Productivity and Capital Measurement Error*

Allan Collard-Wexler
Duke University
NBER

Jan De Loecker
KU Leuven
NBER and CEPR

April 11, 2022

Abstract

Any productivity analysis relies on correctly allocating the contributions of capital, and other factors of production, to output. The presence of measurement error in constructing the capital stock is well-known, and poses an enormous challenge for any empirical analysis of productivity. It may lead to under-estimating the role of capital, and therefore capital deepening in explaining output differences across firms, industry dynamics and ultimately economic growth. We introduce an identification scheme, and an accompanying estimation procedure, that jointly deals with measurement error in capital and the standard simultaneity bias due to unobserved productivity shocks. Our strategy is based on the insight that once persistent productivity differences across producers are controlled for, a set of valid instruments become available that in standard IV approaches would fail. Within our framework we propose two instruments: the replacement value of capital and lagged investment. We validate our approach through Monte Carlo experiments, and trace out the relationship between the bias in the capital stock and the size of the measurement error. We apply our approach to three datasets covering three distinct economies: China, India and Chile, as well as cross-country data from the World Bank. We estimate capital coefficients that are typically two times larger than those using standard approaches that only control for simultaneity, suggesting a sizable measurement error in capital. We underscore the importance of obtaining reliable capital coefficients, and document the bias in the productivity premium of technology, size, and ownership in the context of our applications. Throughout, we document a smaller role for unobserved productivity differences, suggesting a more prominent role for capital in explaining output differences.

Key words: *Productivity; productivity premia; capital stock; measurement error; simultaneity; IV*

*This paper was initially circulated under the title *Production function estimation with measurement error in inputs*. Contact information: collardwexler@gmail.com and jan.deloecker@kuleuven.be. De Loecker gratefully acknowledges support from the FWO Odysseus Grant (G0F7216N). We thank Steve Berry, Penny Goldberg, John Haltiwanger, Matt Masten, Shakeeb Khan, Ariel Pakes, John Sutton, Mark Schankerman, Jo Van Biesebroeck, Chad Syverson, and Daniel Xu, and participants at NBER SI, Carnegie-Mellon, Chicago, Zurich, LSE, Imperial College, and the Cornell-Penn State Econometrics workshop for helpful conversations. Ana Paula Cusolito, Jo Van Biesebroeck, and Daniel Ackerberg generously shared their code. Sherry Wu and Yiling Jiang provided excellent research assistance.

1 Introduction

Production functions are a central component in a variety of economic analyses. These often first need to be estimated using data on individual production units, and there is a vast literature on identification in the presence of unobserved productivity shocks, but more limited work on how to deal with the challenges of measurement of output and inputs.[1] There is reason to believe that, more than any other input in the production process, there are severe errors in the recording of a producer's capital stock. These errors can be potentially large and extremely difficult to reduce through improved collection efforts since the errors are expected to accumulate over time as the capital stock ages, and is renewed. These errors can potentially suggest a much lower contribution of capital, relative to productivity, in explaining output and output growth.

Unlike most inputs in production (such as labor, energy, or materials), capital purchases are typically consumed over a long period, and the assets are rarely resold.[2] This introduces a few key challenges. Unlike any other input in production, capital expenditures need to be allocated inter-temporarily, in light of the asset's expected lifespan. This is not an issue with other inputs in the production process whose consumptions occur contemporaneously within the time of purchase. Therefore the value of the capital stock (at any given point in time) needs to be computed without observing a transaction that indicates the economic value.

The current state of the art is to either assume that accounting values accurately reflect the underlying economic value of capital stock, or to compute capital stock assuming a common and fixed depreciation schedule across producers.[3] These approaches thus introduce doubt to most, if not all, productivity analyses, hereby covering both producer-level correlates of productivity and aggregate productivity; e.g. understanding the contribution of information technology, managerial practices, integration in global markets, and the forces behind allocative efficiency and reallocation.[4]

While it is well-known by practitioners that the measurement of capital assets poses a problem for the estimation of production functions, and the associated output elasticities and productivity

[1] Separate review articles have been (partly) dedicated to discuss productivity measurement, and the role of measurement error in inputs – see Bartelsman and Doms (2000), Syverson (2011), and De Loecker and Syverson (2021). See Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg, Caves, and Frazer (2015) for specific applications.

[2] There are a few industries that depart from this pattern, such as the large market for leased aircraft used by airlines Gavazza (2011) or ships that transport bulk cargo Kalouptsidi (2014).

[3] The accounting value of the capital stock (reported as the book value of capital) is often observed in production data. In rare instances the book value can be contrasted to the computed capital stock measure (using perpetual inventory methods), and they turn out not to line up. In the U.S. Census data on manufacturing, the perpetual inventory and capital stock measures differ by 15 to 20 percent for example, see Becker, Haltiwanger, Jarmin, Klimek, and Wilson (2006).

[4] See Brynjolfsson and Hitt (2003), Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013), De Loecker (2007), Olley and Pakes (1996); Collard-Wexler and De Loecker (2015); Baily, Hulten, and Campbell (1992).

estimates, there is scant evidence on the severity of this problem, let alone how to address it successfully. Griliches and Mairesse (1998) features a discussion on the role of measurement error in the context of production function estimation. In particular, the presence of measurement error in capital is, perhaps indirectly, reflected in the well-documented fact that when estimating production functions with firm fixed-effects, capital coefficients are extremely low, and sometimes even negative.[5] One interpretation is that capital is a fixed factor of production, and, therefore, the variation left in the time series is essentially noise. However, this also implies that changes in capital, which capture depreciation, are potentially contaminated by measurement error. In an in-depth study of measurement issues related to capital, Becker, Haltiwanger, Jarmin, Klimek, and Wilson (2006) find that different ways of measuring capital that ought to be equivalent, such as using perpetual inventory methods or inferring capital investment from the capital producing sectors, lead to different results for a variety of outcomes, such as parameter estimates of the production function, and the investment and capital patterns.

An additional complication comes from the fact that the characteristic of which we wish to know its productivity premium, (take firm size, for example), is expected to be strongly correlated with the capital stock as well. Therefore even the simple correlation between a producer's characteristic and productivity is highly sensitive to the estimated capital coefficient. Consequently, the presence of measurement error in capital may therefore substantially alter the conclusions around the sources and drivers of productivity growth.

This paper shows that commonly used estimation techniques in the productivity literature fail in the presence of plausible amounts of measurement error in capital. We propose a novel approach that introduces an IV procedure that nests the most common methods in the literature to estimate production functions, including control function and dynamic panel approaches. In particular, we introduce an identification scheme, and an accompanying estimation procedure, that jointly deals with measurement error in capital and the standard simultaneity bias due to unobserved productivity shocks. Our strategy is based on the insight that once persistent productivity differences across producers are controlled for, a set of instruments become available that in standard IV approaches would fail. Within our framework we propose two instruments: lagged investment and the replacement value of capital.

We first validate our approach through Monte Carlo experiments, and trace out the relationship between the bias in the capital stock and the size of the measurement error. Second, we apply our procedure to three datasets covering three distinct economies: China, India and Chile, as well as cross-

[5] Griliches and Mairesse (1998) state, "*In empirical practice, the application of panel methods to micro-data produced rather unsatisfactory results: low and often insignificant capital coefficients and unreasonably low estimates of returns to scale.*".

country data from the World Bank. We find capital coefficients that are typically two times larger than those using standard methods that only control for simultaneity, suggesting sizeable measurement error in capital, at least interpreted through the lens our Monte Carlo experiments. Third, we underscore the importance of obtaining reliable capital coefficients, and document the bias in the productivity premium of technology, size, and ownership in the context of our applications. Throughout, we document a smaller role for unobserved productivity differences, suggesting a more prominent role for capital (and the underlying accumulation process) in explaining output differences. In one case, looking at the productivity advantages of firms with electric generators in India, the difference between generator and non-generator plants disappears entirely when we apply our instrumenting strategy. In addition we find that the correlation between size and productivity falls by a third when we use our correction for measurement error across a variety of data sets, underscoring the sensitivity of allocative efficiency analysis to measurement error in capital.

This is not the first paper to consider the role of capital measurement error on the estimation of production functions. An earlier paper by Tybout (1992) does this in the context of a maximum likelihood estimator, while Galuscak and Lizal (2011) does so in a Levinsohn and Petrin (2003) framework. Both of these papers attempt to use components of capital that are claimed to be measured without error, such as machinery or depreciations, however, the overall concern still stands. A more modern approach to measurement error is taken by Kim, Petrin, and Song (2016) providing an identification proof utilizing non-classical measurement error theory, along with restrictions on the markovian behavior of the structural error term, to filter out measurement error. However, these non-classical measurement error models are challenging to implement, so, to our knowledge, no estimator has yet been used on either real or simulated data (as in a Monte-Carlo). In addition, the topic of measurement error in control variables is also addressed by Hu, Huang, and Sasaki (2020). Van Biesebroeck (2007) evaluates the performance of various production function estimators in the presence of measurement error, although not with a specific focus on measurement error in capital.

In this paper, we provide an estimator that is robust to the presence of such measurement error, in the context of endogenous input choices. Our goal is not to offer an alternative approach to improve the measurement of capital, but to provide an estimator that is insulated from it, at least in the context of the errors-in-variable structure.

The remainder of the paper is organized as follows. Section 2 relates our approach to the existing literature in greater detail, and we connect the presence of measurement error to the broader debate around sources of productivity differences across producers, and factors of productivity growth. In Section 3, we introduce the identification strategy, and we present the associated estimation algorithm. We present the impact of measurement error in capital in a controlled Monte Carlo setting in Section

4, providing a mapping from the size of this error to the bias in the estimated capital coefficients. In Section 5 we apply our approach to three distinct datasets covering producers in China, India, Chile, as well as the World Bank Enterprise Data, and document the bias in the estimated capital coefficients. Section 6 illustrates the impact of this bias in obtaining reliable estimates of the productivity premium of technology, ownership, and firm size, in the context of our applications. The last section concludes.

2 The Role of the Capital Coefficient in Productivity Analysis

Dealing with measurement error in capital is critical to obtain reliable estimates of the capital coefficient. These coefficients feed into, not only into the calculation of the productivity residual, but also in computing the marginal product of capital, and decomposing output growth into its underlying sources. We recall an older debate on the relative importance of capital accumulation versus total factor productivity growth in explaining periods of high-growth throughout the world, at various points in time. The essence of the debate can be traced back to the measurement of the capital stock, and its growth, in the national accounts.[6] If a country, a region, or an individual producer for that matter, experience a growth in output, it can come from factor accumulation or productivity growth. Economists care greatly as to which one it is, for the plain and simple reason that one is free, and the other is not. Factor accumulation, in the case of capital, requires investment, on top of offsetting depreciation, and it is privately-held, and purchased in the capital market at prevalent market prices; and thus imply a user cost of capital. In contrast, if output rises due to productivity growth, say due to technological progress, all factors become more productive.

Against this background, a separate literature developed taking the newly arrived micro-datasets to estimate production functions. A major focus was on developing approaches that can deal with the transmission bias, leading to otherwise biased factor shares of the various inputs. This literature, starting with Olley and Pakes (1996), was also interested in productivity growth episodes, and detecting the drivers behind it. One important take-away from this literature is that the estimate of the factor share of capital is greatly affected by whether we recognize that a producer's input choices are systematically related to productivity, and that as industries mature a non-random set of producers is forced to exit. In the case of US telecom equipment producers, Olley and Pakes (1996) report a capital

[6]This debate between productivity and factor accumulation was particularly lively in the aftermath of the East Asian growth miracle. The industrial revolution in several East Asian countries, during (roughly) 1960 and 1990, led to periods of rapid growth; Taiwan, Singapore, Hong Kong and Korea recorded average annual output-per-worker growth of 4.3, 4.2, 4.7 and 4.9 percent, respectively. The conflicting views on the source of this phenomenal output growth are reflected in several prominent papers. Hsieh (2002) put forward evidence, using duality theory, that indeed the source behind the output growth was TFP, while Young (1995) concluded that factor accumulation was the main driver behind this success. A crucial ingredient is thus the relative contribution of capital and productivity to output.

coefficient of 0.35, compared to the within-estimator (the benchmark at the time) of 0.06. While no attempts in this study, or in any of the subsequent work, were made to connect this to aggregate implications, it is clear that any conclusion, on what the drivers of output growth might be, will be highly sensitive to which point estimate is used. In a similar type of analysis Collard-Wexler and De Loecker (2015) demonstrate the importance of the capital coefficient in detecting the potential productivity effects from the introduction of a new technology (in their case the minimill in steel production).[7]

While we broadly relate to the literature on the relative importance of productivity and capital, as prominently featured in the literature mentioned above, we do not engage in a revision of a particular debate, but our findings on the magnitude of the capital coefficient do impact this discussion. Furthermore, we expect measurement error in capital to be problematic for any analysis of micro-level correlations of productivity. If we systematically underestimate the factor share of capital, we also underestimate its marginal product. This underestimate of the role of capital will be loaded into the productivity residual. Thus, any firm characteristic that is correlated with capital, such as firm size, export participation, research and development or management, among others, will falsely appear to be associated with higher productivity. We highlight this in light of our four data sets, covering producers in China, Chile, India, and the World Bank data (WBES), where we demonstrate the sensitivity of the productivity premium of ownership, size and technology respectively. Our analysis points to an important bias in *any* productivity analysis where the interest lies in studying the link between productivity and a variable of interest that is correlated with capital.

3 Identification and Estimation Strategy

We start by presenting the identification strategy to accommodate both unobserved productivity shocks and measurement error in capital. This strategy suggests a straightforward estimation routine that does not add any complexity to the standard estimation routines such as the class considered by Ackerberg, Benkard, Berry, and Pakes (2005) and Ackerberg, Caves, and Frazer (2015) (henceforth ACF).

3.1 Identification under measurement error and productivity shocks

We consider a production function of the type:

$$q_{it} = f(m_{it}) + \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it}, \tag{1}$$

[7] See De Loecker and Syverson (2021) for a detailed discussion of these two studies.

where $q_{it}, l_{it}, k_{it}, m_{it}$ and ω_{it} denote (log) output, labor, capital, materials and productivity, respectively. This captures the class of Hicks-neutral production functions including the Gross Output and Leontief Cobb-Douglas specifications – i.e., $f(m_{it}) = \beta_m m_{it}$ and $f(m_{it}) = 0$, respectively.[8] In levels these two production functions are given by

$$Q_{it} = \begin{cases} \beta_0 M_{it}^{\beta_m} L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it} & \text{(Gross Output)}, \\ \min\{\beta_m M_{it}, \beta_0 L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}\} & \text{(Leontief)}. \end{cases} \tag{2}$$

In what follows we consider the standard DGP considered by ACF that allows to identify the production function that is perfect competitive output markets and competitive input markets with common factor prices. This allows us to focus on the role of measurement error in capital. In Appendix A.2 we briefly discuss departures from this perfect competition framework.

The laws of motion Capital and productivity follow the (standard) law of motions:

$$K_{it} = (1-\delta)K_{it-1} + I_{it-1} \tag{3}$$

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it} \tag{4}$$

This productivity process is assumed throughout the dynamic panel data literature, while the control function approach typically assumes a first-order (exogenous) Markov process, that allows for a non-linear process of the form $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$. In practice, an AR (1) process is almost always used. Moreover, when the properties of new estimators, such as ACF, are evaluated via Monte Carlo experiment, these also assume an AR (1) process.[9]

Measurement error in capital We rely on the commonly assumed errors-in-variable structure for capital, where the observed log of the capital stock (k) is the sum of the log true capital stock (k^*) and the measurement error (ϵ^k):

$$k_{it} = k_{it}^* + \epsilon_{it}^k, \tag{5}$$

where we use the * notation to denote variables measured without error – the one that is typically observed by the firm – and the unstarred notation to denote the observed value. This linear form of measurement error allows us to use IV to rid the model of bias from error-in-variables.

