Production, Investment and Wealth Dynamics under Financial Frictions: An Empirical Investigation of the Self-financing Channel*

Alvaro Aguirre
Central Bank of Chile

Matias Tapia
Central Bank of Chile

Lucciano Villacorta
Central Bank of Chile

October 2021

Abstract

We develop a new empirical framework to assess the self-financing channel. Using administrative data, we estimate firm-level productivity and its effect on firms' decisions. Our framework is robust to financial frictions, while we characterize and quantify the biases of standard methods (e.g., Olley-Pakes). We uncover a distribution of investment and wealth accumulation propensities to productivity shocks. These propensities are heterogeneous in the stock of wealth and productivity level: (i) investment propensities are larger for high-wealth firms, and (ii) propensities for wealth accumulation are larger for high-productivity and low-wealth firms. Our estimates support the existence of self-financing, but its impact is limited.

JEL classification: C33, E23, O11, L0
Keywords: Production function estimation; productivity process; Olley-Pakes; panel data; latent variables, financial frictions; wealth dynamics; investment; self-financing.

*We are grateful to Daniel Ackerberg, Manuel Arellano, Stephane Bonhomme, Paco Buera, Andrea Caggese, Emmanuel Farhi, Ivan Fernandez-Val, Manuel Garcia Santana, Benjamin Moll, Diego Perez, Josep Pijoan-Mas, Ludwig Straub, Chad Syverson, Alonso Villacorta, Gianluca Violante, Fabrizio Zilibotti and attendees at seminars at the 26th International Panel Data Conference, the 2020 World Congress of the Econometric Society, European Economic Association 2020, Santiago Macroeconomic Workshop 2020, LACEA-LAMES 2019, PUC Chile, Central Bank of Chile, and U Chile for their comments, and Diego Huerta and Cristian Valencia for excellent research assistance. The views expressed in this paper are exclusively those of the authors and do not necessarily reflect the position of the Central Bank of Chile or its Board members. E-mail: aaguirre@bcentral.cl, mtapia@bcentral.cl, lvillacorta@bcentral.cl

1 Introduction

A large body of literature has studied the potential role of financial frictions in explaining cross-country differences in aggregate income, investment, and productivity. Among other consequences, incomplete access to external financing can prevent productive firms with low levels of wealth from operating at their optimal scale. This can lead to an inefficient allocation of factors that lowers aggregate productivity. However, a strand of the literature emphasizes the ability of firms to overcome financial frictions by accumulating wealth and build collateral after persistent productivity shocks [see Buera et al., 2015, for a review]. This endogenous response, known as the self-financing channel, has the potential to mitigate the aggregate adverse effects of financial frictions. Importantly, the self-financing channel is reflected in the propensities of firms' investment and wealth accumulation to productivity shocks and how these propensities depend on the amount of available collateral. Moreover, the scope of the mitigating effect of the self-financing channel might depend on the persistence and volatility of the firm-level productivity process.[1]

Accordingly, a precise empirical assessment of the wealth, the investment and productivity processes is essential to understand the quantitative effects of financial frictions. However, an analysis of these objects using micro data is currently absent from the literature. Crucially, standard approaches to estimate firm-level productivity and production function are invalid in the presence of financial frictions. This paper explores empirically the strength of the self-financing channel by developing a novel empirical framework that is robust to the presence of financial constraints in order to jointly estimate the firm's production function, the firm-productivity process, and the wealth accumulation and investment processes. We implement this new estimation method using data on manufacturing firms obtained from the census of the administrative records of formal firms in Chile from 2006 to 2016. Besides including data on inputs and output at the firm level, a novelty of our database is that it provides balance sheet information (measures of firms' wealth) for both public and private firms. Using our framework, we characterize the firm's relevant policy functions for investment and wealth accumulation decisions, and use these functions to identify crucial objects—such as the response of investment and wealth accumulation to productivity shocks. These objects provide empirical evidence of the presence of financial frictions and the self-financing channel and help discipline quantitative macro models of firm dynamics with financial frictions.

