A production function estimator with proxy variables and firm fixed effects

Andrey Stoyanov[*], Nikolay Zubanov[†], and Yoonseok Lee[‡]

[*]York University, Department of Economics, Faculty of Liberal Arts and Professional Studies, 1084 Vari Hall, 4700 Keele St., M3J 1P3, Toronto, Canada.

[†]Corresponding author. Goethe University Frankfurt, Faculty of Economics and Business Administration, Grueneburgplatz 1, 60323 Frankfurt am Main, Germany, email: zubanov@econ.uni-frankfurt.de, tel. +496979834806.

[‡]Syracuse University, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse, NY 13244, USA.

Abstract

We show, in a very plausible theoretical setting, that control function estimators (CFEs) of firm production function, such as Olley-Pakes, may be biased. The bias will occur, in particular, when investments respond differently to short- and long-lasting changes in productivity. We modify the original CFE approach to allow for this differential response by introducing firm fixed effects to the control function. Applying our modified CFE to the data, we find that it does better than the existing CFEs in terms of controlling for persistent unobserved heterogeneity in productivity. Our findings imply that allowing firm fixed effects in the control function enhances its ability to capture firm productivity, and hence improves statistical quality of productivity estimates.

Keywords: production function; control function estimator; panel data.

1 Introduction

Much of the empirical literature on firm-level productivity relies on estimating the production function. A well-known problem in this literature is what Griliches and Mairesse (1998) call the *transmission bias* on input elasticity estimates, caused by a correlation between factor inputs and unobserved firm productivity. Estimators based on the control function approach, introduced in Olley and Pakes (1996), are a popular solution to this bias. The idea of this approach is to control for the correlation between factor inputs and unobserved firm productivity by proxying the latter with a function of observed firm characteristics that reflect a firm's reaction to productivity changes. Several studies have since emerged, extending the classical Olley-Pakes (OP) estimator to address its various limitations. However, the control function approach as a whole still relies on a number of assumptions which are rarely tested. The critical one is the *scalar unobservability* assumption, that the control function fully absorbs firm productivity. In this paper we show that this assumption fails on two datasets popular with empirical researchers because the control function fails to absorb persistent firm-specific productivity component. We demonstrate that this problem can emerge in a very plausible theoretical setting, and propose that introducing firm fixed effects to the control function solves it.

If the scalar unobservability assumption holds, the production function equation residuals will include only firm-unobserved productivity and an occasional measurement error. However, contrary to this expectation, we find, using the manufacturing firm data from Denmark and Chile, that the control function estimator's (CFE) residuals obtained from these data contain a sizeable time-invariant component. We discuss two alternative explanations for its presence – persistent measurement error and higher-order Markov process in productivity – but find the failure of the scalar unobservability assumption to be the most plausible one.

We identify two theoretical possibilities when the control function will not fully capture the time-invariant productivity component. The first occurs when the econometrician does not observe some of the factors relevant to firms' input decisions, for example, management quality. The second possibility occurs when input choices respond differently to short- and long-lasting productivity changes. In both cases, there will be a productivity component in the residuals that is persistent and correlated with factor inputs, thus leading to biased estimates of the production function. Substantial and persistent productivity differences between firms have long been observed in micro data (Bartelsman and Doms, 2000; Syverson, 2011), and the omission of the firm fixed effects in productivity in the existing CFEs has been recognized as one of its serious limitations (Eberhardt and Helmers, 2010), which may cause the transmission bias due to this omission. The transmission bias in the existing CFEs would apply mostly to the capital input's coefficient, since it is capital that is more highly correlated with the persistent productivity component than is labor.[1]

To address this transmission bias problem in the existing CFEs, we introduce firm fixed effects in the estimation procedure. We show that in the presence of fixed effects in productivity and differential investment reaction to short- and long-term variation in productivity, both phenomena highly relevant empirically, the correctly specified control function should include a firm-specific component. We approximate this component with a firm fixed effect, because estimating the fully specified control function is not feasible on a typical micro panel. Therefore, our modification is easy to implement using standard software packages.

We argue that our modified control function with fixed effects estimator (CFFE) can reduce the transmission bias in the existing CFEs. In support of our argument, applying CFFE to the data, we find much weaker persistency in the residuals compared to the standard CFE. We also observe that the coefficient on capital, presumably most affected by the transmission bias, goes down compared to its standard CFE estimate. However, there are also costs in terms of greater susceptibility of a fixed effects estimator, such as ours, to attenuation bias caused by measurement error. This downward bias is a particular concern with inputs that vary little with time, such as capital input. We compare the biases in the estimators with and without fixed effects and identify the circumstances under which each estimator gives a lower bias than the other. Generally, CFFE will perform better when the capital is less persistent, more precisely measured, and when the persistency in the standard CFE residuals is high.

In the rest of the paper, we outline the existing CFE procedures using the OP estimator as the leading example (section 2.1) and its extensions (section 4), and show how a failure of the scalar unobservability assumption can lead to the transmission

[1] Since labor adjustment costs are relatively low, its response to transient productivity shock will be higher than capital's. On the other hand, both inputs will respond equally to variation in productivity's permanent component. Therefore, for any given share of the permanent component in the total productivity, it will be more correlated with capital than with labor.

bias (section 2.3). Then, in section 2.4, we explain how introducing fixed effects in the existing CFEs can mitigate this bias. We apply CFFE to the Chilean and Danish manufacturing firm data in section 3, showing in particular the large reduction in residual persistency from the standard CFE to CFFE. In section 4 we discuss the role of measurement error in factor inputs in the choice between the standard CFE and CFFE, and a higher-order Markov process in productivity as an alternative explanation to the residual persistency we have observed. Section 5 concludes.

2 Control function-based estimators (CFE) of production function

Consider a Cobb-Douglas production function (in logs):

$$\begin{aligned} y_{it} &= l_{it}\beta_l + k_{it}\beta_k + u_{it} \\ u_{it} &= \omega_{it} + e^1_{it} = a_{it} + b_i + e^1_{it} \end{aligned} \tag{1}$$

where i, t are the firm and year indicators, respectively, y_{it} is output, l_{it} is the vector of *static* inputs, such as labor and materials, which can vary freely at each t, k_{it} is a vector of *dynamic* inputs, such as capital, which is partly determined by its previous stock. The term u_{it} is the empirical equivalent of the Hicks-neutral productivity. It includes a component observable to the firm and thus affecting its input choices, ω_{it}, and the component unobservable to the firm, e^1_{it}, which is typically assumed to be serially uncorrelated productivity. In the empirical specification, e^1_{it} will also include a measurement error in output (state variables are assumed to be measured without an error), which may be serially correlated. The observed productivity term ω_{it}, which from now on we will refer to as the total factor productivity (TFP), can be further decomposed into time invariant productivity effect, b_i, and time-variant component a_{it}.