[8]To avoid the unlikely case that the data on materials is generated by unexpected large swings in the price of materials after labor and capital are installed, such that the Leontief FOC would fail to hold, we do not consider the more general Leontief structure $G(M_{it})$, but rather $\beta_m M_{it}$. However, our approach can accommodate this more general case.

[9]Our approach does, however, accommodate endogenous processes of the form $\omega_{it} = \rho_1\omega_{it-1} + \rho_2 d_{it-1}$, where d_{it} is an action taken by the firm to impact future productivity, e.g. export status or innovation. See De Loecker (2013) for a discussion, and Braguinsky, Ohyama, Okazaki, and Syverson (2015) for a recent application.

We assume that the ϵ_{it}^k is classical measurement error — i.e., it is uncorrelated with true capital stock k^*. We do, however, allow for ϵ_{it}^k to be serially correlated over time (within a producer).[10] This means that k_{it-1}, or any other lagged value of capital are not valid instruments. Indeed, since capital is constructed using historical information on the cost of assets, it is unlikely that there is no serial dependence in measurement error of these values.

We further assume that this measurement error is orthogonal to the instrument z_{it}[11]:

$$\mathbb{E}(z_{it}\epsilon_{it}^k) = 0. \tag{6}$$

Later in the paper, we will use as instruments measures of the replacement value of capital, last period's investment choice, or policy variables that move investment.[12]

Simultaneity with Errors-in-Variables Let us consider the empirical counterpart of the production function. We follow the literature and include measurement error in output. Measured output is $q_{it} = q_{it}^* + \epsilon_{it}$ where q^* is true output and ϵ_{it} is output measurement error. Thus,

$$q_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \eta_{it}, \tag{7}$$

where η_{it} is a composite error term including productivity, and measurement error in capital and output:

$$\eta_{it} \equiv \omega_{it} - \beta_k \epsilon_{it}^k + \epsilon_{it}. \tag{8}$$

The search for an instrument (z_{it}) in this setting consists of satisfying both :

$$\mathbb{E}(k_{it}z_{it}) \neq 0 \text{ and } \mathbb{E}(\eta_{it}z_{it}) = 0. \tag{9}$$

This implies that not only that the instrument needs to be orthogonal to the capital and output measurement error ($\epsilon_{it} - \beta_k \epsilon_{it}^k$), but also orthogonal to the productivity term ω_{it} — to avoid the standard simultaneity bias. Both simultaneity and error in variables lead to endogeneity. The literature has long recognized that an IV for the latter is difficult to find, since candidates that are correlated with inputs are, more often than not, also correlated with productivity, and therefore fail to satisfy the exclusion restriction. Input prices are a typical example, since they are expected to be correlated with the inputs, but at the same likely correlated with productivity.

[10] In the Monte Carlo experiments we assume that it follows an AR (1) process, $\epsilon_{it}^k = \rho^k \epsilon_{it-1}^k + u_{it}$, but this is not necessary for our identification results.

[11] Error in variables will lead to the usual attenuation bias problem Hausman (2001). Asymptotically, the estimate of capital will be downward biased by $\sigma_k^2/(\sigma_k^2 + \sigma_\epsilon^2)$, the usual noise to signal formula.

[12] We can allow for measurement error in the instrument z_{it} , as long as this measurement error is not correlated with the measurement error in capital ϵ_{it}^k.

Our setup adds a second source of endogeneity through the presence of capital measurement error, and we now explain how our approach can deal with both sources of endogeneity – i.e., $\mathbb{E}(k_{it}\omega_{it}) \neq 0$; the simultaneity problem, and $\mathbb{E}(k_{it}\epsilon^k_{it}) \neq 0$; the errors-in-variable problem.

First, merely using an instrument for capital in a linear IV context would not work. For instance, suppose we use as an instrument a second measure of capital, based on the replacement value of capital, which we denote k^R_{it}. The underlying true capital stock depends on past investment choices. These choices, in turn, are based on past value of productivity, which are typically highly correlated with current values of productivity. Thus, $E[k^R_{it}\omega_{it}] > 0$. For the same reason, using investment also does not work as a pure IV.[13]

Second, standard approaches — think ACF — rely on a moment condition where productivity shocks (ξ_{it} – i.e., the innovation in the productivity process, equation (4)) are orthogonal to the capital stock (k_{it}). However, when capital is measured with error, this productivity shock contains this measurement error. This implies that the standard moment conditions fail to identify the capital coefficient (β_k).

Our approach, however, relies on a different conditional orthogonality condition:

$$\mathbb{E}(\eta_{it}z_{it}|\omega_{it-1}) = 0. \tag{10}$$

This expression highlights that we require our instrument to be orthogonal to contemporaneous and lagged measurement error in capital (stated in equation (6)). In practice, we exploit the implicit (log) linearity of the production function and the underlying productivity process, present in the most common approaches, to deal with both the simultaneity bias and measurement error in inputs. Using either control function or dynamic panel techniques allows us to isolate the productivity shock, on which we can form moments for estimation. The moment conditions can be formed in a very flexible way to account for both measurement error in inputs, here capital, and various model specifications, including the speed of adjustment of inputs and, the market structure of output and input markets.

Other inputs The coefficients of the other inputs, in our case labor and materials, are identified using standard moment conditions (depending on what one is willing to assume about factor market frictions and adjustment costs). However, since inputs are correlated, we do expect these coefficient to be biased in the presence of measurement error in capital. Under a simple Cobb-Douglas production in labor and capital, and in the absence of productivity shocks, one can show that the presence of

[13] Indeed, in our Monte Carlo experiments in Section 4 the IV estimator is higher than the OLS estimator. This pattern is mimicked by our applications, so we present this IV estimator, to illustrate the bias it causes, but also, to show what is being leveraged in the data to yield consistently higher capital coefficients.

measurement error in capital will lead to an upward bias of the labor coefficient. The intuition behind this result is that labor will pick up the true capital variation through the positive correlation of labor and (true) capital, and thus overestimate the impact of labor on output variation.[14] This has further implications for returns to scale. Correcting for the presence of measurement error in capital leads to ambiguous returns to scale implications; depending both on the severity of the error in capital recording and the strength of the correlation between labor and capital in the data. In our setting, with additional unobserved productivity shock, signing the bias is not straightforward.

Selection We focus on the endogeneity of inputs, but do not model the exit process explicitly. There is value in combining the insights of Olley and Pakes (1996) on using a selection model to account for the non-random exit of producers, with the identification strategy in this paper. However, capital is a key variable in any model of firm survival, and therefore measurement error in capital immediately impacts the ability to correctly estimate, and predict, the survival probability needed to correct for firm exit.

We therefore appeal to the standard observation that relying on unbalanced panels controls for selection based on the unobserved productivity term. This does, however, not capture selection based on other relevant firm characteristics, including capital for example.[15]

Units of output and inputs In most settings, firms charge different prices for their output while paying different prices for inputs, which leads to an additional complication since researchers typically have access to only (deflated) revenues and expenditures on inputs.[16] We believe this to be a very important concern. The starting point of our analysis, however, is to assume we have correctly converted the revenue and expenditure data to the comparable units in a physical sense, and this is precisely the setup of Ackerberg, Caves, and Frazer (2015), and prior work. See Appendix A.2 for more discussion.

3.2 Extracting the productivity shock

The key step in our approach is to extract the productivity shock, thereby eliminating the persistent part of the productivity process. The latter not only creates the transmission bias for inputs facing adjustment costs (of which the capital input is a prominent example), but it also severely limits the

[14] See chapter 11 of Maddala and Lahiri (1992) and Pischke (2007) for a formal treatment of the signal-to-noise bias in multivariate regression analysis.

[15] For an application of a selection control in the context of additional relevant state variables, see Collard-Wexler and De Loecker (2015).

[16] See De Loecker and Goldberg (2014) for a detailed discussion of the implications and potential solutions.

availability of valid instruments to combat the measurement error in capital. Once the productivity shock ($\tilde{\xi}_{it}$) is extracted, we can introduce instruments that are strongly correlated with the underlying capital stock.

There are two distinct approaches to isolate this productivity shock ($\tilde{\xi}_{it}$) from the overall productivity unobservable (ω_{it}): control functions and dynamic panels. The control function approach (ACF) relies on an inverted factor demand equation to isolate the productivity shock by following a two-step procedure. After a first stage, where output is projected on inputs and other relevant factor demand shifters, a productivity shock is extracted using the law of motion on productivity. The dynamic panel data approach, instead, proceeds by replacing the unobserved productivity term by its specific (linear) process. In terms of our moment condition, equation (10), these two approaches can then be written as:

$$\mathbb{E}(\tilde{\xi}_{it} z_{it}) = 0, \tag{11}$$

where $\tilde{\xi}_{it}$ is the extracted productivity shock, that contains the measurement error in capital, in addition to the standard productivity shock. We briefly remind the reader of how these two approaches proceed to extract this shock, and focus specifically on how the various errors enter. We provide more details in the next Section where the estimation algorithm is introduced.

Control function approach Productivity is replaced by the inverted input demand equation, and after a first stage projection, the productivity shock is obtained (for a given value of the parameters) using:

$$\tilde{\xi}_{it}(\beta) = \omega_{it}(\beta) - \rho\omega_{it-1}(\beta), \tag{12}$$

Where $\omega_{it}(\beta) = q_{it} - \beta_0 - \beta_l l_{it} - \beta_k k_{it}$ and in the presence of measurement error:

$$\tilde{\xi}_{it} = \xi_{it} - \beta_k \epsilon^k_{it} + \rho\beta_k \epsilon^k_{it-1} \text{ (Control Function)} \tag{13}$$

Dynamic panel approach Productivity is replaced by the productivity process in the production function:

$$\begin{aligned} q_{it} &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \rho\omega_{it-1} + \xi_{it} + \epsilon_{it} - \beta_k \epsilon^k_{it} \\ &= \beta_0(1-\rho) + \beta_l(l_{it} - \rho l_{it-1}) + \beta_k(k_{it} - \rho k_{it-1}) - \rho q_{it-1} + \tilde{\epsilon}_{it} \end{aligned} \tag{14}$$

Where the composite error term is given by:

$$\tilde{\epsilon}_{it} = \xi_{it} + \epsilon_{it} - \rho\epsilon_{it-1} - \beta_k \epsilon^k_{it} + \rho\beta_k \epsilon^k_{it-1} \text{ (Dynamic Panel)} \tag{15}$$

The two approaches are thus very similar, with the main difference (under our AR(1) productivity process) coming from the inclusion of measurement error in output in the productivity shock under the dynamic panel. The moment conditions for the control function are:

$$E\left[(\xi_{it} - \beta_k \epsilon_{it}^k + \rho \beta_k \epsilon_{it-1}^k) z_{it}\right] = 0, \tag{16}$$

while for the dynamic panel, they are:

$$E\left[(\xi_{it} + \epsilon_{it} - \rho \epsilon_{it-1} - \beta_k \epsilon_{it}^k + \rho \beta_k \epsilon_{it-1}^k) z_{it}\right] = 0. \tag{17}$$

Alternative approach: Factor shares An often used alternative to production function estimation is the factor-share approach. In general, and in order to accommodate imperfect competitive output markets, the output elasticities are directly computed from the data, through the ratio of an input's expenditure over total costs.[17] The latter includes the payments to capital which rely not only on an estimate of the user cost of capital, but also on a measure of the capital stock. Therefore, the factor-share approach is subject to a potential measurement error problem as well, albeit it is expected to impact the output elasticities differently since these are not estimated, rather they are computed over a sample of producers and a period of time. While this may sound as good news for this approach, this advantage comes at the cost of having to assume constant returns to scale and no adjustment costs in any factor of production, conditions that can be problematic.[18]

3.3 Control function estimation

We present the main estimation algorithm that nests the standard approach in the literature, but considers serially correlated measurement error in capital.

The estimation procedure consists of two steps, and follows the discussion in ACF closely.[19] The main difference is that we exploit the (log) linear production structure in capital, labor and productivity, which dictates a (log) linear control for productivity. The log linearity of the production function guarantees that the control for productivity (through the first-order condition used to control for unobserved productivity) will also be log-linear, and allows to treat the error in capital measurement.

[17] See De Loecker and Syverson (2021), Foster, Haltiwanger, and Syverson (2008) and De Loecker, Eeckhout, and Unger (2020) for applications of the approach.

[18] Under constant returns to scale, no adjustment cost and perfect competitive output markets, revenue shares (the ratio of an input's expenditure in total revenue) can be used to estimate the production function Asker, Collard-Wexler, and De Loecker (2014), while under imperfect competition, the user cost of capital — which is rarely measured directly — needs to be observed.

[19] The Not-for-Publication Appendix F provides precise details on these estimators, and our webpage has code for these in STATA.

Under the Leontief production function, the cost minimizing material choice gives a simple log-linear control for productivity given by:

$$\omega_{it} = \ln \beta_m + m_{it} - \beta_l l_{it} - \beta_k k_{it} - \beta_0 \text{ (Leontief)}. \tag{18}$$

Note that this *control* is independent of the presence input price variation, and imperfect competitive output or input markets. It merely states that the firm is optimally deploying the mix of inputs given the technology. Alternatively, under the Gross Output production function we follow ACF-LP and invert the material demand equation (under profit maximization) and (also) obtain a log linear expression for productivity given by:

$$\omega_{it} = -\ln \beta_m + (1-\beta_m) m_{it} + p_t^m - p_t - \beta_l l_{it} - \beta_k k_{it} - \beta_0 \text{ (Gross Output)}, \tag{19}$$

where all lower cases denote logs, and p^m and p are the material input and output price, respectively.[20] Note that in either the Leontief or Gross Output, we obtain log-linear control functions.[21]

In both specifications, we obtain a log-linear first-stage given the log linearity of the control function for productivity. The investment control function approach of Olley and Pakes (1996) does, in general, not lead to a log-linear control for productivity.[22]

The estimation routine is identical across the two production technologies, with one difference that under the gross output specification the material coefficient is computed directly from the first order condition. The identification and estimation of the material coefficient thus follows immediately from the assumption of perfect competitive output markets. Taking the first-order condition of profits with respect to materials immediately generates $\beta_m = \frac{P^m M}{PQ}$, as an estimate. Due to the potential measurement error in output (ϵ), taking the average of the material-revenue share would yield a biased estimate. Instead, we take the median material revenue share across the sample to obtain a consistent estimate of β_m. While the coefficient on materials is often not of direct interest in the case of the Leontief production function, it is given by the ratio of (physical) materials-to-output, $\frac{M}{Q}$, where again we consider the median across firms and time.[23]

[20] In this setting the input demand equation does depend on the specific features of output and input markets, such as input price variation and imperfect competition.

[21] Replacing the productivity shock in the production function with the control for productivity implies that, strictly speaking, the first stage becomes redundant. See footnote 17 on page 2444 of Ackerberg, Caves, and Frazer (2015) for a discussion of this issue.