One of the main contributions of this paper is to develop a tractable econometric framework that enables us to uncover features of the firm's productivity process and assess how wealth accumulation and investment decisions respond to productivity shocks in the data without a structural estimation. In contrast to fully-specified structural approaches, which require the specification of functional forms for preferences and financial frictions, we model our empirical policy rules non-parametrically, leaving functional forms unrestricted.[2] Although we do not estimate a full structural model, our empirical

[1]For instance, Buera and Shin [2011] and Moll [2014] focus on how the degree of persistence of productivity shocks affects the strength of the self-financing channel.

[2]As emphasized by Buera et al. [2021]: "macro models have tended to rely on strong structural assumptions, e.g., assumptions on functional forms and distributions of unobservables, and on somewhat stylized calibration strategies, and

specification and identification strategy are motivated by the economic insights of heterogeneous-agent models with financial frictions, in the spirit of Buera and Shin [2011], Buera et al. [2011], Moll [2014] and Midrigan and Xu [2014]. In our empirical policy rules, the marginal effect of productivity is allowed to be nonlinear and heterogeneous across firms and contingent on the level of collateral, enabling us to reveal a distribution of micro-level investment and savings propensities in response to productivity shocks. Consequently, our modeling approach can provide a rich picture of the joint relationship between investment decisions, wealth accumulation and productivity shocks drawn directly from the data.

An important challenge to our framework is that firm productivity is an unobservable variable. There is a large body of literature devoted to the estimation of firm production functions and, consequently, measures of productivity at the firm level [Syverson, 2011]. Prevalent methods rely on a proxy variable approach to recover productivity using the firm's input decisions [see Ackerberg et al., 2015, for a review]. For instance, in their seminal contribution, Olley and Pakes [1996] (OP hereafter), recover productivity by inverting an investment demand function, which is then used as a nonparametric control in the production function regression. An additional contribution of our paper is to show neatly how these methods fail when financial frictions are present, as they deliver biased estimates. In response to this problem, we propose a novel empirical strategy that is robust to the presence of financial constraints, by exploiting the theoretical insights behind the self-financing channel, and jointly estimating the production function, productivity process, and investment and wealth accumulation processes.

1.1 An overview of our methodology

Our novel empirical framework consists of a firm production function, a non-linear firm investment policy rule, and a non-linear firm wealth accumulation policy rule. These three equations depend on the latent firm-level productivity process. As in the proxy variable framework, we assume that productivity follows a flexible non-linear Markovian process of order one. Our framework has *three main departures* from the proxy variable approach initiated by OP. First, to control for financial constraints, we include the firm's stock of wealth as an additional state variable in the investment equation, in accordance with the insights of theoretical models with financial frictions in which wealth is pledged as collateral. This is precisely the main reason why OP fails, since, in the context of these models, investment variability cannot be completely explained by productivity and initial capital. Intuitively, the OP method assigns differences in investment across firms in the data to differences in unobserved productivity. However, under financing constraints, differences in investment between firms are not only the result of productivity gaps, but also might be driven by differences in access to credit.[3] Second, besides the investment equation, we jointly estimate the wealth accumulation

thus economists often view it as disconnected from micro empirical research".

[3]It is worth noting that the failure of OP in the presence of omitted variables in the investment policy function is well known [see for instance Shenoy, 2020]. Our contribution in this regard is to clearly describe the consequences of this failure and to propose a new empirical framework that can solve this issue–a framework that is conceptually consistent with the widely used macroeconomic models with financial frictions.

policy function. This function is our main focus, as it plays a fundamental role in understanding the scope and implications of the self-financing channel. Moreover, under financial frictions, the behavior of firms' wealth may be more informative about unobserved productivity dynamics than the evolution of investment, since the collateral of constrained firms responds more than their investment to productivity shocks. Third, in order to provide more reliable empirical estimates of these two policy functions, we allow for the existence of unobservable shocks in the policy rules, in addition to the latent productivity shock.[4]

Identification and estimation of our nonlinear model cannot be handled within the proxy variable framework, since our nonlinear policy rules are more flexible and include unobservable shocks in addition to the latent productivity process. Furthermore, a key aspect of our model is to identify and estimate the non-linear policy functions as relevant objects of economic interest in themselves. We show that nonparametric identification of the production function, the productivity process and the policy functions of our model can be attained, following recent developments in nonlinear panel data models with latent variables [Hu and Schennach, 2008, Hu and Shum, 2012, Arellano and Bonhomme, 2017, Arellano et al., 2017].