Estimating equation (1) with OLS would result in inconsistent estimates of β_l and β_k because a firm's choice of input factors depends on observed productivity. For example, a positive TFP shock will stimulate the firm to hire more workers and increase investments in physical capital, while firms with higher levels of TFP, b_i, will persistently employ more of all factors of production. Furthermore, static factor inputs with low adjustment costs, such as labor and material inputs, are more responsive to short-term variation in ω_{it}, resulting in potentially stronger bias in β_l. One approach to deal with the bias in the production function's estimates is by proxying TFP with a function of observables, called *control function*. This section gives an overview of existing control function-based estimators, outlines the key assumptions required for their consistency, and shows how adding firm fixed effects in the control function can relax some of these assumptions.

2.1 An example: the Olley-Pakes (OP) estimator

The Olley and Pakes (1996) estimator deals with the endogeneity problem arising from the correlation between input factors and productivity term ω_{it} by proxying the latter with observables that carry information on ω_{it}. The estimation procedure relies on several assumptions, common to other CFEs we discuss later. While the fully-specified OP contains a correction for endogenous firm entry and exit, we focus on the control function part of the estimation procedure.

Assumption 1 *k_{it} at time t is predetermined.*

Assumption 2 *("Scalar unobservability") The investment function i_{it} is fully determined by the dynamic inputs k_{it}, observed TFP ω_{it}, and, possibly, other observable variables z_{it}.*

Under assumptions 1 and 2, the firm investment level that solves the dynamic profit maximisation problem can be represented as the function of the state variables (k_{it}, z_{it})

and TFP:

$$i_{it} = \phi\left(\omega_{it}, k_{it}, z_{it}\right) \tag{2}$$

Assumption 3 *The investment function i_{it} is monotonic in ω_{it}.*

Assumption 3 allows to specify the control function by inverting the investment function (2) for ω_{it}:

$$\omega_{it} = g_1\left(k_{it}, i_{it}, z_{it}\right) \tag{3}$$

Putting the control function (3) back in the production function equation (1) gives the first-stage OP regression

$$y_{it} = l_{it}\beta_l + k_{it}\beta_k + g_1\left(k_{it}, i_{it}, z_{it}\right) + e^1_{it} \tag{4}$$

Since $g_1(\cdot)$ is unknown, one estimates (4) with a sieve series estimator using a polynomial $\tilde{g}_1(\cdot)$ in (k_{it}, i_{it}, z_{it}) as an approximation for the true $g_1(\cdot)$. Doing so prevents identification of β_k but does allow to obtain the coefficient estimates for static factor inputs, $\widehat{\beta}_l$, and the composite term that captures TFP and input of dynamic factors of production

$$\widehat{\Phi}_{it} = k_{it}\beta_k + g_1\left(k_{it}, i_{it}, z_{it}\right) \tag{5}$$

Estimating β_k requires an additional assumption:

Assumption 4 *The time-varying part a_{it} of the TFP term ω_{it} follows a first-order Markov process:*

$$a_{it} = E\left[a_{it}|a_{it-1}\right] + e^2_{it},$$

where e^2_{it} is an innovation term.

Then, for a given β_l, it follows from (1) that

$$E\left[y_{it} - l_{it}\beta_l|k_{it}, \omega_{it-1}\right] = k_{it}\beta_k + E\left[\omega_{it}|k_{it}, \omega_{it-1}\right] + E\left[e^1_{it} + e^2_{it}|k_{it}, \omega_{it-1}\right]$$

Since k_{it} is pre-determined (Assumption 1), it is independent of innovation in productivity, giving

$$E\left[e_{it}^1|k_{it}, b_i\right] = E\left[e_{it}^2|k_{it}, b_i\right] = 0$$

The regression

$$y_{it} - l_{it}\widehat{\beta}_l = k_{it}\beta_k + E\left[\omega_{it}|\omega_{it-1}\right] + e_{it},$$
$$e_{it} = e_{it}^1 + e_{it}^2$$

is then well-defined and has no endogeneity problem. Letting $E\left[\omega_{it}|\omega_{it-1}\right] = f_2\left(\omega_{it-1}\right)$ and noting that $\omega_{it-1} = \Phi_{it-1} - k_{it-1}\beta_k$ from (5), β_k can be consistently estimated from the second stage regression

$$y_{it} - l_{it}\widehat{\beta}_l = k_{it}\beta_k + f_2\left(\widehat{\Phi}_{it-1} - k_{it-1}\beta_k\right) + e_{it}, \tag{6}$$

with the unknown $f_2\left(\cdot\right)$ again approximated by a polynomial.

2.2 Extensions of the Olley-Pakes estimator

Since OP was introduced, several of its limitations have been identified, motivating other CFEs as its extensions. One such limitation is that in practice many firms report zero investments, which casts doubt on the monotonicity of the investment function (part of Assumption 2). In particular, the presence of capital adjustment costs could violate the monotonicity, making the investment function noninvertable. To address this concern, Levinsohn and Petrin (2003) proposed to include intermediate inputs, such as materials which are always positive, in the control function. Like investments, these inputs are chosen optimally by firms given the state variables, but the adjustment costs of the intermediate inputs are arguably lower, so that the monotonicity assumption is less likely to fail. Levinsohn and Petrin's (2003) extension of OP estimates the value added specification, which is the approach we follow as well.