[22] For an example of what assumptions need to be made to obtain a log-linear investment function, see the setup in the Monte Carlo, and in particular the log linear investment policy function derived in equation B.12 in the Appendix.

[23] The presence of measurement error in output thus invalidates relying on average factor shares. See Appendix A.1 for more details and also see Asker, Collard-Wexler, and De Loecker (2014) and De Loecker, Eeckhout, and Unger (2020) for an application of this estimator.

Estimation routine To encompass both production technologies we denote *output* by y_{it}, where it is understood that under the gross output specification it is given by $y_{it} = q_{it} - \hat{\beta}_m m_{it}$, while under the Leontief specification it is output ($y_{it} = q_{it}$), and $\hat{\beta}_m$ is the estimated material coefficient.

Replacing the unobserved productivity term by the control function (under both models considered above) gives:

$$\begin{aligned} y_{it} &= \theta_t + \theta_l l_{it} + \theta_k k_{it} + \theta_m m_{it} + \tilde{\epsilon}_{it} \\ &\equiv \Phi_{it} + \tilde{\epsilon}_{it} \end{aligned} \tag{20}$$

where θ_t captures the constant term and the time-varying prices (under the gross output specification), $\tilde{\epsilon}_{it} = \epsilon_{it} - \theta_k \epsilon_{it}^k$, and we use θ_h ($h = \{l, m, k\}$) to denote the reduced-form parameters, on each input, capturing both the production function coefficients and the parameters governing the control for unobserved productivity.

The first stage consists of running a simple linear IV-regression where capital is instrumented with z_{it}. The moment conditions to obtain predicted output are given by:

$$\mathbb{E}\left[\tilde{\epsilon}_{it}\begin{pmatrix} l_{it} \\ z_{it} \\ m_{it} \end{pmatrix}\right] = 0. \tag{21}$$

By using two stage least squares, we highlight that the capital stock is predicted using the instrument (z_{it}), labor, materials and year dummies.[24] Given parameters $\hat{\boldsymbol{\theta}}$, we predict $\hat{\Phi}_{it} = \hat{\theta}_t + \hat{\theta}_l l_{it} + \hat{\theta}_k k_{it} + \hat{\theta}_m m_{it}$, where $\hat{\Phi}_{it} = \Phi_{it} + \theta_k \epsilon_{it}^k$.

The moment conditions introduced in the previous section can now be formed by observing that the joint productivity-measurement error of capital term (conditional on the parameter vector $\boldsymbol{\beta}$) is obtained using $\omega_{it}(\boldsymbol{\beta}) + (\theta_k - \beta_k)\epsilon_{it}^k = \hat{\Phi}_{it} - \beta_0 - \beta_l l_{it} - \beta_k k_{it}$.[25]

Using the law of motion on productivity the moment conditions can be constructed to estimate the labor and capital coefficient, respectively:

$$\mathbb{E}\left[\tilde{\xi}_{it}(\boldsymbol{\beta})\begin{pmatrix} l_{it-1} \\ z_{it} \end{pmatrix}\right] = 0. \tag{22}$$

In contrast with the standard approach (ACF) where $\mathbb{E}(\tilde{\xi}_{it} k_{it}) = 0$ is used to identify the capital coefficient, we rely on our instrument being orthogonal to the productivity shock.[26]

[24] We abstract away from the time specific intercept in presenting the moment conditions.

[25] In practice, we will estimate a process for productivity given by $\omega_{it} = \rho_0 + \rho_1 \omega_{it-1} + \xi_{it}$. Thus, β_0 and ρ_0 are not separably identified, and we will just estimate ρ_0.

[26] As in ACF, we concentrated out β_0 in the estimation of the function $\omega_{it} = \rho\omega_{it-1} + \alpha$, where α is an intercept.

Our approach, however, preserves the flexibility of the ACF approach with respect to selecting an alternative DGP, and in the absence of measurement error both estimators (ACF and ours) coincide. In the next section, we introduce our Monte Carlo analysis, to demonstrate the performance of our estimator.

3.4 Selecting instruments

So far we have only treated the instruments in general, highlighting that the virtue of conditioning on the persistent part of productivity alleviates the burden of searching for an instrument that is orthogonal to productivity all together. However, in this paper we do more than just state this condition, we also introduce, and evaluate two instruments, using both Monte Carlo experiments and applications to four well-known productivity datasets.

The previous discussion on the sources of measurement error in capital immediately suggests several instruments. Since the issue is the difficulty in depreciating older capital assets, taking a shifter of capital stock that does not contain depreciation is a good angle of attack. We propose to use either the replacement value of capital, (lagged) investment, or policy variables that shift investment, to identify the marginal product of capital. This solicits the cross-sectional variation present across producers, rather than the year-on-year variation in capital stock, where poorly measured depreciations attenuate the capital coefficient. In other words, our framework relies on either instances where measured capital differences (across producers and time) line up with observed investment activity, or when producers evaluate the value of the capital stock used in production at a given point in time.

Replacement value of capital Some datasets, later in the paper we will be using the World Bank Enterprise Survey, directly ask a question as to the replacement value of capital. Since this replacement question is about the cost of purchasing the firm's current capital, it does not include any error due to depreciation, but will be correlated with the firm's true capital stock k^*.[27] This instrument does not rely on a timing assumption or a particular model of capital formation. Instead it starts from the observation that, in practice, some producer-level survey asks producers about the economic value of the capital stock, as measured by the replacement value thereof, sometimes also labeled the replacement value. In principle, this is the cost of replacing all existing assets with identical but new assets or an appropriate equivalent. Unlike the capital stock, this measure fluctuates with overall market-level conditions that determine the value of assets used in production. As such, this

[27] The idea of measuring the replacement value of a company's capital stock is famously used in the definition of Tobin's Q. Most applications of this theory, however, have had to use imperfect proxies for the replacement value of capital, rather than direct questionnaire items.

measure is closely related to the discussion on the use of appropriate capital price deflators. Again, satisfying the exclusion restriction in this case requires this value to be orthogonal to the measurement error in capital. We motivate this instrument as a more traditional instrument, since it is a second, and alternative measure of the underlying true capital stock.

Lagged investment While past investments need to be appropriately depreciated, current investment, which is directly related to the price of capital in the market for investment goods, does not suffer from the same issue with depreciation. Thus, we also use investment as an instrument for capital. However, we need to argue that this instrument is plausibly orthogonal to the measurement error in capital. A first and important observation is that by considering lagged investment, we immediately eliminate the concern that productivity shocks (ξ_{it}) are correlated with the chosen instrument given the assumption of a one period time to build. We are considering the time to build assumption for investment used in Olley and Pakes (1996): investment is chosen at time $t-1$, but only gets added to capital at time t.

In the general case where we consider serially correlated measurement error in capital, the orthogonality condition then states that lagged investment and measurement error to capital (both at t and $t-1$) are uncorrelated. We rely on the notion that gross investment expenditures are based on transactions reflecting accurately the purchase of capital goods, and these are not subject to unknown depreciation schedules, while accurately capturing the vintage of capital goods, and are not subject to the accumulated measurement errors from the past. In other words, we treat the investment data just like the wage bill and material input expenditures, and rule out (explicitly) that investment is correlated with the errors injected in the construction of the capital stock. This does not, however, rule out that investment itself may be measured with error, as long as this error is uncorrelated with the measurement error, our approach is unchanged.

Under the perpetual inventory method approach, however, the user computes the capital stock using data on investments (typically for different asset groups) and assumes a particular depreciation schedule for these assets. If the source of the measurement error in capital comes uniquely from the error in the depreciation rates, the measurement error no long enters the model as a separable linear error, potentially jeopardizing the IV strategy. We perform an additional Monte Carlo experiment under a structurally derived measurement error model in Section C in the Appendix, and find that our approach still performs reasonably well under considerable measurement error levels.

Other instruments It is also possible to use variation in the effective tax rates on capital on different firms as another instrument that moves true capital, but is uncorrelated with measurement error. However, we have not found a case where these tax rates have sufficient power to be used in estimation. This is an important advantage for both the replacement value of capital and investment instruments we proposed: substantial power. There is a strong correlation in all the datasets we have examined between these instruments and measured capital stock.

4 Monte Carlo Analysis

We evaluate our estimator in a series of Monte Carlo analyses in which the main interest lies in comparing the capital coefficient across methods as we increase the level of measurement error in capital. We follow Ackerberg, Caves, and Frazer (2015) closely, starting with their data-generating process, and add measurement error to capital. We depart from ACF by adding time-varying investment costs in the investment policy function in order to generate additional time-series variation in capital stock.

We refer the reader to Appendix B for the details of the underlying model of investment, but the main features of our setup are as follows. We rely on a constant returns to scale production function with a quadratic adjustment cost for investment. This model yields closed-form solutions for both labor and capital, where labor is set using a static first-order condition given firm-specific wages.[28] Productivity and wages follow an AR(1) process, and we consider a quadratic adjustment cost for investment: $\phi_{it} I_{it}^2$, where ϕ_{it} — which should be considered as the price of capital — itself follows a first-order Markov process. We solve the model in closed form, extending the work in Syverson (2001) and discussed in Appendix B.2, and this generates our perfectly measured Monte Carlo dataset on output, inputs, investment and productivity.

We then overlay measurement error on this dataset composed of AR(1) processes with normally distributed shocks:

$$\begin{aligned} \epsilon_{it} &= \rho^q \epsilon_{it-1} + u_{it}^q \\ \epsilon_{it}^k &= \rho^k \epsilon_{it-1}^k + u_{it}^k, \end{aligned} \tag{23}$$

where $u^q \sim \mathcal{N}(0, \sigma_q^2)$, and $u^k \sim \mathcal{N}(0, \sigma_k^2)$. In other words, we allow for serially correlated measurement error in output and capital.

We analyze the impact of increasing ϵ^k, which is governed by the variance σ_k^2. We distinguish between the role of measurement error *within* a given Monte Carlo, and the overall distribution of estimated coefficients *across* 1,000 Monte Carlo runs.

[28] ACF also deal with two other data-generating processes, other than the approach we described (called DGP1 in their paper). In Appendix B.4, we also consider optimization error in labor (DGP2) and an interim productivity shock between labor and materials as in ACF (DGP3), along with optimization error in labor. We find similar results from any of the DGP's considered by ACF.

In addition, we generate the replacement value of capital instrument using the same measurement error process as measured capital stock, but with independent errors.

Table B.1 in the Appendix shows the parameters used in our Monte Carlo. We pick the same parameters for the size of the dataset, production function, and processes for productivity and wages as in ACF. For the process for the price of capital, denoted ϕ_{it}, we pick parameters that match the cross-sectional dispersion of capital ($std.[k_{it}] = 1.6$) and the time-series variation in capital ($Corr.[k_{it} - k_{it-1}] = 0.93$) in the Annual Survey of Industries in India (discussed in Section 5.2), choosing an autocorrelation term for the process of ϕ of 0.9 and a shock standard deviation of 0.3.[29]

Finally, we pick parameters for the measurement error in inputs and output. We choose a measurement error for output with a standard deviation of 30 percent, and a low autocorrelation of 0.2. For the error in capital (ϵ^k), we choose a serial correlation coefficient of 0.7, so a fairly high persistence, and a standard deviation of 0.2. Note that this assumption on the time series process for capital measurement error yields a difference between k and k^*, which has a standard deviation of 30 percent.

Table 1 presents the estimated capital and labor coefficients from this Monte Carlo exercise, with (true) parameters of the production: $\beta_k = 0.4$ and $\beta_l = 0.6$. We compare the performance of estimators that use investment to instrument for mismeasured capital proposed in this paper, that we call Control IV, and of those that do not, that we call Control estimator.

The Control IV estimators, both investment and replacement value of capital yield very similar results (both unbiased estimators) — i.e., the mean parameters are the same as the true ones — and, precise estimates, with a standard deviation of 0.02. In contrast, the standard Control function estimator yields downwardly biased estimates of the capital coefficient, with an average of 0.32 versus a true value of 0.4. As in ACF, the OLS estimator for capital is downwardly biased because of the simultaneity problem, with $\beta_k^{OLS} = 0.05$. The IV estimators are similarly biased, but show a substantially larger capital coefficiens than the OLS ones. Figure 1 plots the distribution of estimates of both the labor and capital coefficient showing that the Control and Control IV estimates have similar variances, but there is very little overlap in the distributions of the estimated coefficients.

4.1 The impact of measurement error

In Figure 2, we plot the average estimate of β_k over 100 replications against the standard deviation of capital measurement error σ_k for both the Control IV and Control estimators. This Monte Carlo simulation shows that standard estimator becomes progressively more biased away from the true value of β_k as the measurement error in capital increases. It is of course difficult to guess the relevant range

[29] Indeed, it is this last moment that the ACF Monte Carlo has difficulty replicating: it predicts a serial correlation coefficient of capital of 0.997, which is much more than in any producer-level dataset we are aware of.

Table 1: Monte Carlo Results
($\beta_k = 0.4, \beta_l = 0.6$)

	Coefficient			
	Capital		Labor	
OLS	0.05	(0.01)	0.95	(0.01)
FE	0.03	(0.01)	0.90	(0.01)
IV Investment	0.07	(0.01)	0.94	(0.01)
IV Replacement	0.07	(0.01)	0.94	(0.01)
Control	0.32	(0.02)	0.52	(0.02)
Control Investment IV	0.40	(0.02)	0.60	(0.02)
Control Replacement IV	0.40	(0.02)	0.60	(0.02)

Notes: Mean and standard deviation over 1,000 Monte Carlo replications presented of DGP 1 in ACF with capital measurement error. Control refers to the control function approach used in ACF, while Control IV refers to the control function that uses investment and replacement value of capital as an instrument. The estimates without control functions are OLS, IV with capital instrument by investments, FE with firm fixed effects.

of this standard deviation, but the main takeaway is that our IV-based estimator is insulated from this problem. The simulations do suggest that standard methods deliver an estimate of half the magnitude for a standard deviation in the capital measurement error σ_k of about 0.2, which corresponds to a standard deviation between k and k^* of 0.28 in the stationary distribution.

It is important to note that our estimator is robust to capital measurement error, while still undoing the simultaneity bias that typically plagues the production function estimation. Therefore, applying our estimator when the capital stock is accurately measured also provides consistent estimates of the production function coefficients.

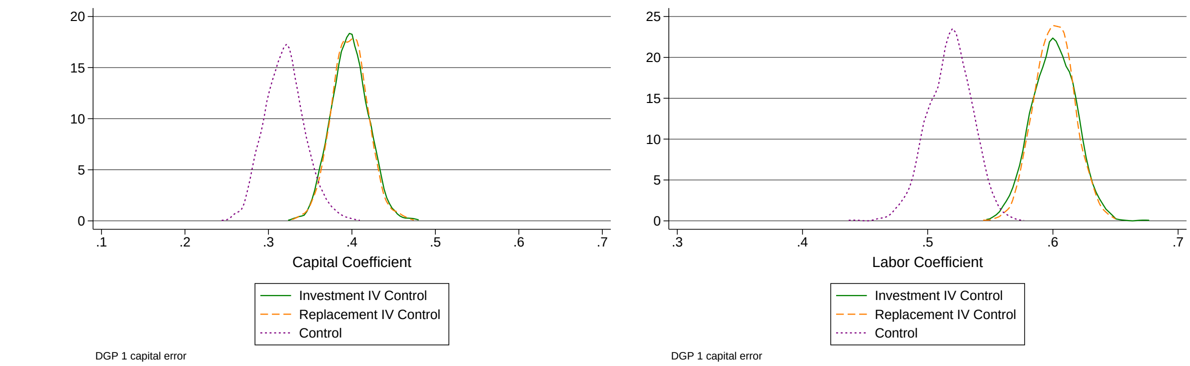
4.2 Alternative sources of measurement error

As discussed before, we have considered what we refer to as a reduced form for the measurement error in capital. That is, we consider the standard representation of an errors-in-variable, whereby the measurement error is (log) additive — here $k = k^* + \epsilon^k$. In Appendix C, we discuss an alternative source of measurement error in capital, derived structurally from the measurement error in depreciation rates: $K_{it} = (1 - \delta_{it})K_{it-1} + I_{it}$ and $\delta_{it} = (\delta + \epsilon^d_{it})$; i.e., there is measurement error in depreciation rates.