From an instrumental variable perspective, both the wealth accumulation policy rule and the investment policy rule can be thought of as noisy measures of unobserved productivity. If conditional independence holds, such that the production function and both policy rules are independent conditional on productivity and observed state variables, the wealth policy rule can be used as an instrument for investment (the noisy measure of productivity) in the production function regression implied by OP. Intuitively, due to the self-financing channel, a positive co-movement between investment decisions and wealth accumulation decisions is related to changes in productivity that can be used to identify the parameters of the production function.

We show that for parsimonious, yet flexible, versions of the policy functions, an IV estimation strategy within the proxy variable framework delivers consistent estimates of the model, following the arguments in the nonparametric identification strategy. For more general policy functions, we consider a tractable estimation strategy that is well-suited to non-linear panel data models with latent variables by adapting the approach in Arellano et al. [2017] to a production function setup. The estimation approach is a stochastic expectation-maximization (EM) algorithm that combines simulation methods and GMM estimation. A key aspect of our method is that it combines the production function with the information on investment decisions and wealth accumulation dynamics to construct the posterior distribution of the productivity process. An important advantage of our empirical methodology based on nonlinear reduced form models (compared to a full structural estimation) is econometric transparency in the sense of Andrews et al. [2017], Andrews et al. [2020] and Bonhomme [2020]. First, we formally discuss identification and clearly show how the conditional independence assumption and the Markovian assumption–justified by the economic insights of structural models with financial frictions–enable us to construct dynamic restrictions that are used to identify the nonlinear reduced-form model,

[4] This is in contrast to the proxy variable approach, which assumes that the policy rules are deterministic functions of productivity and observables.

despite the presence of latent productivity. Second, our IV estimator is transparent, as it directly connects our estimates to the relevant moments and variation in the data that "drive" the estimator (see the discussion in Andrews et al. [2020]). Although the empirical model cannot provide direct policy counterfactuals, its estimated parameters may be used directly or indirectly to calibrate structural models that are able to do so. For example, our production function and productivity estimates can be used to directly parametrize the firm's production function and the productivity process in a structural model, while our empirical policy rules can be used as matching targets for other key parameters related to preferences, adjustment costs to capital and financial constraints.[5]

1.2 Results

Our non-linear framework uncovers new empirical results for both the production function literature and the macro literature with financial frictions. Regarding the production function estimates, the results show that the estimated average effect of capital in the production function increases from 0.35 when using OP to 0.43 when we consider financial frictions in the estimation. By contrast, the estimated marginal effect of labor in the production function decreases from 0.65 in OP to 0.44 when controlling for financial frictions. Using a firm dynamic model with collateral constraints, we show analytically the source of the biases associated with the OP estimator.[6] Intuitively, financial constraints generate differences in investment, capital and output between equally productive firms. The OP approach interprets differences in observed investment across firms as differences in unobserved productivity. Even when the implied variation in output is driven by variations in capital, the OP approach assigns such variation to variations in productivity, as implied by the proxy equation. As a result, the OP productivity proxy captures an important part of the effect of capital on output, underestimating the marginal effect of capital. If financial frictions are less severe in the labor market, the labor coefficient is upwardly biased, as OP interprets a financially constrained firm with low investment as a low-productivity firm that hires "too many" workers and produces "too much" output relative to its proxy-OP productivity. Hence, it will assign a large role to labor in the determination of output, overestimating the labor elasticity. Furthermore, the differences in the estimates of factor elasticities translate into significant differences in the measure of returns to scale. In particular, OP results are consistent with constant returns to scale, whereas our estimates imply decreasing returns to scale with a span of control of 0.87–consistent with the standard calibration of quantitative macroeconomic models. We also use an extended version of the model to generate simulated data,

[5] In the spirit of Nakamura and Steinsson [2018], our empirical policy functions provide "identified moments" such as the average causal effect of productivity on investment and on wealth accumulation that are useful for estimating parameters of structural models or discriminating between structural models.