Another potential problem with OP is that the labor input, which is assumed to be static, may in fact be dynamic thanks, for example, to the presence of hiring costs or other frictions. If l_{it} is dynamic, it becomes a deterministic function of TFP and the state variables (k_{it}, z_{it}) and is therefore collinear with the control function in the first-stage equation (4), leaving β_l unidentified. Of the two solutions to this problem, proposed by Ackerberg et al. (2006) and Wooldridge (2009), we choose the latter as less computationally demanding. Wooldridge (2009) modifies the OP procedure by estimating the structural parameters of the production function using the following moment conditions

$$E\left(e_{it} | k_{it}, \{l_{is}, k_{is}, i_{is}, z_{is}\}_{s \leq t-1}\right) = 0 \tag{7}$$

Thus, all coefficients are estimated in one stage by applying GMM to

$$y_{it} = l_{it}\beta_l + k_{it}\beta_k + f_2\left(g_1\left(k_{it-1}, i_{it-1}, z_{it-1}\right)\right) + e_{it} \tag{8}$$

Clearly, β_l and β_k are both identifiable in (8) provided the moment conditions (7) are satisfied. Another advantage of this procedure is that, since all the coefficients are estimated in one step, standard errors can be easily obtained using the standard formula and bootstrapping is not required.

In our estimation procedure, we combine the choice of the instruments for the control function in the LP estimator with the flexibility of the Wooldridge (2009) GMM estimator by running

$$y_{it} = l_{it}\beta_l + k_{it}\beta_k + f_3\left(k_{it-1}, m_{it-1}, z_{it-1}\right) + e_{it} \tag{9}$$

with the moment conditions

$$E\left(e_{it}|k_{it}, l_{it-1}, k_{it-1}, m_{it-1}, z_{it-1}\right) = 0, \tag{10}$$

where v is log value added and m is log materials input. (Notice that, unlike Wooldridge (2009), we use a sieve rather than fixed polynomial approximation in (9).) Yet another advantage of this estimator – most important to our work – is that it can easily accommodate firm fixed effects, which may improve CFE results as we argue in the next section.

2.3 Failure of the scalar unobservability (Assumption 2) and its consequences

The consistency of CFEs depends critically on how closely the firm-observed TFP $\omega_{it} = g_1(\cdot)$ is approximated by its empirical proxy $\tilde{g}_1(\cdot)$. The closeness of $g_1(\cdot)$ and $\tilde{g}_1(\cdot)$ is important because the approximation error

$$r_{it} = g_1(\cdot) - \tilde{g}_1(\cdot),$$

containing polynomials in $(k_{it-1}, m_{it-1}, z_{it-1})$, will be part of the regression error term e_{it}, threatening the validity of the moment conditions (10). If Assumption 2 is satisfied, the approximation error can be made arbitrarily small by increasing the degree of the polynomial in $\tilde{g}_1(\cdot)$.[2] Indeed, we have tried both third- and fourth-degree polynomial approximations of $g_1(\cdot)$ and found that the estimation results are virtually identical (see also Stoyanov and Zubanov, 2012). If Assumption 2 fails, the approximation error can no longer be ignored. The failure of Assumption 2 can happen in two cases: 1) omitted

[2] Another assumption whose failure can set the approximation error r_{it} out of control is Assumption 4, that TFP follows a first-order Markov. We discuss the consequences of this assumption's failure in extensions.

observables in z, and 2) different investments' responses to permanent (b_i) and transient (a_{it}) changes in TFP. In Case 1, when only a subset z' of the relevant observables (z', z'') is included in the control function, r will contain polynomials in $(k_{it-1}, m_{it-1}, z''_{it-1})$, which will not disappear as the order of polynomial approximation increases. Their presence will invalidate the moment conditions (10) and lead to inconsistent estimates.

In Case 2, even when all the relevant observables are included, differences in investments' (or materials') response to different TFP components a_{it} and b_i will also result in a non-reducible approximation error correlated with factor inputs. By specifying the investment function as in (2), the existing CFEs assume that investments depend on the sum $a_{it} + b_i$ and hence respond equally to a given difference in a_{it} or b_i. However, this assumption may not be true. For instance, factor adjustment costs will weaken the reaction of a given firm to a transitory productivity shock a_{it} (Cooper and Haltiwanger, 2006), but will have no effect on the same firm's scale of operations chosen in response to b_i. Hence, there will be a stronger response of investments to a variation in b_i, observed between firms, than to the same variation in a_{it} over time in a given firm. By restricting this response to be equal for b_i and a_{it}, one relegates part of the response to b_i to the approximation error, which will result in a time-persistent component $r(b_i)$ in the error term. Because the presence of $r(b_i)$ violates the moment conditions, the bias it causes to the regression estimates cannot be dealt with by any of the control function-based estimators surveyed above. The solution we propose is to control for it by adding fixed effects in the control function.

2.4 The CFE with firm fixed effects (CFFE)

Our solution is best illustrated with an example of the investment function that includes all the relevant observables z and a linear combination of b_i and a_{it} among its arguments:

$$i_{it} = \phi\left([a_{it} + \gamma b_i], k_{it}, z_{it}\right) = \phi\left([\omega_{it} + (\gamma - 1)b_i], k_{it}, z_{it}\right), \tag{11}$$

where $\gamma > 1$ is set to allow for investment's response to b_i to be stronger than to a_{it} (Case 2 in the previous section), in violation of Assumption 2. Inverting (11) for $[\omega_{it} + (\gamma - 1)b_i]$ gives the control function

$$\omega_{it} = g_1\left(k_{it}, i_{it}, z_{it}\right) + (1 - \gamma)b_i, \tag{12}$$

which one can approximate with a $\tilde{g}_1(\cdot)$ and added firm fixed effects, and proceed to the second stage as in Section 2.1 or apply a GMM estimator to

$$v_{it} = l_{it}\beta_l + k_{it}\beta_k + f_2\left(g_1\left(k_{it-1}, m_{it-1}, z_{it-1}\right)\right) + \xi_i + e_{it} \tag{13}$$

as in Section 2.2. The firm fixed effects ξ_i will absorb the time-invariant component $r(b_i) = (1 - \gamma)b_i$ in the approximation error. Hence, if other assumptions (1, 3 and 4) are satisfied, the CFE with added fixed effects will produce consistent estimates.

A more general investment function specification,

$$i_{it} = \phi\left(a_{it}, b_i, k_{it}, z_{it}\right),$$

gives upon inversion for a_{it} the control function for the time-varying TFP component

$$a_{it} = g_1\left(k_{it}, i_{it}, z_{it}, b_i\right),$$

implying

$$\omega_{it} = g_1\left(k_{it}, i_{it}, z_{it}, b_i\right) + b_i, \tag{14}$$

where the fixed effect b_i is no longer additive. A situation giving rise to the control function in (14) occurs, for example, when some time-invariant components of z, such as firm location or management quality, have been omitted (Case 1 in the previous section), which often happens in micro data.