This form of measurement error does not map into a log-additive structure, and we evaluate our

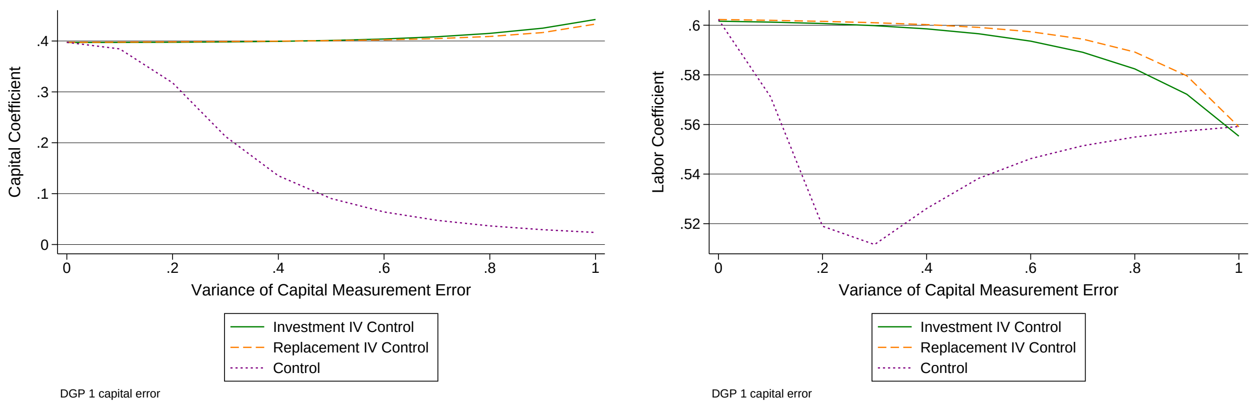
Figure 1: The distribution of the estimated capital and labor coefficient in a Monte Carlo ($\beta_k = 0.4$, $\beta_l = 0.4$)

(a) Capital (b) Labor



Notes: We plot the distribution of the estimated capital coefficient across 1000 Monte Carlo replications, with $\beta_k = 0.4$, and $\sigma_k = 0.2$.

Figure 2: Impact of measurement error on capital coefficient in a Monte Carlo ($\beta_k = 0.4$):



Notes: We plot the estimated capital coefficient as a function of the standard deviation in the capital measurement error. Average of 100 Monte Carlo replications per value of σ_k.

estimator in the presence of this alternative setup. The main takeaway from Appendix Figure C.1 is that our estimator outperforms the other approaches but, given the formal violation of the moment conditions, leads to a small bias of the capital coefficient for large values of the standard deviation of the capital measurement error.

Another potential source of measurement error stems from variation in capital utilization across producers, over time, see Gorodnichenko and Shapiro (2011). This poses a challenge as the variation in unobserved capital utilization is expected to be positively correlated with unobserved productivity. This implies that the exclusion restriction we propose is not expected to hold without making additional assumptions on the process of utilization.

The evidence from the Monte Carlo unequivocally favors our estimator in the presence of measurement error in capital and, moreover, suggests that the bias can be quite severe for moderate measurement error in capital. To verify how large this problem is in real data, we now apply our estimator, exactly as performed in our Monte Carlo analysis, to four different datasets of manufacturing plants, in India, China, Chile, and the World Bank.

5 Applications to Plant-level Data

We apply our approach to four distinct producer-level datasets: the World Bank Enterprise Data, the Annual Survey of Industries (henceforth ASI) from India, the NBS data from China, and the Chilean sample of manufacturing plants. These datasets have been used extensively to study productivity dynamics, but at the same time have distinct features related to the measurement of capital. The World Bank data asks a question on both total assets and the replacement value if one needed to repurchase all of a firm's capital. The data on Chinese establishments report the book value of plants and investment, while the Indian and the Chilean census data report both the book value and the (constructed) capital stock using the perpetual inventory method. In addition, the economic environments are different in important ways. There is substantial investment during the process of economic transition in China, while in India and Chile the presence of small firms and persistent frictions paints a very different industry dynamic over time. We expect these differences to materialize in the estimated coefficient, and in the role and importance of measurement error in the capital stock.

The World Bank Enterprise Survey (henceforth WBES) are a collection of firm-level surveys undertaken by the World Bank in a variety of countries from 2006 to 2020, used, for instance, in Asker, Collard-Wexler, and De Loecker (2014) and Cusolito, Lederman, and Pena (2021). The second dataset we use is from the Annual Survey of Industries from India. This a plant-level survey for over 600,000 plants over a twenty-year period. Allcott, Collard-Wexler, and O'Connell (2016) have previously used and described this dataset. The third dataset is the Chilean census, as used in Pavcnik (2002) and Levinsohn and Petrin (2003), and covers plants in the Chilean manufacturing sector for the period 1979-1986. Finally, we evaluate the impact of our correction to the Chinese manufacturing database provided to us by Brandt, Van Biesebroeck, and Zhang (2012). All variables are deflated using industry-specific price deflators and Appendix D describes each dataset briefly, and presents basic summary statistics.

5.1 World Bank Enterprise data

Table 2 shows estimates of the production function for the World Bank Data, which we present separately, since we have two instruments for capital: the replacement value of capital stock, and lagged investment. Throughout, we compare estimated production function coefficients (on labor and capital) from our approach to those obtained by simple OLS, IV (without the simultaneity control) and the control function approaches including the IV one presented in this paper. We show both the structural Leontief model ($q = \beta_l l + \beta_k k$) and for the more usual Gross Output production function ($q = \beta_l l + \beta_m m + \beta_k k$) where we use a first-order condition to estimate the material coefficient. We focus our discussion primarily on the Gross Output specification, since it uses a production function that is far more commonly studied in the empirical literature, but the pattern of results is similar for the Leontief specification.

In Table 2, the capital coefficient is 0.33 in the OLS specification, but falls to 0.14 when we include firm fixed effects, which is a common result in the literature when focusing on the variation within a firm.[30] Our next specification, IV (investment), considers a two-stage least squares regression of output on capital and labor, where we instrument for capital either the replacement value of capital, lagged investment, or both of these. The capital coefficient rises from 0.33 in OLS to 0.47 in the replacement value IV, to 0.58 with the investment instrument.[31] The coefficients on the IV with replacement value of capital are particularly interesting, since they suggest a large bias from measurement error given we are simply using an alternative measure for capital. However, investment and unobserved productivity are very likely to be positively correlated, so the increase in the capital coefficient in the IV regression might also reflect the endogeneity of investment. Indeed, this is the key insight of the investment control function developed by Olley and Pakes (1996), the additional explanatory power of investment in predicting output.

Turning to the control function approaches, the results echo the differences seen between OLS and IV. The control function without instruments, finds a capital coefficient of 0.27, rising to 0.38 with the replacement value of capital IV, to 0.55 with the investment IV. This pattern reflects the differences we found in the IV estimates without a control function. Notice that the structural Leontief

[30] See Griliches and Mairesse (1998).

[31] When we use both replacement value and investment the coefficient on capital is quite similar to the replacement value IV — both are 0.47 — which is unsurprising as replacement value is highly correlated with capital stock as compared with investment.

Table 2: Production Function Coefficients for World Bank Enterprise Data

	Leontief		Gross Output with FOC Materials		Observations
	Capital	Labor	Capital	Labor	
OLS	0.33	0.88	0.17	0.55	22,787
	(0.01)	(0.01)	(0.01)	(0.01)	
FE	0.14	0.59	0.05	0.38	20,597
	(0.02)	(0.05)	(0.02)	(0.04)	
IV⋄ Replacement Value of Capital	0.47	0.72	0.25	0.47	18,545
	(0.01)	(0.02)	(0.01)	(0.01)	
IV⋄ Investment	0.58	0.52	0.32	0.32	1,346
	(0.06)	(0.08)	(0.01)	(0.01)	
Control	0.27	1.02	0.16	0.59	2,193
	(0.02)	(0.04)	(0.01)	(0.01)	
Control-IV Replacement Value	0.38	0.87	0.23	0.51	1,937
	(0.03)	(0.05)	(0.01)	(0.02)	
Control-IV Investment	0.55	0.58	0.39	0.58	1,131
	(0.11)	(0.15)	(0.08)	(0.14)	

Notes: The Leontief specification is $Q_{it} = \min\{\beta_m M_{it}, L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}\}$, and the Gross Output with FOC Materials is given by $M_{it}^{\beta_m} L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}$, where the material coefficient is computed using the median of the material revenue share ($\beta_m = Median[\frac{P_t^M M_{it}}{P_t Q_{it}}]$) which is $\beta_m = 0.40$. Standard errors in parenthesis, clustered by firm. The F-stat for the investment instrument is 106, while the F-stat for the replacement of capital instrument is 17.895. ⋄: IV without a control for simultaneity.

estimates confirm the patterns of the Gross Output production functions.[32] This is solely due to the requirements that the data is a panel, and the large number of observations that are missing for investment (Levinsohn and Petrin, 2003).

We have found that the increase in the capital coefficient from our procedure is due to instrumenting for capital in both the first and second stage; we obtain substantially smaller coefficients if we only instrument in one of the stages.

5.2 Country and industry estimates

Table 3 shows estimates of the production function for India, China, and Chile. Note that in contrast to the World Bank data, we do not have a measure of the replacement value of capital, so the IV we are using is lagged investment. In addition, because there are more observations in these datasets, it is possible to estimate coefficients by industry and by country.

We start by reporting the median for the various estimators in Table 3, but also include the 25th and 75th percentile across industries to give an idea of the range of estimates. Again, using fixed effects lowers the capital coefficient substantially. In the case of the Gross Output production function, the capital coefficient falls from a median of 0.12 for OLS to 0.06 for firm FE in India, and from 0.10 to 0.05 for Chile, with no difference for China. By itself, this does not conclusively show that there is measurement error in capital. However, if capital is fixed over a long period of time, we cannot identify its marginal product using the time series variation within producers. China is a clear case where capital is changing at an unprecedented rate, quite atypical from the capital evolution of most countries. When we instrument capital using investment, we find substantially higher capital coefficients compared to OLS, of 0.06 versus 0.18 in China, 0.12 versus 0.17 in India, and 0.10 versus 0.22 in Chile, respectively. This reinforces our prior that instrumenting for capital with lagged investment may lead to a higher capital coefficient.

We now turn to the estimators that control for the simultaneity of inputs, and the impact of measurement error in capital. We find the capital coefficient increases from a median of 0.05 for an ACF control function to 0.12 for our Control IV approach for China. Likewise, the median coefficient increases from 0.09 to 0.19 for India, and from 0.09 to 0.16 for Chile. The fact that the capital coefficient roughly doubles is striking, but completely in line with the predictions from our Monte Carlo, say in Figure 2, if the standard deviation of the measurement error σ_k is around 0.2.

[32] The number of observations falls tremendously when we use either lagged investment as an IV, or when we use ACF approaches, going from 22,787 observations for OLS, to 1,346 for investment IV, or 2,193 observations for the control function. 47 percent of observations have non-positive investment. The OLS and IV estimates are fairly similar if we condition on the data where no observations are missing, such as the 1,131 observations used in the IV control function estimates.

Table 3: Production Function Coefficients

Gross Output with FOC Materials

	China		India		Chile	
	Capital	Labor	Capital	Labor	Capital	Labor
OLS	0.06	0.12	0.12	0.25	0.10	0.43
	[0.04 0.09]	[0.09 0.17]	[0.09 0.15]	[0.18 0.32]	[0.04 0.06]	[0.33 0.51]
FE	0.06	0.11	0.06	0.19	0.05	0.30
	[0.03 0.09]	[0.07 0.15]	[0.04 0.09]	[0.14 0.24]	[0.08 0.13]	[0.22 0.38]
IV	0.18	0.03	0.17	0.16	0.22	0.19
	[0.11 0.23]	[0.00 0.06]	[0.14 0.25]	[0.12 0.21]	[0.19 0.37]	[0.03 0.40]
Control	0.05	0.12	0.09	0.27	0.09	0.38
	[0.02 0.10]	[0.07 0.21]	[0.07 0.11]	[0.23 0.34]	[-0.01 0.12]	[0.33 0.57]
IV Control	0.12	0.10	0.19	0.19	0.16	0.21
	[0.08 0.17]	[0.04 0.19]	[0.12 0.20]	[0.17 0.30]	[0.13 0.35]	[0.03 0.59]

Leontief

	China		India		Chile	
	Capital	Labor	Capital	Labor	Capital	Labor
OLS	0.28	0.45	0.38	0.70	0.24	0.89
	[0.22 0.33]	[0.38 0.54]	[0.36 0.80]	[0.64 0.80]	[0.19 0.31]	[0.85 0.97]
FE	0.26	0.40	0.17	0.60	0.12	0.71
	[0.22 0.30]	[0.36 0.45]	[0.16 0.30]	[0.36 0.45]	[0.07 0.14]	[0.66 0.81]
IV	0.64	0.13	0.60	0.40	0.55	0.45
	[0.54 0.74]	[0.46 0.99]	[0.54 0.74]	[0.46 0.99]	[0.45 0.74]	[0.15 0.65]
Control	0.25	0.60	0.33	0.76	0.22	1.01
	[0.10 0.40]	[0.33 0.89]	[0.27 0.41]	[0.65 0.88]	[0.14 0.30]	[0.84 1.06]
IV Control	0.31	0.73	0.44	0.68	0.29	0.97
	[0.20 0.42]	[0.46 0.99]	[0.20 0.42]	[0.46 0.99]	[0.12 0.38]	[0.87 1.32]

Notes: We report the median coefficient across all industries for each dataset. In parenthesis we show the 25th and 75th percentile for each coefficient. The Leontief specification is $Q_{it} = \min\{\beta_m M_{it}, L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}\}$, and the Gross Output with FOC Materials is given by $M_{it}^{\beta_m} L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}$, where the material coefficient is computed using the median of the material revenue share ($\beta_m = Median[\frac{P_t^M M_{it}}{P_t Q_{it}}]$).

The Leontief production function shows a similar pattern of increasing capital coefficients when moving from the control to the Control IV approach. The capital coefficient increases from 0.25 to 0.31 for China, 0.33 to 0.44 for India, and 0.22 to 0.29 for Chile.

As for the labor coefficient, in general we find that the cases where the capital coefficient increases, the labor coefficient offsets this increase. For instance, for the Gross Output production function, the control function labor coefficient decreases from 0.12 to 0.10 when we use our IV approach in China, 0.27 to 0.19 for India, and 0.38 to 0.21 for Chile. Likewise, the labor coefficient in the IV estimates are lower than the OLS ones, falling from 0.12 to 0.03 for China, 0.25 to 0.16 for India, and 0.43 to 0.19 for Chile respectively for the case of the Gross Output production function. Note as well that the MonteCarlo experiments were less dispositive as to the effect of capital measurement error on the labor coefficient.