[6] Although we focus our attention on constraints on investment in capital, our conceptual argument is more general, as financial frictions could also be present in expenditures on intermediate inputs or labor, generating a similar effect. Therefore, the same biases could emerge in alternative methodologies, as in Levinsohn and Petrin [2003], in the sense that the true policy functions for intermediate inputs or labor would also depend on a measure of the firm's wealth/assets. The extent of the bias under those methodologies is an empirical issue, as it would depend on the relative importance of credit constraints in the hiring of different inputs.

confirming our theoretical insights and empirical results.

In terms of the firm-level productivity process, we show that OP significantly underestimates both the dispersion and the persistence of productivities relative to our approach. The 90th to 10th ratio of the firm productivity distribution with our methodology is twice as large as with OP in any given year. Moreover, the standard deviation of productivity under OP is 0.16, which increases to 0.42 when we control for financial frictions. These results are also consistent with the theoretical implications of financial frictions: Relatively productive firms, which are expected to be more financially constrained according to the canonical model, show larger investment gaps with respect to their optimal levels. This leads OP to underestimate these firms' productivity relative to unproductive firms and to shrink the estimated productivity distribution. Regarding persistence, the first-order autoregressive estimated parameter grows from 0.56 in OP to 0.82 in our model. Given that the OP estimated productivity is a combination of the true productivity process and financial constraints, the underestimation of persistence by OP may suggest that constraints are less persistent thanks to self-financing.

The literature on production function estimations uses policy rules as auxiliary equations to control for unobserved productivity. However, these policy functions themselves have not been an object of interest in this literature. By contrast, we pay special attention to the estimated policy functions because they are key to understanding the role of financial frictions and the self-financing channel.

The estimated investment policy function enables us to assess the transmission of firm productivity shocks to investment decisions and document how sensitive this transmission is to financial frictions. Our estimated investment propensities in response to income shocks suggest that financial frictions depend nonlinearly on wealth, capital and productivity. For all capital levels, the marginal effect of productivity on investment is monotonically increasing in wealth. Propensities at the lowest levels of wealth are significantly lower than those for firms with the highest wealth, suggesting that collateral constraints play an important role. For instance, for highly leveraged firms, the elasticity of investment to productivity shocks more than doubles when we move from the bottom to the top of the wealth distribution. Moreover, in line with earning-based constraints models, the investment propensity in response to productivity shocks is also heterogenous in the firm productivity level with a larger propensity for more productive firms. However, the relationship between investment propensity to productivity shocks and initial productivity also interacts with the initial stock of capital. For instance, for very highly productive firms but with low capital, the investment propensity is at its highest value and less sensitive to the amount of collateral. However, the investment propensity is very sensitive to the amount of collateral for firms with low capital and productivity below the median of the productivity distribution. By contrast, for firms with a stock of capital above the median of the capital distribution, the propensity is increasing in wealth for all productivity values. In fact, for highly leveraged firms, the investment propensity in response to productivity shocks is at its lowest for firms at the bottom of the wealth distribution, independent of their initial level of productivity. For this firms, the propensity dramatically increases as we move along the wealth distribution.

The estimated wealth accumulation policy shows that there is a significant and positive effect of productivity shocks on future wealth, which suggests that the self-financing channel is active in the data. Interestingly, we show that the effect of productivity on wealth accumulation is heterogeneous in the stock of wealth. For highly productive but constrained firms at the bottom end of the wealth distribution, the elasticity of productivity on wealth accumulation is close to 1. Thus, for very productive but constrained firms, the transmission of persistent income shocks to savings is almost complete. This response weakens significantly as we move upwards along the wealth distribution. This result is consistent with the economic mechanisms driving the self-financing channel in models with financial frictions: Low wealth firms, which are more constrained, have higher incentives to save in order to self-finance future investments when they experience positive and persistent productivity shocks.

Finally, in order to asses the strength of self-financing we follow Banerjee and Moll [2010], and use our estimated empirical model to compute the convergence time of the marginal product of capital between two firms with the same productivity that start with different levels of wealth. On the one hand, we find evidence of self-financing in the data as we show that the MPK of these firms converge over time. On the other hand, this channel is not too powerful. For instance, when we compare firms at the 10th-percentile with firms at the 90th-percentile of the wealth distribution, the MPK of poor firms is around three times the MPK of rich firms at the initial period and it takes more than 40 years to see convergence in their MPKs. Still, half of the initial gap in their MPKs disappears after ten years.