Approximating ω_{it} in (14) with a third-degree polynomial in $g_1(\cdot)$'s arguments, we obtain (omitting z for brevity)

$$\begin{aligned}\omega_{it} &\approx \textstyle\sum_{j_1=0}^{3}\sum_{j_2=0}^{3-j_1}\pi_{j_1j_2}(k_{it})^{j_1}i_{it}^{j_2} + \sum_{j_1=0}^{3}\sum_{j_2=0}^{3-j_1}\pi_{i;j_1j_2}(k_{it})^{j_1}i_{it}^{j_2} \\ &= G_1(k_{it}, i_{it}) + G_{2i}(k_{it}, i_{it}),\end{aligned} \tag{15}$$

where $\pi_{j_1j_2}$ is the coefficient on the corresponding power combination, and $\pi_{i;...}$ is the same, specific to each firm. Firm-specific terms in (15), summarized in $G_{2i}(\cdot)$ above, appear because of the presence of the non-additive b_i in the control function (14).

To reduce the burden on the data, we use a restricted version of the control function where $G_{2i}(\cdot)$ is approximated with the firm fixed effect,

$$G_1(\cdot) + G_{2i}(\cdot) \approx \sum_{j_1=0}^{3}\sum_{j_2=0}^{3-j_1}\pi'_{j_1j_2}(k_{it})^{j_1}i_{it}^{j_2} + \mu_i, \tag{16}$$

where μ_i contains b_i and the firm-averages of the interactions in (15). We prefer the fixed-effects approximation in (16) over that in (15), which uses firm-specific polynomial coefficients, for two reasons. First, consistent estimation of the unrestricted G_{2i} in (15) requires a long time dimension, which is typically unavailable in micro data. Second, using (16) makes a minimal modification to the existing CFEs. The consequence of using (16) instead of (15) is, of course, a less precise approximation of the firm-observed

TFP, ω_{it}. However, critically for our purpose, (16) is just as good as (15) in capturing the time-persistent TFP components.

Having specified the control function using (15) and (16), one obtains the analogue of the OP second-stage equation (6) with fixed effects:

$$y_{it} - l_{it}\widehat{\beta}_l = k_{it}\beta_k + f_2\left(G_1(k_{it-1}, i_{it-1}) + \mu_i\right) + e_{it}, \tag{17}$$

Here, too, $f_2(\cdot)$ approximated with a polynomial in $G_1(\cdot)$ and μ_i can be written as the sum of the same two components as in (15) in lags of their arguments. Replacing the second component with a fixed effect, as we did in (16), we obtain

$$y_{it} = l_{it}\widehat{\beta}_l + k_{it}\beta_k + f_2\left(G_1\left(k_{it-1}, i_{it-1}, z_{it-1}\right)\right) + \text{fixed effect}_i + e_{it}, \tag{18}$$

which can be estimated with GMM using the moment conditions

$$E\left(e_{it}|k_{it}, l_{it-1}, k_{it-1}, i_{it-1}, z_{it-1}, \text{fixed effect}_i\right) = 0$$

or

$$E\left(e_{it}|k_{it}, l_{it-1}, k_{it-1}, m_{it-1}, z_{it-1}, \text{fixed effect}_i\right) = 0$$

for the value added specification.

Importantly for empirical researchers, the practical implementation of CFFE is fairly uncomplicated. Approximating $f_2(\cdot)$ with a polynomial makes equation (18) linear, allowing it to be estimated with GMM using the Stata command `xtivreg2`.

While running CFFE helps remove the upward bias in β's due to correlation between factor inputs and b_i, it may also exacerbate the downward bias due to the presence of measurement error (Eberhardt and Helmers, 2010). This attenuation bias will be particularly strong when the measurement error is serially uncorrelated. Therefore, applying CFFE involves a tradeoff between a reduction in the positive bias and an increase in the attenuation bias. One should be cautious of it in empirical analysis and use our estimator only when the standard CFE estimation results suggest the presence of a time-persistent TFP component (b_i). In section 4.1 we illustrate this tradeoff by comparing the biases in the estimates with and without fixed effects in the presence of persistency in TFP as well as measurement error in factor inputs.

3 Empirical results

In this section we use two popular datasets in the firm productivity literature to show that the control function in the standard CFE fails to account for persistency in TFP, leaving it to the error term. We then demonstrate that the firm fixed effects in our modification of CFE absorb this persistency, preventing a correlation between factor inputs and the error term, and the resulting bias.

3.1 Data

The first dataset comes from *Instituto Nacional de Estadistica* and covers all Chilean manufacturing plants with more than 10 employees during the years between 1979-1996. These data have been used in many studies of firm-level productivity, including Levinsohn and Petrin (2003), Ackerberg et al. (2006), and Gandhi et al. (2011), as well as in applications of productivity analysis in other contexts, most notably Pavcnik (2002), Petrin and Levinsohn (2013), Kasahara and Rodrigue (2008). For each of the 10,927 plants in our sample, the data include four-digit ISIC industry code identifier, gross output, material inputs, capital stock and investments, and labor input measured in person-years, all converted into real values using industry price deflators. A more detailed description of the data is available in Levinsohn and Petrin (2003).

The second dataset comes from the Danish Business Statistics Register, maintained by Statistics Denmark, and includes all firms registered in Denmark. In this study we use information on manufacturing firms for the years 1995-2007. The variables observed in each year include four-digit NACE industry identifier, total output, value added, fixed assets, investments, material inputs, and employment (headcount measure). Munch and Skaksen (2008), Loecker and Warzynski (2012), and Lentz and Mortensen (2008) have used these data.