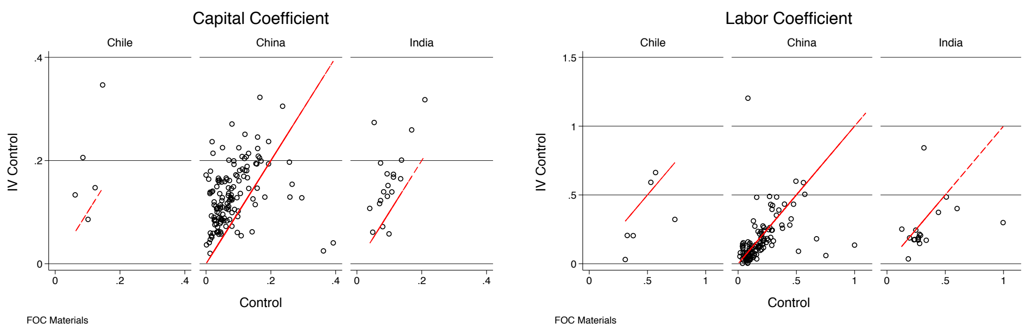
As mentioned before, the presence of measurement error in the capital stock is expected to affect estimates of returns to scale. This is because, in general, all output elasticities are potentially biased through the correlation of every input with capital, thereby injecting a signal-to-noise bias. When we sum the labor and capital coefficients under our procedure, we find that returns to scale parameters tend to be higher. A notable case is the estimated returns to scale in the Chinese manufacturing sector. Brandt, Van Biesebroeck, and Zhang (2012) document that deploying standard production function estimators leads to surprisingly low returns to scale parameters which in turn imply high productivity growth estimates. We first of all confirm that finding, but importantly, the estimated returns to scale go up considerably when using the IV-Control approach. Depending on the specification for the production function, we find an increase of 0.05 or 0.19 in the estimated scale parameter, for the gross output and Leontief specification, respectively, suggesting constant rather than decreasing returns to scale.

Finally, in Figure 3 we plot the industry-specific capital and labor coefficients by country, with the top panel showing the Gross Output production function, and the bottom panel showing the Leontief. The vertical axis shows the capital coefficient from our IV estimator, while the horizontal axis shows the capital coefficient that does not instrument with investment, and the red solid line is the 45 degree line.

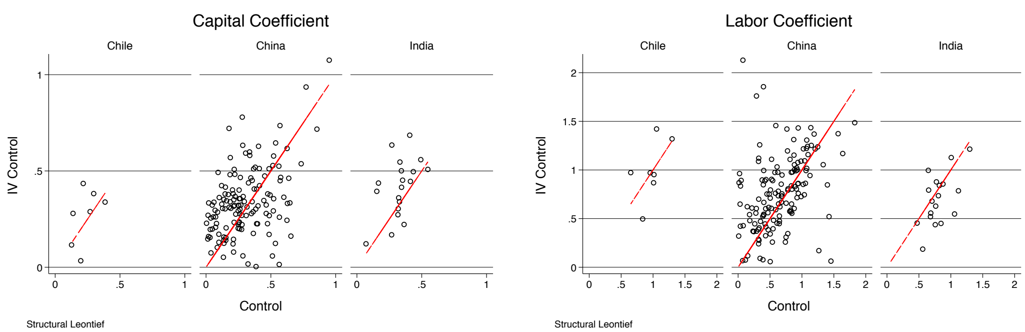
The bulk of the observations for the capital coefficient are above the 45-degree line (in red), indicating that our IV estimates are higher than the non-IV estimates for capital. This is particularly true for the Gross Output production function. Note as well, that there is nothing mechanical about our higher capital coefficients: there are several industries where we find a lower capital coefficient using the IV control approach versus the control function. The labor coefficients are also a bit lower, but this pattern is less striking. These differences are also frequently statistically significant. We have

Figure 3: Industry-specific Coefficients

(a) Gross Output



(b) Leontief



Notes: Each observation is a two-digit industry (as classified by the respective national industry classification), and we plot the capital coefficient obtained from our procedure (i.e., IV) against the alternative control function. The red line is the 45-degree line.

computed the standard errors for the Control and Control IV estimators using a bootstrap procedure, and report these in Table E.1 in the Appendix. We test for the capital coefficient of the Control IV estimator to be either larger or smaller than the Control estimator, which we do at the 90 percent level. For the Gross Output production function, 17 percent of industries have Control IV capital coefficients that are significantly higher, while none of the industries have Control IV capital coefficients that are significantly smaller. For the Leontief production function, we find 15 percent of the industries with significantly larger Control IV capital coefficients, and only 4 with significantly smaller coefficients.

Taking these results at face value suggests that measurement error in capital leads to obtaining capital coefficients that are significantly lower — i.e., half the magnitude. This has important consequences for any subsequent productivity analysis, as we discuss in the next section.

6 Measuring Productivity Premia

In this section we quantify the implications of the substantial bias in the estimated capital coefficient. We focus on the productivity premium of three distinct firm characteristics: differences in technology in India and the WBES, the role of private ownership in the Chinese economy, and firm size in all of our datasets.

Throughout, we compute productivity based on the gross output specification, where the material coefficient is estimated using the first order condition approach (using the material input revenue share – denoted $\hat{\beta}_m$). In addition we impose a constant returns to scale production function, such that $\beta_m + \beta_l + \beta_k = 1$, which allows us to focus on a single parameter of the production function β_k. We focus on constant returns to scale since these are commonly assumed in the literature and allow us to abstract from estimating returns to scale. That being said, in our datasets, we generally find higher returns to scale using our Control IV estimator, and in some cases, increasing returns to scale. Under this structure, we can write the productivity estimate as a function of the estimated capital coefficient:

$$\tilde{\omega}_{it} = q_{it} - \hat{\beta}_m m_{it} - \beta_k k_{it} - (1 - \hat{\beta}_m - \beta_k) l_{it}, \tag{24}$$

where $\tilde{\omega}_{it} \equiv \omega_{it}(\beta_k) + \epsilon_{it} + \beta_k \epsilon_{it}^k$. When we run regressions using $\tilde{\omega}$ as a dependent variable, this additive error will not introduce a bias, which is not the case when we compute correlations.

In all what follows, we compute a productivity premium based on a regression of productivity on the relevant characteristic, including full industry and year fixed effects. Our main goal is to illustrate how a large bias in estimates of the production function translate into economically meaningful differences in productivity.

6.1 Technology Premia

We first revisit the impact of technology on productivity with two specific examples. The first case is Allcott, Collard-Wexler, and O'Connell (2016)'s investigation of the effect of power outages on productivity and output in India. The effect of reducing power outages is mitigated for plants that own generators, since these plants do not need to shut down during outages. The second case looks at the productivity effects of information technology, such as discussed in Atrostic and Nguyen (2005); Brynjolfsson and Hitt (2003) and, a very closely related paper using the WBES by Cusolito, Lederman, and Pena (2021).

Panel (a) of Figure 4 shows the premium for electric generators on the left panel, and the premia for the firm having a webpage on the right panel. In each of these figures we compute the premium as a function of the capital coefficient β_k. We also plot the Control and Control IV estimate of β_k for the relevant dataset from Section 5.[33] For electric generators, the Control estimator computes a productivity premia of 7 percent above non-generators, while the Control IV estimator computes a productivity premia of -2 percent. To understand why increasing the capital coefficient from $\beta_k = 0.09$ for the Control estimate, to $\beta_k = 0.19$ for the Control IV estimate, has this effect on productivity premia, it is useful to look at the relative capital to labor ratio for electric generators versus non-generators. This ratio is 80 percent higher for electricity generators, which is unsurprising, as firms with greater capital stock have larger incentives to purchase generators.

Panel (b) of Figure 4 shows the productivity premia for firms that have a website. Again, more capital intensive firms are more likely to have a website. This leads to a predicted productivity premia of 0.34 for the Control estimator, versus 0.23 for the Control IV estimator.

Now, clearly, there are endogeneity issues in the interpretation of these productivity premia: firms do not buy generators or build websites arbitrarily. We are merely showing how the correlation between technology and productivity changes when we use different estimators. Any causal analysis will take these correlation as inputs.

In the case of generators, the productivity premium of firms that own a generator over those that do not disappears once we use our Control IV estimator. Likewise, there is a 50 percent drop in the productivity premium of firms that own websites when we move from the Control to the Control IV estimator. In both applications, the presence of measurement error in capital leads to a severe overestimate of the importance of technology in explaining productivity differences between firms.

6.2 Ownership and Productivity Premium

There is growing interest into the effects of management on productivity, for instance, the recent work of Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013), and relatedly the ownership of these firms, such as the impact of foreign direct investment. An important example of this management effect can be found in the large performance differences between Chinese firms that are either state-owned (henceforth SOE), versus private or even foreign, such as in Chen, Igami, Sawada, and Xiao (2021).

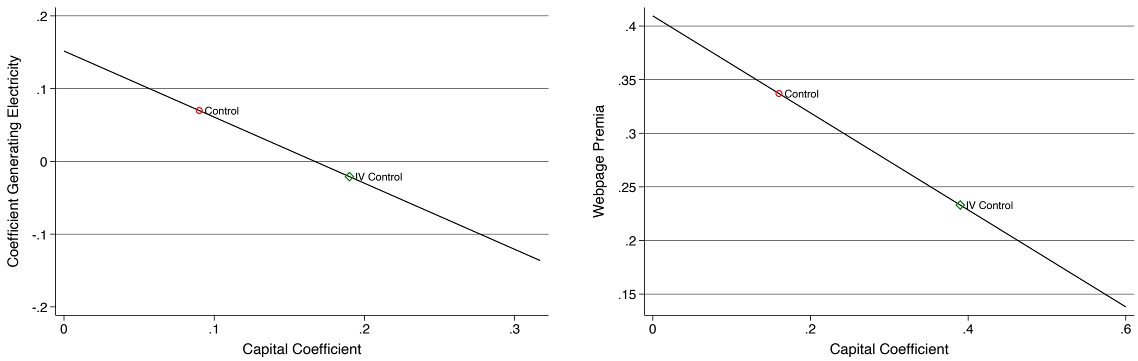
Figure 5 shows the productivity difference between state owned-firms, either firms that are privately owned in China, or foreign-owned firms, again as a function of the capital coefficient β_k. An interesting

[33] We present the average coefficient across all industries since this allows us to abstract from the distribution of estimated parameters across different industries, but are indicative of impacts of biased capital coefficients on productivity premia.

Figure 4: Technology Productivity Premia

(a) Premia for Electric Generators

(b) Webpage Premia



Notes: The data in panel (a) is drawn from the ASI in India, where electric generators are identified from plants with positive power production. The data in panel (b) is from the WBES, where firms are asked if they have a website. Control and Control IV present the average capital coefficient estimated in each dataset.

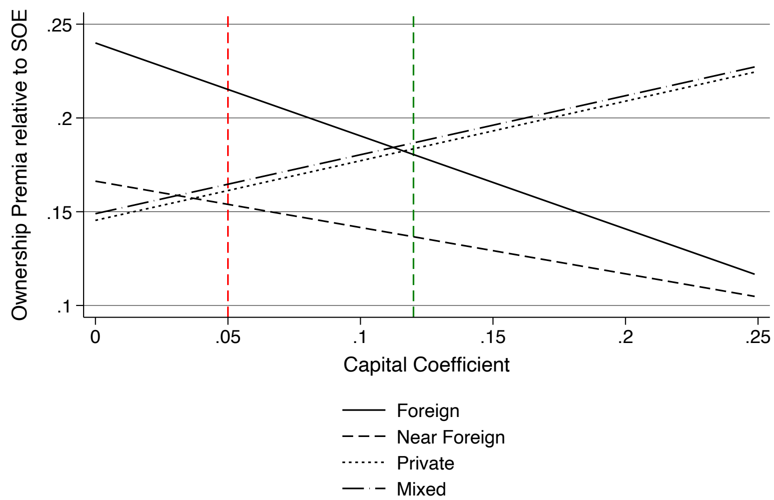
difference is that foreign firms are more capital intensive than state-owned firms (as measured by the capital to labor ratio), while private Chinese firms are less capital intensive than their state-owned rivals. This means that the higher the capital coefficient, the less productive private and mixed firms look, and the more productive foreign firms look.[34] Indeed, the productivity premia of foreign firms relative to SOE is 0.22 using the Control estimator versus 0.18 for the Control IV estimator. Conversely, the productivity premia rises from 0.16 in the Control estimator to 0.18 in the Control IV estimator when comparing private firms to SOEs. Among private firms, we find that the FDI premium, that is the productivity difference between foreign private firms and Chinese private firms, shrinks from 5.5 % to 0.0%; it vanishes, when we correct for measurement error in capital.

6.3 Size Productivity Premium

Finally, a large body of work has leveraged the alignment between firm size and productivity to understand the allocative efficiency of an economy. In short, the more inputs are assigned to high productivity firms, the higher is the aggregate productivity of an economy. This type of analysis is found in the work of Olley and Pakes (1996); Bartelsman, Haltiwanger, and Scarpetta (2013); Hsieh and Klenow (2009) among others, and different metrics are used in each of these papers. However, all

[34]In the Chinese context there is considerable evidence that private and SOE firms have very different access to capital that could be driving the large difference in capital intensity.

Figure 5: Productivity Premia for different ownership in China



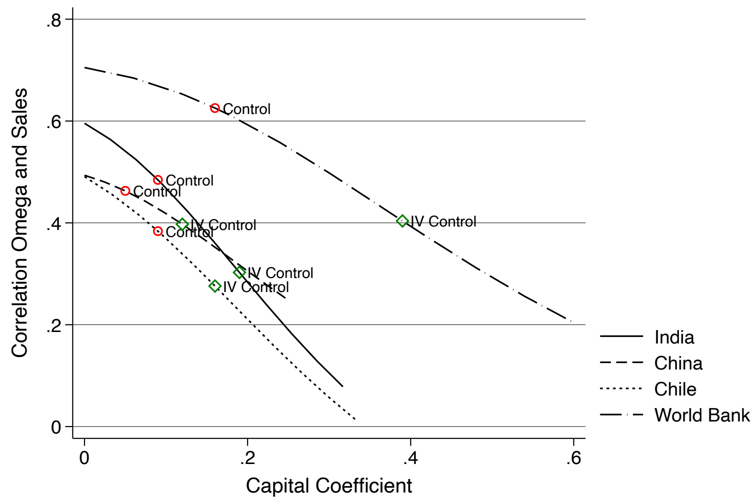
Notes: Productivity Premia plotted relative to State-Owned Firms. Control estimate plotted in red at $\beta_k = 0.05$, while Control IV investment plotted in green at $\beta_k = 0.12$. Near Foreign denotes firms whose owners are in Hong Kong, Macau, and Taiwan.

these metrics have in common the idea that larger firms ought to be more productive. Therefore, we will compute a simple statistic: the correlation between productivity and size as measured by total sales.

Figure 6 shows the correlation between productivity and sales plotted against the capital coefficient β_k for our four datasets, China, India, Chile, and the WBES. As before, we have plotted the average estimated capital coefficient for each dataset, for both the Control and the IV Control estimators. For India, the control estimator of β_k computes a correlation of 48 percent, falling to 30 percent with the Control IV estimator. Similar effects are found for China (46 to 39 percent), Chile (38 to 28 percent), and the World Bank (63 to 40 percent). The reason for this effect is simple: large firms tend to be more capital intensive.