1.3 Related literature

Our paper makes contributions to three different streams of literatures. First, it connects with the empirical literature that estimates production functions at the firm level using the proxy variable approach [Olley and Pakes, 1996, Levinsohn and Petrin, 2003, Ackerberg et al., 2015, Doraszelski and Jaumandreu, 2013, 2018, Gandhi et al., 2020, Shenoy, 2020]. Among these papers, ours is closest to Olley and Pakes [1996] and Ackerberg et al. [2015], which studied value-added production functions using the investment equation as the proxy variable. We build on these papers to develop a framework that is robust to financial frictions. Our paper differs from these papers in several aspects. First, our paper studies the biases that appear when the proxy variable approach is used to estimate the production function and the productivity process under the environment of macro models with collateral constraints. Second, our paper uses the insights and economic mechanisms presented in those models to propose a novel strategy that is robust to financial frictions. In this sense, our paper is the first paper that uses the self-financing channel to identify the firm productivity process and the firm production function. In terms of the methodology, we allow for more flexible policy rules including transitory shocks, unlike the proxy variable approach. We propose a new sequential identification scheme that leads to two novel estimators that jointly exploit the information in the investment and the wealth accumulation policy rules. Finally, an important difference of our framework is the identification and estimation of the investment and wealth accumulation policy rules, one of the main contributions of this paper.[7] Our empirical framework shares the spirit of the empirical consumption-

[7] The papers that use the proxy variable approach use (only) the investment policy rule as an auxiliary equation to control for unobserved productivity in the production function regression. However, this policy function has not been an

household income framework [e.g. Blundell et al., 2008, Kaplan and Violante, 2010, Arellano et al., 2017, Straub, 2019], which exploits panel data to estimate the degrees to which consumption decisions respond to unobserved household income shocks, but applied to a firm setup. A crucial econometric difference between these frameworks lies in the estimation of the net income process. In the household framework, income shocks and their effect on consumption are extracted directly from the household income data after removing demographic characteristics that are assumed to be orthogonal to the income shocks. By contrast, to estimate the unobserved firm productivity process and its effect on investment and savings, we need to estimate the production function parameters where the regressors are endogenous and correlated with unobserved productivity.

Second, our paper connects to the macro-finance literature that studies the aggregate effects of financial frictions. We are closer to the set of papers focusing on collateral constraints and the self-financing channel [e.g. Buera and Shin, 2011, Buera et al., 2011, Song et al., 2011, Buera and Shin, 2013b, Caggese and Cuñat, 2013, Manova, 2013, Moll, 2014, Midrigan and Xu, 2014, Khan and Thomas, 2013], as we guide our empirical specification by the general implications of these models, i.e. self-financing by incumbents undoes the effect of financial frictions and allows firms to invest closer to the optimal level.[8] Our main contribution is to empirically estimate the saving and investment decisions of firms, which in these papers are an endogenous outcome of structural models calibrated with micro-data and built under different assumptions. As suggested by Hopenhayn [2014], this may be the source of the disparity of magnitudes reported for the aggregate effects of frictions. Our estimations may help to discipline these models. We provide empirical estimates of key elasticities and, unlike these papers, we exploit microeconomic data not only on real variables, but also on financial variables, for Chilean manufacturing firms. Ours is the first paper to provide empirical evidence of the self-financing channel studied in this literature. Our results show that, on one hand, self-financing is an active and relevant force at firm-level, allowing restricted firms to overcome financial constraints and converge towards their optimal capital over time. On the other hand, our estimated parameters suggest that self-financing operates slowly over time. Keeping firm productivity constant, simulations show persistent differences in the marginal product of capital between low-wealth and high-wealth firms with the same level of productivity, with gaps only closing after several decades. This supports the notion that the self-financing channel might be less strong than suggested elsewhere in the literature [e.g. Banerjee and Moll, 2010].

This paper also connects to two strands of research in corporate finance. One area of literature, starting with Fazzari et al. [1987], tries to identify financially constrained firms through the sensitivity of firms' investment to cash flows beyond profitability. Typically, profitability is captured by the

object of interest in those papers, and there has been no discussion on how to identify and estimate it when its relationship with productivity is not deterministic. We do this not only for investment, but also for the wealth accumulation policy function as well.

