)



Notes: The graphs show the estimated distributions (Epanechnikov kernel, bandwidth differs across series for improved visibility) of the labour (left column) and capital (right column) slope estimates for the (i) OLS (blue solid line), (ii) 2FE (black dash-dots), (iii) FD (green dots) and (iv) LP (red dashes) estimators. Methodology following Levinsohn & Petrin (2003).

Figure 8: Simulated Data — Difference in TFP estimates (Sample 1)



Notes: In each scatter plot we graph TFP-level estimates based on the OLS regression (y-axis) against those from the named regression model. The kernel graph provides kernel density estimates for the TFP-level estimates, which have been 'within-transformed' to centre them around zero.

Table 10: TFP level correlations — Sample 1

	OLS	OLS†	FE	2FE	LP	ACF		OLS	OLS†	FE	2FE	LP	ACF
‘raw’							‘within’						
OLS	1.000						OLS	1.000					
OLS†	1.000	1.000					OLS†	1.000	1.000				
FE	0.745	0.750	1.000				FE	0.862	0.864	1.000			
2FE	0.974	0.976	0.871	1.000			2FE	0.986	0.987	0.929	1.000		
LP	0.771	0.773	0.975	0.878	1.000		LP	0.853	0.852	0.978	0.908	1.000	
ACF	0.981	0.978	0.625	0.913	0.670	1.000	ACF	0.986	0.984	0.788	0.947	0.801	1.000
	OLS	OLS†	FE	2FE	LP	ACF		OLS	OLS†	FE	2FE	LP	ACF
‘demeaned’							both						
OLS	1.000						OLS	1.000					
OLS†	1.000	1.000					OLS†	1.000	1.000				
FE	0.736	0.741	1.000				FE	0.847	0.849	1.000			
2FE	0.972	0.974	0.868	1.000			2FE	0.984	0.985	0.923	1.000		
LP	0.761	0.763	0.974	0.873	1.000		LP	0.836	0.835	0.976	0.898	1.000	
ACF	0.979	0.977	0.609	0.907	0.654	1.000	ACF	0.984	0.981	0.762	0.938	0.776	1.000

Notes: We present the Pearson correlation coefficients for the TFP level estimates (‘raw’, i.e. not logs) computed from the six estimators indicated ($n = 1,742$, $N = 318$). OLS † indicates the TFP estimates from an OLS model with year dummies. For ‘within’ we use TFP estimates in deviation from the firm-specific TFP-mean, for ‘demeaned’ we use TFP estimates in deviation from the cross-section (time-specific) TFP-mean, for ‘both’ we use TFP estimates in deviation from the cross-section and firm-specific TFP mean (we add the sample mean as is practice for a ‘2FE’ transformation).

Table 11: Shocks to TFP — Sample 1

Source of TFP estimates	OLS	OLS†	FE	2FE	LP	ACF
	(1)	(2)	(3)	(4)	(5)	(6)
TFP level year 2000	3.511 (63.09)**	3.530 (63.43)**	5.161 (83.80)**	4.007 (71.35)**	5.009 (80.47)**	3.074 (54.64)**
year 2001	-0.155 (1.99)*	-0.156 (2.00)*	-0.183 (2.12)*	-0.164 (2.08)*	-0.180 (2.07)*	-0.148 (1.88)
year 2002	0.047 (0.59)	0.047 (0.59)	-0.001 (0.02)	0.034 (0.42)	-0.003 (0.04)	0.058 (0.71)
year 2003	0.176 (2.20)*	0.176 (2.20)*	0.131 (1.48)	0.164 (2.03)*	0.125 (1.40)	0.183 (2.27)*
year 2004	0.321 (3.97)**	0.321 (3.97)**	0.309 (3.46)**	0.319 (3.91)**	0.301 (3.33)**	0.320 (3.91)**
year 2005	0.488 (6.01)**	0.489 (6.02)**	0.462 (5.14)**	0.485 (5.91)**	0.443 (4.88)**	0.486 (5.91)**
year 2006	0.622 (7.56)**	0.623 (7.58)**	0.605 (6.64)**	0.622 (7.50)**	0.578 (6.28)**	0.614 (7.38)**
year 2007	0.697 (7.88)**	0.697 (7.88)**	0.687 (7.01)**	0.694 (7.77)**	0.688 (6.94)**	0.700 (7.82)**
R-squared	0.10	0.10	0.08	0.10	0.08	0.09

Notes: We run OLS regressions ($n = 1,742$) using TFP estimates backed out from various ‘first stage’ production function models as the dependent variable: in column (1) OLS, (2) OLS † with year dummies, (3) one-way (firm) fixed effects, (4) two-way fixed effects, (5) LP, (6) ACF. We report absolute t -statistics in parentheses, **, * indicate statistical significance at the 1% and 5% level.

Table 12: ‘Determinants’ of TFP illustration — Sample 1

Source of TFP estimates	OLS	OLS†	FE	2FE	LP	ACF
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Patents only						
patent stock	0.0004 (1.00)	0.0005 (1.01)	0.0160 (4.38)**	0.0016 (2.07)*	0.0131 (4.54)**	0.0001 (0.25)
Constant	64.201 (38.36)**	65.447 (38.28)**	372.006 (26.35)**	106.067 (36.69)**	315.569 (28.32)**	41.934 (38.17)**
R-squared	0.00	0.00	0.01	0.00	0.01	0.00
PANEL B: Add year dummies						
patent stock	0.0004 (1.02)	0.0004 (1.03)	0.0159 (4.44)**	0.0015 (2.13)*	0.0131 (4.61)**	0.0001 (0.24)
Constant	46.754 (10.84)**	47.694 (10.82)**	272.044 (7.33)**	77.614 (10.38)**	230.604 (7.89)**	30.314 (10.70)**
R-squared	0.07	0.07	0.04	0.07	0.05	0.07
PANEL C: Add region dummies						
patent stock	0.0003 (0.68)	0.0003 (0.69)	0.0147 (4.15)**	0.0013 (1.80)	0.0122 (4.37)**	0.0000 (0.11)
Constant	70.525 (12.37)**	71.989 (12.36)**	451.923 (9.10)**	119.297 (12.06)**	395.266 (10.16)**	45.038 (12.02)**
R-squared	0.13	0.13	0.08	0.12	0.10	0.13

Notes: We run OLS regressions ($n = 1,742$) using TFP estimates backed out from various ‘first stage’ production function models as the dependent variable: in column (1) OLS, (2) OLS † with year dummies, (3) one-way (firm) fixed effects, (4) two-way fixed effects, (5) LP, (6) ACF. ‘patent stock’ is calculated as depreciation-free accumulation of patent applications recorded by the European (EPO) and UK patent offices since 1982. In order to save space we do not report the coefficients on the year dummies in Panels B and C or on the regional dummies in Panel C. Dependent and independent variables are in levels, not logarithms. The pattern of results is very similar when we use all variables in logs (+0.001) and a dummy for firms which never recorded any patent applications, and when we use the flow variable of patent applications (or its log plus a ‘no patent applications’ dummy). Due to the small magnitudes of standard errors we report absolute t -statistics in parentheses, **, * indicate statistical significance at the 1% and 5% level.