Interpreting the level of the correlation between output and productivity is complex, since even in an economy where inputs are distributed at random, firms with higher productivity will produce more output. However, this is distinct from what we do here, evaluating how the capital coefficient changes the correlation between size and productivity. In addition, we are merely showing that a class of allocative efficiency metrics are substantially affected by the capital coefficient one has estimated,

Figure 6: Size Productivity Premia and the Capital Coefficient



Notes: The range plotted for each dataset corresponds to the range of estimated capital coefficients. Control and IV Control present the average capital coefficient estimated in each dataset.

even when restricting attention to constant returns to scale production functions. Most worryingly, countries and industries where capital measurement is the most problematic will have the largest bias in capital coefficients, which leads to overestimates of the allocative efficiency.[35]

7 Concluding remarks

This paper revisits the estimation of production functions in the presence of measurement error in capital. Our starting point is that appropriately measuring capital is one of the most difficult tasks that go into estimating a production function. There is, however, rather surprisingly little work that deals directly with measurement error in capital, or any input for that matter.

Introducing an estimator that relies on a hybrid IV-control function approach, we build on what have now become standard techniques to address the simultaneity bias, and add an IV strategy to correct for the measurement error of capital. We propose either the replacement value of capital

[35]Relatedly, Rotemberg and White (2021) show that differences in the treatment of outliers across countries, which is a very specific form of the measurement error considered in this paper, leads to dramatically different conclusions on allocative efficiency metrics across different economies.

or lagged investment as an instrument for the capital stock, while still controlling for the standard simultaneity bias. The latter is crucial to capture the well-known simultaneity concern that arises due to he presence of unobserved productivity shocks. Our approach therefore captures leading estimators in the literature that have come to be widely used throughout various subfields of economics.

From an applied point, our estimator is as easy to implement as the standard methods used in the literature, including control function and dynamic panel techniques, and allows for a great deal of flexibility in incorporating recent advances in the estimation of production function in the context of market imperfections, both in output and input markets, the presence of adjustment frictions for factors, and endogenous productivity processes.

Monte Carlo simulations show that our estimator performs well, even in cases of rather large measurement error. We apply our estimator to Indian, Chilean and Chinese producer-level data, along with cross-country data from the World Bank data. We estimate capital coefficients that are double those obtained with standard techniques. This indicates that correcting for measurement error in capital can be a first-order concern, and it has immediate implications for the literature that studies productivity dynamics, firm growth, investment, productivity dispersion, and the covariates of productivity growth. We illustrate this impact using distinct settings, including technology upgrading, ownership and the size premium, and we find substantially lower productivity premia in all cases when correcting for measurement error in capital. Any analysis where the conclusions rest on estimates of the marginal product of capital, by itself, or as an input into measurement of productivity, can be misleading in the presence of errors in the recording and measurement of the capital stock.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2005): "Econometric tools for analyzing market outcomes," *Handbook of Econometrics*, 6.

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification properties of recent production function estimators," *Econometrica*, 83(6), 2411–2451.

Allcott, H., A. Collard-Wexler, and S. D. O'Connell (2016): "How do electricity shortages affect industry? Evidence from India," *The American Economic Review*, 106(3), 587–624.

Asker, J., A. Collard-Wexler, and J. De Loecker (2014): "Dynamic inputs and resource (mis) allocation," *Journal of Political Economy*, 122(5), 1013–1063.

Atrostic, B. K., and S. V. Nguyen (2005): "IT and productivity in US manufacturing: do computer networks matter?," *Economic Inquiry*, 43(3), 493–506.

Baily, M. N., C. Hulten, and D. Campbell (1992): "Productivity dynamics in manufacturing plants," *Brookings Papers on Economic Activity. Microeconomics*, 1992(23), 187–267.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013): "Cross-country differences in productivity: The role of allocation and selection," *The American Economic Review*, 103(1), 305–34.

Bartelsman, E. J., and M. Doms (2000): "Understanding productivity: Lessons from longitudinal microdata," *Journal of Economic Literature*, 38(3), 569–594.

Becker, R., J. Haltiwanger, R. Jarmin, S. D. Klimek, and D. Wilson (2006): "Micro and macro data integration: The case of capital," in *A new architecture for the US national accounts*, pp. 541–610. University of Chicago Press.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 128(1), 1–51.

Bond, S., and M. Söderbom (2005): "Adjustment costs and the identification of Cobb Douglas production functions," Discussion paper, IFS Working Papers, Institute for Fiscal Studies (IFS).

Braguinsky, S., A. Ohyama, T. Okazaki, and C. Syverson (2015): "Acquisitions, productivity, and profitability: evidence from the Japanese cotton spinning industry," *The American Economic Review*, 105(7).

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012): “Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing,” *Journal of Development Economics*, 97(2), 339–351.

Brynjolfsson, E., and L. M. Hitt (2003): “Computing productivity: Firm-level evidence,” *Review of Economics and Statistics*, 85(4), 793–808.

Chen, Y., M. Igami, M. Sawada, and M. Xiao (2021): “Privatization and productivity in china,” *RAND Journal of Economics*.

Collard-Wexler, A., and J. De Loecker (2015): “Reallocation and technology: Evidence from the US steel industry,” *The American Economic Review*, 105(1), 131–71.

Cusolito, A. P., D. Lederman, and J. Pena (2021): “The effects of digital-technology adoption on productivity and factor demand: Firm-level evidence from developing countries,” Discussion paper, World Bank.

De Loecker, J. (2007): “Do exports generate higher productivity? Evidence from Slovenia,” *Journal of International Economics*, 73(1), 69–98.

De Loecker, J. (2013): “Detecting learning by exporting,” *American Economic Journal: Microeconomics*, 5(3).

De Loecker, J., J. Eeckhout, and G. Unger (2020): “The rise of market power and the macroeconomic implications,” *The Quarterly Journal of Economics*, 135(2), 561–644.

De Loecker, J., and P. Goldberg (2014): “Firm Performance in a global market,” *Annual Review of Economics*, 6, 201–227.

De Loecker, J., P. Goldberg, A. Khandelwal, and N. Pavcnik (2016): “Prices, markups and trade reform,” *Econometrica*, 84(2), 445–510.

De Loecker, J., and C. Syverson (2021): “An Industrial Organization perspective on productivity,” Discussion paper, National Bureau of Economic Research.

De Loecker, J., and F. Warzynski (2012): “Markups and firm-level export status,” *The American Economic Review*, 102(6), 2437–2471.

Foster, L., J. Haltiwanger, and C. Syverson (2008): “Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?,” *The American Economic Review*, 98(1), 394–425.

Galuscak, K., and L. Lizal (2011): "The impact of capital measurement error correction on firm-level production function estimation," Discussion paper, Czech National Bank.

Gavazza, A. (2011): "Leasing and secondary markets: Theory and evidence from commercial aircraft," *Journal of Political Economy*, 119(2), 325–377.

Gorodnichenko, Y., and M. D. Shapiro (2011): "Using the survey of plant capacity to measure capital utilization," Discussion Paper CES-WP-11-19, US Census Bureau Center for Economic Studies.

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*, ed. by S. Strom, pp. 169–203. Cambridge University Press.

Hausman, J. (2001): "Mismeasured variables in econometric analysis: problems from the right and problems from the left," *Journal of Economic Perspectives*, 15(4), 57–67.

Hsieh, C.-T. (2002): "What explains the industrial revolution in east asia? Evidence from the factor markets," *The American Economic Review*, 92(3), 502–526.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

Hu, Y., G. Huang, and Y. Sasaki (2020): "Estimating production functions with robustness against errors in the proxy variables," *Journal of Econometrics*, 215(2), 375–398.

Kalouptsidi, M. (2014): "Time to build and fluctuations in bulk shipping," *The American Economic Review*, 104(2), 564–608.

Kim, K., A. Petrin, and S. Song (2016): "Estimating production functions with control functions when capital is measured with error," *Journal of Econometrics*, 190(2), 267–279.

Levinsohn, J., and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, 70(2), 317–341.

Maddala, G. S., and K. Lahiri (1992): *Introduction to econometrics*, vol. 2. Macmillan New York.

Olley, G. S., and A. Pakes (1996): "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 64(6), 35.

Pavcnik, N. (2002): "Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants," *Review of Economic Studies*, pp. 245–276.

Pischke, S. (2007): "Lecture notes on measurement error," Working Paper London School of Economics.

Rotemberg, M., and T. K. White (2021): "Plant-to-Table(s and Figures): Processed Manufacturing Data and Measured Misallocation," Discussion paper, New York University.

Syverson, C. (2001): "Output market segmentation, heterogeneity, and productivity," Ph.D. thesis, University of Maryland.

——— (2011): "What determines productivity?," *Journal of Economic Literature*, 49(2), 326–65.

Tybout, J. R. (1992): "Making noisy data sing: Estimating production technologies in developing countries," *Journal of Econometrics*, 53(1-3), 25–44.

Van Biesebroeck, J. (2007): "Robustness of productivity estimates," *Journal of Industrial Economics*, 3(55), 529–539.

Young, A. (1995): "The tyranny of numbers: confronting the statistical realities of the East Asian growth experience," *The Quarterly Journal of Economics*, 110(3), 641–680.

A Factor shares and Alternative Environments

We present a more detailed discussion on the use of the factor share approach to identify and estimate the material coefficient in the baseline setup of our framework. Departures from perfect competitive output and input markets are briefly discussed and we offer an illustration under a specific model of product differentiation under common input prices, and highlight that the main features of our approach are robust to considering such an environment.

A.1 Factor shares

Under the standard DGP of perfect competition it is well-known that the output elasticity of a variable input in production can immediately be measured in the data using the *revenue share* of an input X:

$$\beta_X = \frac{(P^X X)_{it}}{(P_t \tilde{Q})_{it}}. \tag{A.1}$$

where $P^X X$ is the expenditure on the input X, and $P\tilde{Q}$ is observed revenue, including the measurement error in output ($\tilde{Q}_t = Q_{it} \exp(\epsilon_{it})$). In the case of the Cobb-Douglas production function (either Gross Output if we wish to measure the coefficient on materials, or Cobb-Douglas in labor and capital, if we wish to measure the coefficient on labor), we have to confront the substantial dispersion in the revenue share in (any) dataset. Under the assumed setup this dispersion can only come from measurement error in output, and therefore we need to rely on an estimator that is robust to this.

The presence of measurement error in output then calls for a simple estimator:

$$\hat{\beta_X} = Median\left[\frac{(P^X X)_{it}}{(P\tilde{Q})_{it}}\right], \tag{A.2}$$

where the median is used instead of the mean, due to the error in the denominator of this expression. This estimator is a consistent estimator of β_X if we assume that the output measurement error satisfies $Med[\epsilon_{it} = 0]$. This is obtained by letting the median operator go through the observed revenue share:

$$\frac{(P^X X)_{it}}{P_t Q_{it}} \frac{1}{\exp(Median[\epsilon_{it}])}, \tag{A.3}$$

where we use the property of the exponential being a monotone function. Thus, our estimator proposed in equation (A.2) is a consistent estimator of β_X. Likewise, we can estimate the material coefficient β_m under the Gross Output production function, and under Leontief the estimator considers the median of the output share $(\frac{M_{it}}{\tilde{Q}_{it}})$.

Under the Gross Output specification, a special case presents itself whereby only the capital coefficient is required to be estimated using GMM. Using the estimates $\hat{\beta}_l$ and $\hat{\beta}_m$, we can net out the contribution of these inputs to output: $y_{it} = q_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_m m_{it}$. We can then proceed with our estimation routine relying on y_{it} as the measure of output (instead of q_{it} in the first-stage). This discussion outlines how control function methods can be combined with factor-share approaches in our method. However, the factor share approaches crucially rely on the assumptions of flexible input choice and competitive output and input markets.

A.2 Alternative environments

In most settings, we observe firms charging different prices for their output, and paying different prices for inputs, which leads to an additional complication since researchers typically have access to only (deflated) revenues and expenditures on inputs. We believe this to be a very important concern. The starting point of our analysis, however, is to assume we have correctly converted the revenue and expenditure data to the comparable units in a physical sense, and this is precisely the setup of Ackerberg, Caves, and Frazer (2015).

A related but distinct concern is the presence of imperfect competitive output and input markets. As discussed in the review article by De Loecker and Goldberg (2014), both observations lead to two challenges: unobserved output and input price variation plagues identification, and second, the control for unobserved productivity needs to be adjusted to accommodate the departures from perfect competition and identical prices across firms. The approach suggested in this paper is thus valid under commonly assumed models of imperfect competition, either when adopting the Leontief technology, or in the case of the Gross Output production function by extending the input demand equation as in De Loecker and Warzynski (2012) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). We discuss briefly, and illustrate it with a few leading cases, how our approach can be adjusted to incorporate departures from the standard setup in the main text of the paper.

Leontief As discussed in the main text, under the Leontief technology the presence of imperfect competition does not affect the control. The reason is simple: regardless of input price differences, or imperfect competition in the product market, as long as firms cost minimize and produce at the optimal point where $F(L,K)\Omega = \beta_m M$, the control is valid.

Gross Output Under this technology, the control function approach is adopted and alternative environments (such as input price differences and imperfectly competitive product and factor markets) require additional terms to be included in the control to reflect the dependance of optimal input demand on these output and input market conditions. In general, let the vector z_{it} captures these factors, then $\omega_{it} = h(m_{it}, k_{it}, l_{it}, z_{it})$. A few leading cases that are of interest include input price variation, demand-side heterogeneity (e.g. export demand, product-level demand differences), and quality differences. See De Loecker and Goldberg (2014) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a detailed discussion.

The FOC approach to estimating the material coefficient is, however, no longer valid under imperfect competitive output markets, and therefore to estimate the material coefficient β_m under the Gross Output specification, we can no longer rely on the factor share. This is precisely the approach of De Loecker and Warzynski (2012) to estimate markups: comparing the revenue share to the output elasticity of a variable input generates a measure of the markup. We therefore introduce an additional moment condition to identify the material coefficient – i.e., $\mathbb{E}(\tilde{\xi}_{it} m_{it-1}) = 0$. This requires additional structure to the problem by relying on observed, and serially correlated material prices, that are to be included in the control for productivity. Under Leontief the material coefficient remains as before.

Finally, our approach, like ACF, can also handle different timing assumptions on the choice of inputs. For instance, in the case in which labor faces adjustment costs, the labor input now constitutes a state variable, and labor choices will not entirely react to productivity innovation shocks ξ_{it}. In the case of one-period hiring, labor at time t and labor at $t-1$ are exogenous variables. The moments condition is then given by: $\mathbb{E}\left[\tilde{\xi}_{it}\left(\ l_{it}\ \right)\right] = 0$ Thus, the source of identification for the labor coefficient is then precisely that adjustment costs vary across firms, to the extent that these vary with the labor stock (see Bond and Söderbom (2005) for a discussion).

B Monte Carlo: details

In this section, we describe details of the Monte Carlo that we will use to evaluate the performance of our estimator.