[8] In most cases, financial frictions generate a bound on investment that is increasing in current net wealth. Frictions can also be modeled as an interest rate spread that is decreasing in net wealth [e.g. Bernanke et al., 1999, Quadrini, 2000], or with the bound depending also on productivity, as predicted by models of endogenous imperfect markets [e.g. Aguirre, 2017, Brooks and Dovis, 2020]. Our empirical framework is consistent with these different specifications.

Tobin's Q or other observable characteristics of a firm. A second related area of literature discusses the determinants of firms' cash holding decisions and relates them to firm characteristics such as growth opportunities and risk management.[9] In our framework, the investment and asset accumulation policy functions are two of our outcomes, and we are able to identify unobservable productivity not only to control for profitability, but also to estimate non-linear and interaction effects with our measure of collateral. Furthermore, since we follow the structural macro models, we focus on net wealth instead of cash flows. Our results show that net wealth is a significant determinant of investment in our sample of Chilean firms, and that wealth accumulation decisions are affected by the firm's productivity process.[10]

The rest of the paper is organized as follows. Section 2 presents a simple model of firm dynamics with collateral constraints in order to shed light on the OP estimator in a setup with financial frictions. It also motivates the ingredients of the empirical model that we bring to the data. Section 3 introduces this empirical model and its assumptions. Section 4 establishes identification of the production function, the productivity process and the policy functions. Section 5 describes the estimation methods. Section 6 describes the data and presents the main empirical results. It also extends the simple model in Section 2 to generate simulated data that are used to validate our empirical estimation. Section 7 concludes.

2 A Simple Model with Financial Frictions

This section describes a stylized structural model featuring the main ingredients used in the the macro literature focused on firms investing under financial constraints. We do not estimate this model, but instead use it as an instrument to motivate the ingredients of the more general empirical specification taken to the data. In that sense, this simplified model serves two purposes. First, it illustrates the nature of the biases incurred when estimating the production function using standard methods in the presence of financial constraints. We use this setup to derive the sign of the biases that emerge when applying OP to estimate the production function. Second, this setup provides insights on the general form of the firm policy rules for investment and wealth accumulation. These policy rules play a crucial role in our analysis, as they are informative about the extent of collateral constraints and the self-financing channel.

We start with a model where the only source of uncertainty is productivity, and then we extend it to illustrate the potential sources and implications of stochastic shocks in the policy functions. Technical details on econometric issues are omitted at the moment, as they are discussed at length in the next sections when we present the empirical specification.

Following the macro literature on firms subject to collateral constraints [see Buera et al., 2015, for a detailed analysis], we introduce a stylized firm maximization problem that generates predictions

[9] See, for example, Opler et al. [1999] and Almeida et al. [2004]

[10] Lian and Ma [2020] find that, for relatively large firms in the US, earnings are more relevant than the liquidation value of assets as collateral, although this is less so for small firms and varies across countries depending on their financial infrastructure. Our measure of net wealth includes last period retained earnings, and our specification can be easily modified to include total earnings separately from net wealth.

that are well known in the literature, e.g. investment is suboptimal in wealth-poor firms and firms accumulate wealth out of earnings in order to pledge it as collateral to obtain resources to invest in the future. The novelty is to show explicitly how the insights from this class of models relate to the literature on production function estimations.

Although we state the problem recursively, we use time indexes to facilitate the mapping to the empirical model. Lower cases variables denote their values in logs. An incumbent firm with initial wealth A_{it}, capital K_{it} and productivity Z_{it} solves the following dynamic problem to maximize the discounted value of distributed profits D_{it} choosing labor L_{it}, investment I_{it} and next period wealth A_{it+1}:

$$\begin{aligned} V(A_{it}, K_{it}, Z_{it}) &= \max_{A_{it+1}, I_{it}, L_{it}} D_{it} + \beta E\left[V(A_{it+1}, K_{it+1}, Z_{it+1}) | Z_{it}\right], \\ \text{s.t.} \qquad D_{it} + g(A_{it+1}) &= Y_{it} - W L_{it} - (r+\delta) K_{it} + (1+r) A_{it}, \\ Y_{it} &= Z_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} \\ K_{it+1} &= I_{it} + (1-\delta) K_{it}. \end{aligned}$$

where Y_{it} is the value added produced by firm i. Investment, which determines next period's capital, is decided before the firm observes its current productivity draw, while labor is decided contemporaneously with productivity.[11] The function $g(\cdot)$ is assumed to be convex, which given the use of linear preferences, rules out corner solutions.[12] The firm discounts future flows at β, capital depreciates at rate δ, and the firm pays interest rate r for its debt, implicitly defined by $K_{it} - A_{it}$.