3.2 CFE results without and with firm fixed effects

Tables 1 and 2 present the OLS and standard CFE estimation results for the Cobb-Douglas production function (1) for Chile and Denmark, respectively, as well as test results for the presence of a persistent TFP component not captured by the control function. We perform this test by looking into first-order autocorrelation in regression residuals. Theory informs us that the persistent TFP component not captured by the control function, if present, will be part of the first-stage residuals comprised of e_{it}^1 and the approximation error $r(b_i)$ which cannot be identified individually. We therefore separate the first- and second-stage components, $(\widehat{e_{it}^1 + r_{it}})$ and $\widehat{e_{it}^2}$, from the combined error term $\widehat{e}_{it}$. To recover $\widehat{e_{it}^2}$, we first construct the TFP term from the first stage $\widehat{\omega}_{it} = \widehat{\Phi}_{it} - k_{it}\widehat{\beta}_k$ and collect $\widehat{e_{it}^2}$ as the residual from the non-parametric regression of $\widehat{\omega}_{it}$ on $\widehat{\omega}_{it-1}$.

TABLE 1 HERE.

TABLE 2 HERE.

Since the patterns observed in two datasets are remarkably similar, we will focus our exposition on the results for Chile. The middle panel of Table 1 reports the autoregression coefficients $\widehat{\rho}$, $\widehat{\rho}_1$, and $\widehat{\rho}_2$ for the total error term, $\widehat{e}_{it}$, and its components $(\widehat{e_{it}^1 + r_{it}})$ and $\widehat{e_{it}^2}$, respectively. These coefficients indicate strong autocorrelation (0.5 or higher) in the overall error term, with the exception of the OLS with fixed effects residuals (0.26). The residual autocorrelation is mostly driven by the autocorrelation in the first-stage residuals (coefficient ρ_1). Furthermore, the data suggest that the high persistency in the first-stage residuals is due to the presence of a time-invariant component in them. Indeed, roughly half of their variation comes from variation between firms (the bottom lines in Tables 1 and 2), and, as seen on the correlogram in Figure 1, their autocorrelation hardly declines with time. Very similar patterns are observed in the first-stage residuals on the Danish data, and when the production function is estimated separately for three largest manufacturing industries (column 5 to 7). In sum, the high persistency in the CFE residuals suggests that CFE without fixed effects does not fully capture TFP.

FIGURE 1 HERE.

We now apply CFFE to our data. Table 3 shows the estimated factor input elasticities and, in square brackets, their differences with the respective estimates from the Wooldridge (2009) CFE (WOP, Tables 1 and 2, columns 4 to 7), as well as the residual autocorrelation tests. There is a considerable reduction in the magnitude of residual autocorrelation coefficients as compared to the standard CFE estimates: a maximum of 0.235 (Chile) and 0.065 (Denmark) with fixed effects versus an upwards of 0.5 without. Recovering the first- and second-stage residuals, we find that the reduction in their autocorrelation happens in the the first stage (coefficient ρ_1), whereas the second-stage autocorrelations (coefficient ρ_2), already small in the standard CFE case, are affected less.The facts that 1) it was the first-stage residuals that were most strongly autocorrelated, and that 2) this autocorrelation was greatly reduced by incorporating fixed effects in the control function imply together that CFFE can capture a large part of the persistent TFP component that was previously left in the error term.

TABLE 3 HERE.

Turning to the factor input elasticity estimates in Table 3, one observes that, while the labor elasticity estimates are broadly similar between the two estimators, the CFFE capital elasticity estimates are always lower, especially on the Chile sample. We believe the reason has to do with differential response of capital and labor to the two TFP components, a_{it} and b_i. Given that labor input is more responsive to changes the time-varying a_{it} than is capital, there will be an upward bias in the labor's estimated coefficient and a downward bias in capital's when one controls for b_i but does not control for a_{it}, as in the OLS with fixed effects. Analogously, when capital is more strongly correlated with the time-invariant b_i than is labor, there may be an upward bias in capital's and a downward bias in labor's coefficient when one controls for a_{it} but not (fully) for b_i, as in the standard CFE. The elimination of this bias by CFFE will then bring capital's coefficient down more than labor's compared to their standard CFE estimates. However, in addition to the correction for the upward transmission bias, the reduction in the capital's coefficient may also be due to the amplification of the downward attenuation bias, a concern we discuss in the next section.

4 Extensions

Is our modification always an improvement over the existing CFEs? In theory, any description of the data generating process that is more complete than the previous one will produce better quality statistical estimates. In practice, however, the benefits of a more complete data description must be set against the costs of imposing additional estimation burden on the data. This section discusses two issues with CFFE. The first is the measurement error in factor inputs, which may be exacerbated when adding fixed effects. The second concerns alternative explanations to the persistency in the CFE error term, for example, failure of the first-order Markov in TFP (Assumption 4), which may or may not be sufficiently addressed by adding fixed effects.

4.1 Measurement error in factor inputs

Consider a simplified version of the production function regression (1)

$$y = x^{*}\beta + \gamma + \varepsilon, \tag{19}$$

where x^{*} is the only factor input and γ is the time-invariant component of the error term $\gamma + \varepsilon$. Instead of x^{*} one observes $x = x^{*} + \eta$, where η is the measurement error, independent and identically distributed for all firms and years. We abstract from the instrumentation of x that the Wooldridge (2009) CFE involves[3], and partial out regressors other than x through orthogonal projection.

Applying OLS without fixed effects to (19) will result in a bias on the estimate $\hat{\beta}_{\text{no FE}}$ as follows:

$$B_0 = E(\hat{\beta}_{\text{no FE}}) - \beta = \frac{\sigma_{x^{*}\gamma}}{\sigma^2_{x^{*}} + \sigma^2_{\eta}} - \beta\frac{\sigma^2_{\eta}}{\sigma^2_{x^{*}} + \sigma^2_{\eta}}, \tag{20}$$

where $\sigma_{x^{*}\gamma}$ is the covariance between x^{*} and γ and σ^2_{η} and $\sigma^2_{x^{*}}$ are the variances of the measurement error and x^{*}, respectively. The first term in (20) is the transmission bias, due to the correlation between x and the fixed effect γ, and the second term is the attenuation bias due to measurement error η. Running (19) with fixed effects will remove the transmission bias but will exacerbate the attenuation bias, producing

$$B_1 = E(\hat{\beta}_{\text{FE}}) - \beta = -\beta\frac{T-1}{T}\frac{\sigma^2_{\eta}}{\sigma^2_{x^{*}} + \frac{T-1}{T}\sigma^2_{\eta} - \sigma^2_{\bar{x}^{*}}} \tag{21}$$

Notice that the attenuation bias with fixed effects is indeed larger than without, since the denominator in (21) is lowered by the between-firm variance in x^{*}, $\sigma^2_{\bar{x}^{*}}$.