Table B.1: Monte Carlo Parameters

Data Size		
Number of Firms	N= 1,000	Taken from ACF.
Time Periods	T= 10	
Production Function Parameters		
Capital Coefficient	$\beta_k = 0.4$	
Labor Coefficient	$\beta_l = 0.6$	
Depreciation Rate	$\delta = 0.2$	
Productivity Process	ρ_ω=0.7, $\sigma_\omega = 0.3$	
Wage Process	ρ_w=0.3, $\sigma_w = 0.1$	
Cost Capital ϕ	ρ_ϕ=0.9, $\sigma_\phi = 0.3$	Dispersion of Log Capital of 1.6 Autocorrelation of Capital of 0.93
Measurement Error Parameters		
Error in Capital	$\rho^k = 0.7$, $\sigma_k = 0.2$	High Persistence, and 30 percent measurement error in stationary distribution
Error in Output	$\rho^y = 0.2$, $\sigma_y = 0.3$	Low Persistence, and 30 percent measurement error in stationary distribution

We specify laws of motion for each of the variables in the data-generating process.

B.1 Timing

First, we specify the timing assumptions in our model. Investment is chosen with a one period time to build. Materials are chosen statically — i.e., after the firm knows its productivity Ω_{it}. Labor is chosen statically in for DGP 2 and DGP 3, and in an interim period for DGP 1 — i.e., part of the productivity shock is revealed before the firm makes its labor choice.

Second, there are four exogenous state variables, productivity Ω_{it}, wages W_{it}, output prices P_{it}, and the price of capital ϕ_{it}, which all have log AR(1) processes. The only endogenous state variable is capital.

Logged productivity Ω has a first-order Markov evolution:

$$\omega_{it} = \rho^\omega \omega_{it-1} + u^\omega_{it}, \tag{B.1}$$

where $u^\omega \sim \mathcal{N}(0, \sigma_\omega^2)$.

In addition, log wages have a first-order Markov process:

$$w_{it} = \rho^w w_{it-1} + u_{it}^w, \tag{B.2}$$

and likewise for the logged price for output (P):

$$p_{it} = \rho^p p_{it-1} + u_{it}^p, \tag{B.3}$$

where $u^w \sim \mathcal{N}(0, \sigma_w^2)$ and $u^p \sim \mathcal{N}(0, \sigma_p^2)$. For the purposes of the Monte Carlo, we will normalize $p_{it} \equiv 1$, the case of perfect competition.

B.2 Derivation of Investment Policy as in Syverson (2001)

In this section, we derive a closed form for the investment function in Syverson (2001), to show that we can allow a time-varying cost of capital ϕ_{it}. This derivation is very close to Syverson (2001), so our goal is merely to show that this model admits a time-varying cost of capital ϕ_{it}.

Firms have flow profits given by:

$$\Pi_{it} = P_{it}\Omega_{it}L_{it}^{\alpha}K_{it}^{1-\alpha} - W_{it}L_{it} - \frac{\phi_{it}}{2}I_{it}^2, \tag{B.4}$$

where P is the price of output, A is physical productivity, W refers to firm specific wages, and I is investment.

The firm's value function V is given by:

$$\begin{aligned} V(P_{it}, \Omega_{it}, K_{it}, W_{it}, \phi_{it}) = \underset{L_{it},K_{it}}{max}\, & P_{it}\Omega_{it}L_{it}^{\alpha}K_{it}^{1-\alpha} - W_{it}L_{it} \\ & + \beta\mathbb{E}_{it}V(P_{it+1}, \Omega_{it+1}, K_{it+1}, W_{it+1}, \phi_{it+1}) \\ \text{such that } & K_{it+1} = (1-\delta)K_{it} + I_{it} \end{aligned} \tag{B.5}$$

where δ is the depreciation rate of capital.

Labor is chosen using the usual first-order condition $\frac{\partial \Pi_{it}}{\partial L_{it}} = 0$:

$$\begin{aligned} P_{it}\Omega_{it}\alpha L_{it}^{\alpha-1}K_{it}^{1-\alpha} &= W_{it} \\ \rightarrow L_{it} &= \left[\frac{\alpha P_{it}\Omega_{it}}{W_{it}}\right]^{\frac{1}{1-\alpha}} K_{it} \end{aligned} \tag{B.6}$$

And, likewise, investment solves the Euler Equation, $\frac{\partial V}{\partial I} = 0$, giving,

$$\phi_{it}I_{it} = \beta\mathbb{E}_{it}V_K(P_{it+1}, \Omega_{it+1}, K_{it+1}, W_{it+1}, \phi_{it+1}). \tag{B.7}$$

The envelope condition yields:

$$\begin{aligned} V_K(P_{it}, \Omega_{it}, K_{it}, W_{it}, \phi_{it}) =& (1-\alpha)P_{it}\Omega_{it}L_{it}^{\alpha}K_{it}^{-\alpha} \\ & + (1-\delta)\mathbb{E}_{it}V_K(P_{it+1}, \Omega_{it+1}, K_{it+1}, W_{it+1}, \phi_{it+1}). \end{aligned} \tag{B.8}$$

Substituting into the first-order conditions, the envelope condition becomes

$$\phi_{it} I_{it} = \beta \mathbb{E}_{it}\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+1}^{-\frac{\alpha}{1-\alpha}} P_{it+1}^{\frac{1}{1-\alpha}} \Omega_{it+1}^{\frac{1}{1-\alpha}}\right] + \beta(1-\delta)\mathbb{E}_{it}\phi_{it+1} I_{it+1}. \tag{B.9}$$

And then iterating this equation forward — i.e., replacing $\mathbb{E}_{it}\phi_{it}I_{it}$ with the right-hand side in equation (B.9) — yields:

$$\begin{aligned}
\phi_{it} I_{it} =& \beta \mathbb{E}_{it}\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+1}^{-\frac{\alpha}{1-\alpha}} P_{it+1}^{\frac{1}{1-\alpha}} \Omega_{it+1}^{\frac{1}{1-\alpha}}\right] \\
&+ \beta(1-\delta)\mathbb{E}_{it}\beta\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+2}^{-\frac{\alpha}{1-\alpha}} P_{it+2}^{\frac{1}{1-\alpha}} \Omega_{it+2}^{\frac{1}{1-\alpha}}\right] \\
&+ \left[\beta(1-\delta)\right]^2 \mathbb{E}_{t+1}\phi_{it+2} I_{it+2} \\
\phi_{it} I_{it} =& \beta \mathbb{E}_{it}\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+1}^{-\frac{\alpha}{1-\alpha}} P_{it+1}^{\frac{1}{1-\alpha}} \Omega_{it+1}^{\frac{1}{1-\alpha}}\right] \\
&+ \beta(1-\delta)\mathbb{E}_{it}\beta\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+2}^{-\frac{\alpha}{1-\alpha}} P_{it+2}^{\frac{1}{1-\alpha}} \Omega_{it+2}^{\frac{1}{1-\alpha}}\right] \\
&+ \left[\beta(1-\delta)\right]^2 \beta\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} W_{it+3}^{-\frac{\alpha}{1-\alpha}} P_{it+3}^{\frac{1}{1-\alpha}} \Omega_{it+3}^{\frac{1}{1-\alpha}}\right] \\
&+ \left[\beta(1-\delta)\right]^3 \mathbb{E}_{t+2}\phi_{it+3} I_{it+3}.
\end{aligned} \tag{B.10}$$

Writing in the form of geometric series

$$I_{it} = \frac{\beta(1-\alpha)}{\phi_{it}}\alpha^{\frac{\alpha}{1-\alpha}}\mathbb{E}_{it}\sum_{j=0}^{\infty}\left\{[\beta(1-\delta)]^j W_{it+1+j}^{-\frac{\alpha}{1-\alpha}} P_{it+1+j}^{\frac{1}{1-\alpha}} \Omega_{it+1+j}^{\frac{1}{1-\alpha}}\right\} \tag{B.11}$$

Given that we assume that P_{it}, Ω_{it} and W_{it} follow the log-linear AR(1) process with normal error terms, the investment function becomes:

$$\begin{aligned}
I_{it} =& \frac{\beta(1-\alpha)}{\phi_{it}}\alpha^{\frac{\alpha}{1-\alpha}}\mathbb{E}_{it}\sum_{j=0}^{\infty}\{[\beta(1-\delta)]^j W_{it}^{-\frac{\alpha\phi_w^{i+1}}{1-\alpha}} \prod_{s=0}^{j}(u_{t+1+i-s}^w)^{-\frac{\alpha\phi_w^s}{1-\alpha}} P_{it}^{\frac{\phi_p^{i+1}}{1-\alpha}} \\
&\cdot \prod_{s=0}^{j}(u_{t+1+i-s}^p)^{\frac{\phi_p^s}{1-\alpha}} \Omega_{it}^{\frac{\phi_a^{i+1}}{1-\alpha}} \prod_{s=0}^{j}(u_{t+1+i-s}^a)^{\frac{\phi_a^s}{1-\alpha}}\} \\
=& \frac{\beta(1-\alpha)}{\phi_{it}}\alpha^{\frac{\alpha}{1-\alpha}} \prod_{s=0}^{j}\{[\beta(1-\delta)]^j W_{it}^{-\frac{\alpha\phi_w^{i+1}}{1-\alpha}} \prod_{s=0}^{j}\mathbb{E}_{it}(u_{t+1+i-s}^w)^{-\frac{\alpha\phi_w^s}{1-\alpha}} P_{it}^{\frac{\phi_p^{i+1}}{1-\alpha}} \\
&\cdot \prod_{s=0}^{j}\mathbb{E}_{it}[(u_{t+1+i-s}^p)^{\frac{\phi_p^s}{1-\alpha}}]\Omega_{it}^{\frac{\phi_a^{i+1}}{1-\alpha}} \prod_{s=0}^{j}\mathbb{E}_{it}[(u_{t+1+i-s}^a)^{\frac{\phi_a^s}{1-\alpha}}]\}
\end{aligned}$$

Since for $\epsilon \sim \mathcal{N}(0,\sigma^2)$, we have $\mathbb{E}(u^{\frac{\phi^s}{1-\alpha}}) = \exp(\frac{\sigma^2\phi^{2s}}{2(1-\alpha)^2})$; then, the investment function can be further simplified as:

$$I_{it} = \frac{\beta(1-\alpha)}{\phi_{it}} \alpha^{\frac{\alpha}{1-\alpha}} \mathbb{E}_{it} \sum_{i=0}^{\infty} \{[\beta(1-\delta)]^j W_{it}^{-\frac{\alpha\phi_w^{i+1}}{1-\alpha}} \prod_{s=0}^{j} \exp\left(\frac{\alpha^2\sigma_w^2\phi_w^{2s}}{2(1-\alpha)^2}\right) P_{it}^{\frac{\phi_p^{i+1}}{1-\alpha}} \cdot \prod_{s=0}^{j} \exp\left(\frac{\sigma_p^2\phi_p^{2s}}{2(1-\alpha)^2}\right) \Omega_{it}^{\frac{\phi_a^{i+1}}{1-\alpha}} \prod_{s=0}^{j} \exp\left(\frac{\sigma_a^2\phi_a^{2s}}{2(1-\alpha)^2}\right)\}. \tag{B.12}$$

B.3 Process for the Price of Capital

In the original Monte Carlo proposed by Ackerberg, Caves, and Frazer (2015), the authors extend the model proposed by Syverson (2001) by allowing for the price of capital ϕ to differ between firms — i.e., to allow the price for capital to be a firm-specific ϕ_i. This is important, in the context of their Monte Carlo, since it allows a higher cross-sectional dispersion of capital between firms than that generated by reasonable processes of productivity, given the patterns in standard producer-level data.

In this paper, we also need capital to move around more than it does in Ackerberg, Caves, and Frazer (2015), with a process that generates a serial correlation coefficient for capital of 0.99. Instead, we have the following AR(1) process:

$$\phi_{it} = \rho^{\phi}\phi_{it-1} + \sigma^{\phi}u_{it}^{\phi}, \tag{B.13}$$

where $u_{it}^{\phi} \sim \mathcal{N}(0, 1)$.

B.4 Alternative Data-Generating Processes

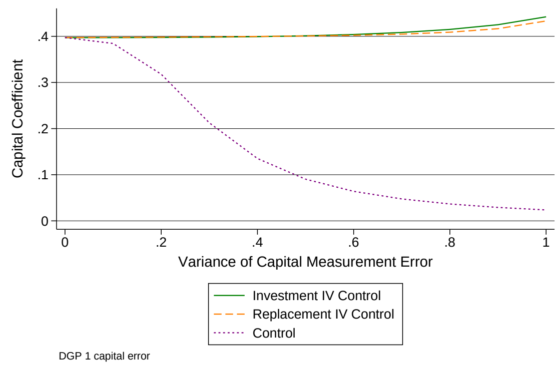
We evaluate the performance of our estimator in two alternative data-generating processes (DGPs), as considered in ACF in their Monte Carlos.

- DGP 1:
  DGP 1 is the case considered in the main Monte Carlos in the paper. Note that labor is chosen a half-period before materials are picked. More precisely, labor is chosen at time $t - 0.5$, and materials are chosen at time t, where the productivity process is adjusted so that the stochastic process for $\omega_{it-0.5}$ is given by:

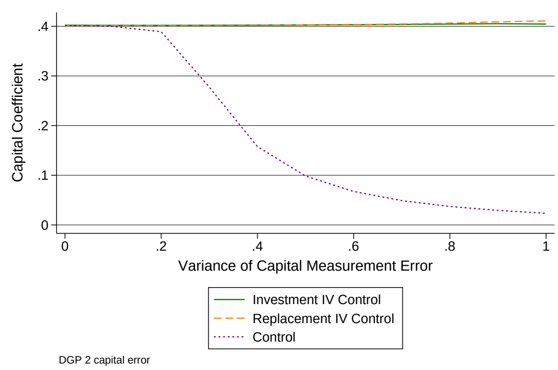
  $$\omega_{it} = \rho^{0.5}\omega_{it-0.5} + \xi_{it}^{b},$$

  where ξ_{it}^{b} is an appropriately adjusted normally distributed shock.
- DGP 2:
  DGP2 refers to the case of optimization error in labor. The variance of the wage distribution is shut down, $\sigma_w = 0$, but instead, firms face an optimization error in labor. Thus, $l_{it} = l_{it}^{*} + \epsilon_{it}^{l}$ where $\epsilon_{it}^{l} \sim \mathcal{N}(0, 0.37)$.
- DGP 3:
  DGP 3, has the same process as DGP 1, but adds in optimization error in labor,$l_{it} = l_{it}^{*} + \epsilon_{it}^{l}$, where $\epsilon_{it}^{l} \sim \mathcal{N}(0, 0.37)$, as in DGP 2.

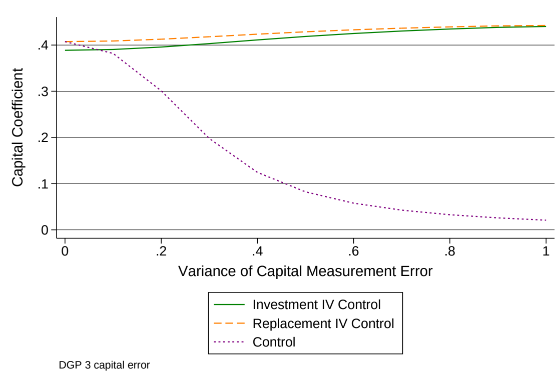
Figure B.1, below replicates Figure 2 for DGP 1, 2 and 3, and shows the sensitivity of our IV and non-IV estimators to the measurement error in capital. Notice that the pattern that we document in the Monte Carlos for DGP 1 in the main paper is the same as what we find for these alternative DGPs: our estimator performs well with varying degrees of capital measurement error, while the standard control function approaches are biased for reasonable amounts of measurement error.