As is standard in the literature, the log of productivity z_{it} follows a Markovian linear process

$$z_{it+1} = \rho z_{it} + \eta_{it}, \tag{1}$$

where $\eta_{it} \sim N(0, 1)$. In the empirical model we allow for a more flexible Markovian process.

Financial Constraints We assume firms face collateral constraints. Although our empirical specification does not depend on the specific nature of the constraint, we consider in this section the case in which collateral defines an upper-bound for debt. This type of constraint rules out equilibrium default and can be obtained as the result of a simple limited-enforcement problem [see e.g. Buera et al., 2011]. Additionally, due to its simplicity it has been widely used in the macro literature. An alternative, also consistent with our empirical framework, is to assume that collateral affects borrowing costs.[13] Both

[11] This timing assumption is relevant in OP and related production function estimation methods, although it is not the most common assumption in the macro literature. Some papers assume capital is chosen within the period, mainly because assuming otherwise enlarges the state-space considerably [see e.g. Midrigan and Xu, 2014].

[12] Although assuming linear preferences is not needed in our empirical framework, it simplifies the illustrative analysis in this section. The inclusion of the convex function g introduces an incentive to smooth assets over time, ruling out corner solutions in which firms retain either all or none of their earnings. This specification combines ease of analysis with the general qualitative implications of models that introduce concavity in preferences.

[13] The constraint on borrowing costs arises in an environment with equilibrium default and intermediaries that offer debt contracts under competitive markets. This implies that the firm faces an interest rate spread when borrowing funds. This spread depends on the amount the firm borrows, since the value of paying back to the intermediary, relative to defaulting, is decreasing on debt [see e.g. Bernanke et al., 1999, Quadrini, 2000, Herranz et al., 2015].

cases generate a wedge in the investment optimality condition that depends on collateral. As we show below this causes serious problems for existing production function methodologies, and it is one of the features we exploit in our empirical specification.

Following Buera et al. [2015] we consider the following specification

$$K_{it+1} \leq \kappa(A_{it}, Z_{it}) \tag{2}$$

This class of models usually assume that only net-worth influences the upper-bound on capital κ. However, a more general specification may also include firm productivity as a determinant of the upper-bound. This will arise endogenously in models in which firm productivity (or value added) is observable for intermediaries, and may increase repayment in the case of default or contain information about default probabilities, as in Aguirre [2017], Brooks and Dovis [2020] or Lian and Ma [2020] [see Buera et al., 2015, for a closer examination]. In the final part of this section we allow for heterogeneity across firms in collateral constraints, incorporating a firm-specific stochastic component.

Optimality Conditions We first consider the FOC with respect to labor. Since the firm observes Z_{it}, we have

$$\beta_l Z_{it} K_{it}^{\beta_k} L_{it}^{\beta_n - 1} = W. \tag{3}$$

Using (3), the FOC with respect to investment can be written as:

$$C_k E(Z_{it+1}|Z_{it})^{\frac{1}{1-\beta_l}} \left(I_{it} + (1-\delta)K_{it}\right)^{\frac{\beta_k}{1-\beta_l}-1} = \beta(r+\delta) + \mu(A_{it}, Z_{it}), \tag{4}$$

where C_k is a constant. The last term in the right hand side is the wedge due to financial frictions. It corresponds to the multiplier of the collateral constraint (2), which is decreasing in both of its arguments. Note that if we had assumed that collateral affects borrowing costs, that term would be the spread, and would have also been a decreasing function of net-worth.