[3]Working with a fully-fledged CFE will introduce an extra parameter – the correlation between the endogenous variable and the instruments – which is not important for the purposes of this section.

To decide which estimator to use, one could compare the absolute values of the OLS and FE biases, $|B_0|$ and $|B_1|$. The FE bias B_1 is always negative, whereas the OLS bias, B_0 can be both negative and positive. When $B_0 < 0$, OLS should always be preferred since then B_0 will always be less than B_1 in magnitude. When $B_0 > 0$, OLS without fixed effects will produce a greater bias if

$$\sigma_{x^*\gamma} > \beta \left(\frac{T-1}{T} \frac{\sigma_x^2}{\sigma_x^2 - \sigma_{\bar{x}}^2} + 1 \right) \sigma_\eta^2, \tag{22}$$

where σ_x^2 and $\sigma_{\bar{x}}^2$ are the overall and between-firm variances of the observed x, respectively. That is, adding fixed effects will produce an estimate closer to the true β when the inequality in (22) holds.

Intuitively, all else equal, $\beta_{\text{no FE}}$ will be less biased than β_{FE} when the true β is large, since β increases the magnitude of the attenuation bias, which is larger in FE (equation (21)). $\beta_{\text{no FE}}$ tends to be less biased than β_{FE} also when x is persistent over time, since then the share of the within-firm variance in the total will be smaller, increasing the attenuation bias in β_{FE}. Because larger T only increases the FE attenuation bias, OLS should be preferred to FE on long panels. Finally, while no measurement error ($\sigma_\eta^2 = 0$) makes fixed effects the preferred option, in a more realistic case of $\sigma_\eta^2 > 0$ OLS will be preferred when the measurement error is large enough.[4]

The above intuitions can be brought to practical use in choosing whether to use FE or not. Two of the five parameters in (22), T and $\frac{\sigma_x^2}{\sigma_x^2 - \sigma_{\bar{x}}^2}$, are directly observable; the rest may be inferred from the data. The magnitude of $\sigma_{x^*\gamma}$ is related to the share of

[4]These intuitions receive some empirical support as we compare the standard CFE and CFFE estimates for capital input obtained on subsamples of our data. These subsamples, drawn from different industry groups, vary in the ratio of the total variance in observed capital to its within-firm component, $\frac{\sigma_x^2}{\sigma_x^2 - \sigma_{\bar{x}}^2}$, and the CFE-estimated capital input elasticity, β. We find (results available on request) that the difference between standard CFE and CFFE capital elasticity estimates increases with both $\frac{\sigma_x^2}{\sigma_x^2 - \sigma_{\bar{x}}^2}$ and β, confirming our theoretical results for the biases with and without fixed effects (equations (20) and (21)), as well as their relative strength (equation (22)).

time-invariant component in the total residual variance (the bottom rows in Tables 1 and 2). The standard CFE estimate of β can be taken as its upper bound. Though including controls, such as capacity utilization, may reduce σ_η^2, it is likely to remain unknown in most empirical applications and will have to be guessed.

4.2 Other sources of time persistency in CFE residuals

There may be sources of persistency in CFE residuals other than the fixed TFP component not captured by the control function, the focus of our paper. One is persistent measurement error in inputs or/and output, which affects the first-stage residuals e_{it}^1. Persistency in the output measurement error is inconsequential, but persistent measurement error in factor inputs will lead to attenuation bias in input elasticities' estimates, just as a common measurement error will. The existing CFEs cannot address the measurement error problem, and their GMM-based versions will require longer lags of instruments in the moment conditions (10). In the extreme case of a fixed effect in the factor input measurement error, GMM will not improve over the standard CFE. Adding fixed effects will absorb the measurement error and may therefore be an improvement in this circumstance.

Another potential source of persistency in CFE residuals is the failure of Assumption 4, that TFP follows a first-order Markov process. A higher-order Markov in TFP clearly invalidates the second-stage CFE results, since their consistency explicitly relies on Assumption 4. We have seen in Tables 1 and 2 that the second-stage residuals were indeed autocorrelated, though their autocorrelation was weaker than in e_{it}^1, and less of their variance was firm-specific. Higher-order Markov process in TFP will contaminate the first-stage CFE results as well, by introducing lags of TFP in the production function (1) which will not be captured by the control function (3). Since the failure of Assumption 4 has consequences for both stages of CFE, it may contribute to the observed persistency of the first- as well as the second-stage residuals.

The presence of a higher-order Markov in TFP can be addressed by extending the existing CFEs to allow for a second-order Markov processes in TFP,

$$\omega_{it} = \mathrm{E}\left[\omega_{it}|\omega_{it-1}, \omega_{it-2}\right] + \xi_{it} = \lambda\left(\omega_{it-1}, \omega_{it-2}\right) + \xi_{it},$$

at the cost of extra assumptions on the control function and at least one additional proxy variable, besides materials, to separately identify the two lags of TFP (Ackerberg et al., 2007; Stoyanov and Zubanov, 2014). We implement the estimator proposed in these studies by choosing capital investment as the additional proxy. While materials, which are used up within one period, respond primarily to short-lasting TFP shocks, investment captures longer shocks and is thus better suited to proxy longer TFP lags.

Table 4 reports the regression estimates and residual autocorrelations for the CFE specification with the TFP following a second-order Markov process but without firm fixed effects. This modification produces second-stage residuals e_{it}^2 that are less strongly correlated than those from the standard CFE. However, the first-stage residual autocorrelation is essentially unaffected in both Chile and Denmark datasets, implying that the failure of Assumption 4 is unlikely to be an important factor contributing to the persistency in e_{it}^1. Therefore, our conjecture that the existing CFEs do not fully capture fixed effects in TFP remains the preferred explanation for the observed persistency in CFE residuals.

TABLE 4 HERE.

5 Conclusion

In this paper, we identify a potential transmission bias in the production function estimators based on the control function approach. This bias occurs because the control function does not fully capture a persistent TFP component, which will happen when not all relevant controls are included or when investments (or materials input) respond differently to short- and long-lasting changes in TFP. We show that allowing for differential response to different TFP components, and hence eliminating the transmission bias in CFE, boils down to introducing firm fixed effects in the control function. Importantly to applied researchers, this modification of the existing CFEs is easy to implement.