(a) DGP 1



(b) DGP 2



(c) DGP 3

Note: We plot the estimated capital coefficient as a function of the standard deviation in the capital measurement error (σ_k^2). Average of 100 Monte Carlo replications per value of σ_k. The true value of $\beta_k = 0.4$.

Figure B.1: Relationship between β_k and Measurement error σ_k in Capital for different DGPs

C Different Processes for Measurement Error in Capital

In our main specifications in this paper, we rely on the process for measurement error having the form:

$$k_{it} = k^*_{it} + \epsilon^k_{it}, \tag{C.1}$$

where ϵ^k_{it} is a mean zero measurement error, which may be serially correlated. While this is the standard formulation for an *errors-in-variables* structure, we contrast this approach to, what we refer to as, a structural derived measurement error for capital. After we introduce this setup, we perform an analogous Monte Carlo analysis to evaluate our estimator, and we make sure that both approaches are directly comparable through their implied variance in the measurement error of capital.

As discussed in Section 2, we consider the main source of the capital measurement error to stem from errors in depreciation $D_{it} = \delta_{it} K_{it-1}$, where the correct measure is given by $d_{it} = \delta^*_{it} K^*_{it-1}$. Applying the same law of motion for capital as before, it is easy to show that the measured capital stock under the error in depreciation rates is given by:

$$K_{it} = \sum_{\tau=0}^{t} I^*_{i\tau} - \sum_{\tau=0}^{t} D_{i\tau-1}, \tag{C.2}$$

where we keep the assumption that we observe investment without error and, for simplicity, also the initial capital stock.[36] The source of measurement error is, thus, from the cumulative depreciation errors, and we capture as follows: $\delta_{it} = \delta^* + \epsilon^d_{it}$. Both the reduced-form and the structural approach generate a wedge between the measured and true capital stock, in levels K and K^*, respectively. After some algebra, we have a direct mapping between the structural measurement error, ϵ^d_{it}, and the reduced form measurement ϵ^k_{it}:

$$\epsilon^d_{it} = \frac{K^*_{it}}{K^*_0} \exp{(\epsilon^k_{it})}. \tag{C.3}$$

This relationship is important when comparing the performance of our estimator across both Monte Carlos: a much smaller standard deviation in the depreciation error, ϵ^d_{it}, is needed to generate a certain standard deviation of the classical measurement error, ϵ^k_{it}.[37]

C.1 Monte Carlo Analysis

All parameters of the Monte Carlo are the same as before, except that we parameterize $\epsilon^d_{it} \sim \mathcal{N}(0, \sigma_d)$. True depreciation is given by $D^*_{it} = \delta K^*_{it-1}$. However, measured depreciation is given by $D_{it} = (\delta + \epsilon^d_{it}) K_{it}$.

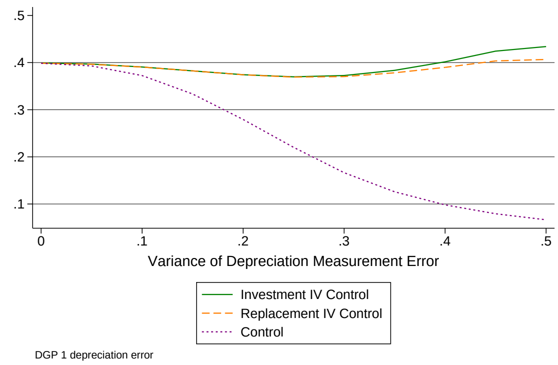
Figure C.1 presents the results of this Monte Carlo for both the IV Control method we propose and the Control method that does not use investment as an instrument. We plot the relationship between the mean estimate of β_k over 100 replications and σ_d.

Figure C.1 shows that under a small amount of measurement error in depreciation — say, on the order of a 0.01 standard deviation shock to a depreciation rate of 0.1 — the IV control function method we have proposed does fairly well, with mean estimates around the true value of 0.4. In contrast, the mean estimates that do not instrument with investment, show a drop of β_k to 0.3, with a 0.01 standard deviation shock to the depreciation rate, in line with the previous results that we showed illustrating that measurement error in capital stock leads to downward bias on the capital coefficient.

[36] This is without loss of generality for the purpose of this Appendix.

[37] We have traced this out in our Monte Carlo analysis, and we obtain about a 20-1 ratio — i.e., we obtain similar implied standard deviations for the measurement error using either $\sigma^k = 0.2$ or $\sigma^d = 0.01$.

Notice that Figure C.1 does not show that our IV Control function estimator is consistent for any value of measurement error in depreciation. Indeed, the process for measurement error in depreciation does not lead to the log additive error structure on capital that we need for estimation. Instead, our goal is merely to point out that our estimator might perform well for alternative measurement error structures, at least for some small deviations from our structure.



Note: We plot the estimated capital coefficient as a function of the standard deviation in the capital measurement error (σ_k^2). Average of 100 Monte Carlo replications per value of σ_d. The true value of $\beta_k = 0.4$.

Figure C.1: Relationship between β_k and Measurement error σ_d in Depreciation

D Data Appendix

We apply our estimator to four datasets, covering manufacturing plants in China, India and Chile, and a large cross-country dataset from the World Bank. There have been numerous productivity studies using these data, and, therefore, are completely standard in which variables are reported, and how they are constructed.

D.1 World Bank data

We use data across dozens of countries from the World Bank Enterprise Survey. We use the programs and procedure outlined in Cusolito, Lederman, and Pena (2021) to process the data. Table D.1 presents summary statistics for the data from the World Bank, where we note that all values have been converted to 2009 U.S. dollars. The primary feature of this dataset is that the survey is not designed to be a panel, so many observations are lost when one conditions on having observations in two adjacent periods. The World Bank data also has a survey item on the replacement value of capital that is unusual for producer level data. This variable is defined as the answer to the following survey question: *Cost For Establishment To Re-Purchase All Of Its Machinery, Land And Buildings.*

We note that the rank correlation between the value of asset versus the repurchase cost of assets is 0.74, indicating substantial, but not perfect, overlap between these measures.

Table D.1: Summary Statistics for World Bank Enterprise Data

Variable	**Mean**	**Std. Dev.**	**N**
Sales	13.4	2.3	22787
Materials	12.1	2.6	23115
Capital	11.4	2.7	23115
Labor	3.6	1.4	23115
Replacement Value Capital	11.9	2.6	20538
Investment	10.5	2.6	12138
Lagged Investment	10.7	2.2	1366
Lagged Replacement Value Capital	12.3	2.4	1937
Lagged Capital	11.7	2.5	2193

D.2 Chilean manufacturing

Annual plant-level data on all manufacturing plants with at least ten workers were provided by Chile's Instituto Nacional de Estadistíca (INE). These data, which cover the period 1979-1986, include production, employment, investment, intermediate input, and balance-sheet variables. Industries are classified according to the four-digit ISIC industry code. Output and input price indices are constructed at the three digit industry and obtained directly from average price indices produced by the Central Bank of Chile. We directly observe total number of employees, total real value of production, total real intermediate input, total real book-value of fixed assets, gross additions to capital (i.e. investment). In total there are 37,600 plant-year observations reporting employment, with a minimum of 4,205 plants in 1983 and 5,814 plants in 1979. Following Levinsohn and Petrin (2003) we rely on the law of motion for capital where investment is productive within the year, reflecting the accounting standards of reported additions to capital stock (i.e. $K_{it} = (1-\delta)K_{it-1} + I_{it}$). See Pavcnik (2002), Levinsohn and Petrin (2003) and Asker, Collard-Wexler, and De Loecker (2014) for productivity studies using these data.

D.3 Chinese manufacturing

The Chinese manufacturing data is described in detail in Brandt, Van Biesebroeck, and Zhang (2012). This is firm-level data for the period 1998–2007 that are the product of annual surveys conducted by the National Bureau of Statistics (NBS). The survey includes all industrial firms that are either state-owned, or are non-state firms with sales above 5 million RMB. There are plant-level identifiers that allow us to put this together into a panel.

There are 1,557,915 plant by year observations for 346,434 unique plants. For plants observed multiple times, only about 1 percent have gaps between yearly observations.

D.4 Indian manufacturing

We use India's Annual Survey of Industries (ASI) for establishment-level microdata; this dataset is described in more detail in Allcott, Collard-Wexler, and O'Connell (2016). Registered factories with over 100 workers (the "census scheme") are surveyed every year, while smaller establishments (the "sample scheme") are typically surveyed every three to five years. The publicly available ASI includes establishment identifiers that are consistent across years beginning in 1998, but we have plant identifiers going back to 1992. We have a plant-level panel for the entire 1992-2010 sample.

The ASI is comparable to manufacturing surveys in the United States and other countries. Variables include revenues, value of fixed capital stock, total workers employed, total costs of labor, and materials. Industries are grouped using India's NIC (National Industrial Classification) codes, which are closely related to SIC (Standard Industrial Classification) codes.

There are 615,721 plant-by-year observations at 224,684 unique plants. 107,032 plants will be immediately dropped from our estimators because they are observed only once. For plants observed multiple times, 60 percent of intervals between observations are one year, while 91 percent of intervals are less than five years.

The mean (median) plant employs 79 (34) people and has gross revenues of 139 million (20 million) Rupees, or in U.S. dollars approximately \$3 million (\$400,000).

E Standard Errors Across Estimators

Table E.1: Standard errors

Leontief

	China		India		Chile	
	Capital	Labor	Capital	Labor	Capital	Labor
OLS	0.02	0.03	0.01	0.01	0.03	0.05
FE	0.03	0.04	0.01	0.02	0.04	0.06
IV	0.05	0.05	0.02	0.03	0.09	0.13
Control	0.13	0.23	0.05	0.11	0.05	0.13
IV Control	0.12	0.24	0.09	0.19	0.20	0.45

Gross Output with FOC Materials

	China		India		Chile	
	Capital	Labor	Capital	Labor	Capital	Labor
OLS	0.01	0.01	0.00	0.01	0.01	0.03
FE	0.01	0.02	0.01	0.01	0.04	0.04
IV	0.02	0.02	0.01	0.01	0.05	0.07
Control	0.13	0.23	0.05	0.11	0.05	0.13
IV Control	0.12	0.24	0.09	0.19	0.20	0.45

Notes: We report the median standard error across all industries for each dataset. The Leontief specification is $Q_{it} = \min\{\beta_m M_{it}, L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}\}$, and the Gross Output with FOC Materials production function is given by $M_{it}^{\beta_m} L_{it}^{\beta_l} K_{it}^{\beta_k} \Omega_{it}$, where the material coefficient is computed using the median of the material revenue share ($\beta_m = Median[\frac{P_t^M M_{it}}{P_t Q_{it}}]$).

F Estimators: Code and Implementation

In this section we describe the estimators proposed in this paper in pseudo-code, with enough detail so that they can easily be coded up by other researchers. Code for these estimators is also available in STATA from the authors.

F.1 Leontief

1. Estimate Φ function.

$$q_{it} = \theta_0 + \theta_l l_{it} + \theta_k k_{it} + \theta_m m_{it} + \epsilon_{it}$$

by 2SLS using exogenous regressors $W = [l_{it}, z_{it}, m_{it}]$, and obtain $\hat{\Phi}_{it} = \hat{\theta}_0 + \hat{\theta}_l l_{it} + \hat{\theta}_k k_{it} + \hat{\theta}_m m_{it}$. Note that the control function (not IV) has $W_{it} = [l_{it}, k_{it}, m_{it}]$ instead — the OLS estimator.

2. For a parameter $\boldsymbol{\beta}$, minimize the criterion $\mathcal{Q}(\boldsymbol{\beta})$ using:

   (a) Compute $\omega_{it} = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it}$

   (b) Estimate the AR(1) process for productivity, $\omega_{it} = \rho_0 + \rho\omega_{it-1} + \tilde{\xi}_{it}$, by OLS, obtain $\hat{\rho}$. Recover productivity shock $\tilde{\xi}_{it} = \omega_{it} - \hat{\rho}\omega_{it-1}$ – where tilde refers to the presence of the measurement error in capital.

   (c) Compute $\mathcal{Q}(\boldsymbol{\beta})$ as the empirical analogue of the moment condition $\mathbb{E}\left[\tilde{\xi}_{it}\begin{pmatrix} l_{it-1} \\ z_{it} \end{pmatrix}\right] = 0$.

$$\mathcal{Q}(\boldsymbol{\beta}) = (\boldsymbol{\xi z})'(\boldsymbol{z}'\boldsymbol{z})^{-1}(\boldsymbol{\xi z}),$$

   where $\boldsymbol{\xi}$ denotes the stacked vector of $\tilde{\xi}_{it}$, and $\boldsymbol{z}$ denotes the stacked vector of $[l_{it-1}, z_{it}]$. Note that the control function (not IV) has $\boldsymbol{z} = [l_{it-1}, k_{it}]$ instead.

   (d) Find $\hat{\boldsymbol{\beta}}$ as the minimizer of $\mathcal{Q}(\boldsymbol{\beta})$.

F.2 Gross Output (FOC Materials)

1. Estimate β_m

$$\hat{\beta}_m = Median\left(\frac{P^M M_{it}}{P\tilde{Q}_{it}}\right).$$

2. Produce output y_{it} netted out from material contribution.

$$y_{it} = q_{it} - \hat{\beta}_m m_{it}$$

3. Estimate Φ function.

$$y_{it} = \theta_0 + \theta_l l_{it} + \theta_k k_{it} + \theta_m m_{it} + \epsilon_{it}$$

by 2SLS using exogenous regressors $W = [l_{it}, z_{it}, m_{it}]$, and obtain $\hat{\Phi}_{it} = \hat{\theta}_0 + \hat{\theta}_l l_{it} + \hat{\theta}_k k_{it} + \hat{\theta}_m m_{it}$.

4. For a parameter $\boldsymbol{\beta}$, minimize the criterion $\mathcal{Q}(\boldsymbol{\beta})$ using:

(a) Compute $\omega_{it} = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it}$

(b) Estimate the AR(1) process for productivity, $\omega_{it} = \rho_0 + \rho\omega_{it-1} + \tilde{\xi}_{it}$, by OLS, obtain $\hat{\rho}$. Recover productivity shock $\xi_{it} = \omega_{it} - \hat{\rho}\omega_{it-1}$.

(c) Compute $\mathcal{Q}(\boldsymbol{\beta})$ as the empirical analogue of the moment condition $\mathbb{E}\left[\tilde{\xi}_{it}\begin{pmatrix} l_{it-1} \\ z_{it} \end{pmatrix}\right] = 0$.

$$\mathcal{Q}(\boldsymbol{\beta}) = (\boldsymbol{\xi}\boldsymbol{z})'(\boldsymbol{z}'\boldsymbol{z})^{-1}(\boldsymbol{\xi}\boldsymbol{z}),$$

where $\boldsymbol{\xi}$ denotes the stacked vector of $\tilde{\xi}_{it}$, and $\boldsymbol{z}$ denotes the stacked vector of $[l_{it-1}, z_{it}]$.

(d) Find $\hat{\boldsymbol{\beta}}$ as the minimizer of $\mathcal{Q}(\boldsymbol{\beta})$.

Instruments We use two different instruments (z_{it}), either we rely on lagged investment (i_{it-1}) or the replacement value of capital (k_{it}^R).