We demonstrate the empirical importance of our modification by reporting a substantial firm-specific component in the CFE residuals obtained on the Danish and Chilean manufacturing data. The presence of this component is a marker of the transmission bias, since a correctly specified control function would absorb all relevant firm heterogeneity. We then show that applying CFFE greatly reduces persistency in the residuals. This advantage notwithstanding, CFFE may also exacerbate the attenuation bias on the coefficients on factor inputs measured with error. We discuss the factors affecting the strength of this bias relative to the transmission bias, and describe situations in which CFFE should be preferred to the existing CFEs.

References

Ackerberg, D., K. Caves, and G. Frazer (2006): "Structural identification of production functions," MPRA Paper 38349, University Library of Munich, Germany.

Ackerberg, D., C. Lanier Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 63.

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

Cooper, R. and J. Haltiwanger (2006): "On the Nature of Capital Adjustment Costs," *Review of Economic Studies*, 73, 611–633.

Eberhardt, M. and C. Helmers (2010): "Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators," Economics Series Working Papers 513, University of Oxford, Department of Economics.

Gandhi, A., S. Navarro, and D. Rivers (2011): "On the Identification of Production Functions: How Heterogeneous is Productivity?" University of Western Ontario, CIBC Centre for Human Capital and Productivity Working Papers 20119, University of Western Ontario, CIBC Centre for Human Capital and Productivity.

Griliches, Z. and J. Mairesse (1998): "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, New York: Cambridge University Press, 169–203.

Kasahara, H. and J. Rodrigue (2008): "Does the Use of Imported Intermediates

Increase Productivity? Plant-Level Evidence," *Journal of Development Economics*, 87, 106–118.

Lentz, R. and D. T. Mortensen (2008): "An Empirical Model of Growth Through Product Innovation," *Econometrica*, 76, 1317–1373.

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341.

Loecker, J. D. and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–71.

Munch, J. R. and J. R. Skaksen (2008): "Human capital and wages in exporting firms," *Journal of International Economics*, 75, 363–372.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–97.

Pavcnik, N. (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 69, 245–276.

Petrin, A. and J. Levinsohn (2013): "Measuring Aggregate Productivity Growth Using Plant-Level Data," *RAND Journal of Economics*, 43, 705–725.

Stoyanov, A. and N. Zubanov (2012): "Productivity Spillovers across Firms through Worker Mobility," *American Economic Journal: Applied Economics*, 4, 168–198.

——— (2014): "The distribution of the gains from spillovers through worker mobility between workers and firms," *European Economic Review*, 70, 17–35.

Syverson, C. (2011): "What determines productivity?" *Journal of Economic Literature*, 49, 326–365.

Wooldridge, J. M. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104, 112–114.

Table 1. Production function estimation results from existing estimators, Chile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimator	OLS	OLS-FE	OP	WOP	WOP	WOP	WOP
Industry sample	All manufacturing industries				Food products (311)	Fabricated metals (381)	Textiles (321)
Labor	1.023 (0.018)	0.813 (0.016)	0.689 (0.017)	0.758 (0.017)	0.582 (0.027)	0.795 (0.045)	0.623 (0.047)
Capital	0.308 (0.010)	0.128 (0.009)	0.530 (0.014)	0.349 (0.018)	0.321 (0.035)	0.221 (0.056)	0.259 (0.048)
N	54,801	54,801	49,228	49,265	14,823	4,171	4,189
ρ	0.696 (0.003)	0.264 (0.007)	0.686 (0.009)	0.647 (0.007)	0.491 (0.014)	0.569 (0.025)	0.576 (0.024)
ρ_1			0.625 (0.006)	0.637 (0.007)	0.496 (0.014)	0.535 (0.024)	0.540 (0.023)
ρ_2			-0.125 (0.008)	-0.030 (0.009)	-0.188 (0.017)	-0.223 (0.025)	-0.148 (0.025)
	The share of between-firm variation in residuals						
e^1	0.57	n/a	0.51	0.53	0.37	0.47	0.46
e^2			0.09	0.09	0.08	0.11	0.11

Note: Standard errors in parentheses are clustered by firm. WOP stands for the Wooldridge (2009) modification of the OP estimator. For OP and WOP estimators, the share of between-firm variation in residuals is calculated for the first-stage regression

Table 2. Production function estimation results from existing estimators, Denmark

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimator	OLS	OLS-FE	OP	WOP	WOP	WOP	WOP
Industry sample	All manufacturing industries				Fabricated metals	Printing	Food products
					(28)	(22)	(15)
Labor	0.917	0.723	0.631	0.658	0.724	0.686	0.655
	(0.004)	(0.003)	(0.006)	(0.005)	(0.010)	(0.014)	(0.012)
Capital	0.139	0.108	0.203	0.102	0.094	0.079	0.105
	(0.003)	(0.003)	(0.002)	(0.003)	(0.005)	(0.007)	(0.008)
N	151,556	151,556	122,938	122,902	24,235	15,061	14,890
ρ	0.777	0.136	0.648	0.554	0.528	0.554	0.535
	(0.002)	(0.008)	(0.007)	(0.008)	(0.015)	(0.019)	(0.031)
ρ_1			0.559	0.582	0.550	0.554	0.541
			(0.006)	(0.007)	(0.015)	(0.015)	(0.022)
ρ_2			-0.217	-0.266	-0.254	-0.248	-0.224
			(0.007)	(0.007)	(0.016)	(0.024)	(0.023)
	The share of between-firm variation in residuals						
e^1	0.75	n/a	0.56	0.56	0.51	0.55	0.52
e^2			0.19	0.16	0.18	0.16	0.20

Note: Standard errors in parentheses are clustered by firm. WOP stands for the Wooldridge (2009) modification of the OP estimator. For OP and WOP estimators, the share of between-firm variation in residuals is calculated for the first-stage regression

Table 3. Production function estimation results from OP with fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Chile				Denmark			
	All industries	Food products	Fabricated metals	Textiles	All industries	Fabricated metals	Printing	Food products
		(311)	(381)	(321)		(28)	(22)	(15)
Labor	0.669	0.606	0.736	0.679	0.780	0.794	0.817	0.730
	(0.025)	(0.053)	(0.065)	(0.068)	(0.011)	(0.022)	(0.025)	(0.032)
	[-0.089]	[0.024]	[-0.059]	[0.056]	[0.122]	[0.070]	[0.131]	[0.075]
Capital	0.164	0.173	0.107	0.142	0.081	0.066	0.074	0.091
	(0.012)	(0.028)	(0.044)	(0.038)	(0.003)	(0.005)	(0.007)	(0.010)
	[-0.185]	[-0.148]	[-0.114]	[-0.117]	[-0.021]	[-0.028]	[-0.005]	[-0.014]
N	49,043	13,272	3,811	4,000	119,448	23,094	14,580	14,397
ρ	0.235	0.219	0.177	0.171	0.058	0.065	0.034	0.054
	(0.008)	(0.014)	(0.028)	(0.025)	(0.009)	(0.018)	(0.020)	(0.022)
ρ_1	0.219	0.215	0.160	0.156	0.055	0.058	0.035	0.035
	(0.008)	(0.014)	(0.027)	(0.025)	(0.009)	(0.018)	(0.022)	(0.022)
ρ_2	-0.003	-0.007	-0.020	-0.028	-0.174	-0.132	-0.205	-0.204
	(0.009)	(0.013)	(0.034)	(0.033)	(0.008)	(0.015)	(0.016)	(0.021)

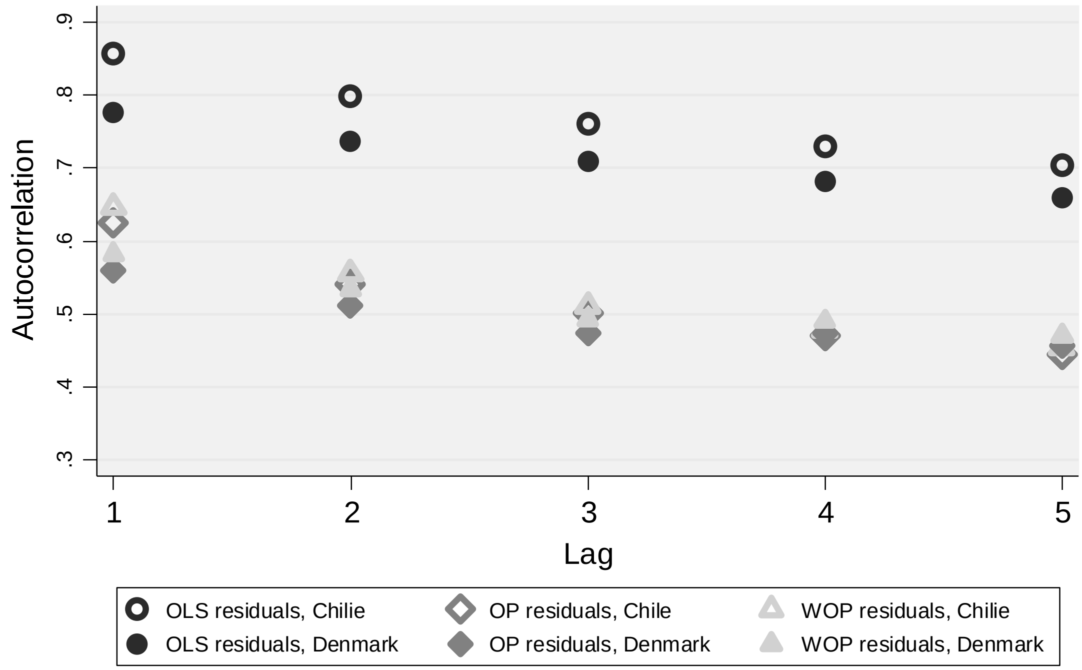
Note: Standard errors in parentheses are clustered by firm. In square brackets are the differences in factor input elasticity estimates between the OP with and without fixed effects. For OP and WOP estimators, the share of between-firm variation in residuals is calculated for the first-stage regression

Table 4. Estimation results from the classical OP with a second-order Markov process in TFP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Chile				Denmark		
	All	Food products	Fabricated metals	Textiles	All	Fabricated metals	Printing	Food products
		(311)	(381)	(321)		(28)	(22)	(15)
Labor	0.504	0.425	0.594	0.527	0.683	0.738	0.714	0.656
	(0.010)	(0.020)	(0.029)	(0.031)	(0.006)	(0.012)	(0.017)	(0.013)
	[-0.254]	[-0.157]	[-0.201]	[-0.096]	[0.025]	[0.014]	[0.028]	[0.001]
Capital	0.366	0.314	0.442	0.231	0.080	0.069	0.066	0.093
	(0.022)	(0.045)	(0.067)	(0.058)	(0.003)	(0.006)	(0.007)	(0.008)
	[0.017]	[-0.007]	[0.211]	[-0.028]	[-0.022]	[-0.025]	[-0.013]	[-0.012]
N	16,081	4,127	1,503	1,261	83,948	16,604	10,080	9,857
ρ	0.730	0.553	0.356	0.588	0.577	0.554	0.579	0.534
	(0.011)	(0.025)	(0.058)	(0.040)	(0.008)	(0.016)	(0.020)	(0.025)
ρ_1	0.707	0.519	0.616	0.589	0.575	0.553	0.577	0.558
	(0.012)	(0.028)	(0.033)	(0.041)	(0.008)	(0.017)	(0.018)	(0.024)
ρ_2	0.129	0.029	-0.044	-0.168	0.001	-0.145	0.010	-0.163
	(0.014)	(0.032)	(0.041)	(0.047)	(0.010)	(0.016)	(0.022)	(0.032)
	The share of between-firm variation in residuals							
e^1	0.63	0.47	0.56	0.55	0.55	0.51	0.53	0.54
e^2	0.21	0.22	0.18	0.23	0.18	0.18	0.20	0.16

Note: Standard errors in parentheses are clustered by firm. In square brackets are the differences in factor input elasticity estimates between the OP with and without fixed effects. For OP and WOP estimators, the share of between-firm variation in residuals is calculated for the first-stage regression

Figure 1. Autocorrelation correlogram for the regression residuals.



Note: total residuals are used for OLS; first-stage residuals for OP and WOP