After taking logs and expectations over Z_{it+1} we can express (4) as:

$$k_{it+1} = c_k + \frac{\rho}{(1-\beta_k-\beta_l)} z_{it} - \frac{\rho(1-\beta_l)}{(1-\beta_k-\beta_l)} \widetilde{\mu}_{it} \tag{5}$$

where $\widetilde{\mu}_{it} = \ln(r + \delta + \mu(A_{it}, Z_{it}))$ and c_k is a constant.

If the constraint is not binding, wealth does not play a role, and, conditioning in initial capital, there is a positive monotonic relationship between investment and productivity, exactly the one exploited by the proxy variable framework. However, when the constraint binds, the multiplier is different from zero and investment is increasing in the stock of wealth for a given level of productivity. In line with the literature on production function estimation, we can express (4) by the following general function

$$i_{it} = h(z_{it}, k_{it}, a_{it}) \tag{6}$$

where $h_z > 0$, $h_k < 0$ and $h_a \geq 0$.

Finally, in an environment with collateral constraints, the firm must decide on wealth accumulation, which is crucial to finance future investment. The FOC in this case is given by:

$$g'(A_{t+1}) = \beta \left(1 + r + E_t \left[\kappa_A \mu(A_{t+1}, Z_{t+1})\right]\right) \tag{7}$$

Hence, even if the constraint does not bind today but is expected to bind in the future, there is an additional benefit from wealth accumulation. An additional dollar of retained earnings allows the firm to increase investment in κ_A dollars when the constraint binds. The marginal benefit is then the expected marginal product of capital net of borrowing costs, the value of the multiplier. Since productivity is persistent, higher productivity today increases the expected marginal product of capital for tomorrow, generating a positive correlation between productivity and wealth accumulation. We are interested in estimating the non-linear relationship between net-worth and the state variables defined in the firm's problem. Similarly to investment, we can define this general relationship as

$$a_{it+1} = g(z_{it}, k_{it}, a_{it}) \tag{8}$$

In section 3 we exploit the positive relationship between productivity and wealth accumulation by explicitly using the wealth accumulation policy function to learn about the firm's productivity process and the firm's production function.

2.1 The bias in the OP estimator under financial frictions

We use the model described above to illustrate the biases that appear when estimating the parameters of the firm production function using standard methods which do not account for financial frictions. In a very influential paper, Olley and Pakes [1996] propose a proxy variable approach to address the endogeneity problem that arises when estimating the parameters β_l and β_k from a value-added production function in logs, using data on value added y_{it}, capital k_{it} and labor l_{it}:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \tag{9}$$

where ε_{it} is measurement error in value added.[14] The main challenge in the estimation of β_l and β_k is that z_{it} is an unobservable variable for the econometrician which is potentially correlated with the observable regressors k_{it} and l_{it}, creating an endogeneity problem in the OLS regression of y_{it} on k_{it} and l_{it}.

The OP approach relies on using the investment policy function as an auxiliary equation to obtain information on the unobserved productivity z_{it}. For example, in the absence of constraints, we can see from the investment policy function (6) that: $i_{it} = h\left(z_{it}, k_{it}\right)$. Under the assumptions that z_{it} is the only unobserved variable for the econometrician in h (known as the scalar unobserved assumption) and

[14] We focus on a model with perfect competition where output prices are homogeneous across firms as in Olley and Pakes [1996], Levinsohn and Petrin [2003], Ackerberg et al. [2015], and Gandhi et al. [2020]. For production function estimation with monopolist competition and heterogeneous markups see De Loecker [2011a,b] and Bond et al. [2021].

that h is monotonic in z_{it}, we can invert the policy function to recover productivity as $z_{it} = h^{-1}(i_{it}, k_{it})$ and construct valid moment conditions. For instance, we can rewrite (9) as:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + h^{-1}(i_{it}, k_{it}) + \varepsilon_{it}. \tag{10}$$

Since ε_{it} is assumed to be uncorrelated with the inputs, OP propose to approximate $h^{-1}(i_{it}, k_{it})$ with a high-order polynomial on investment and capital and run an OLS regression of y_{it} on l_{it}, k_{it}, and the non-linear, time-dependent polynomial $h^{-1}(i_{it}, k_{it})$ to estimate β_l and β_k. However, the OLS regression identifies β_L, but cannot separate β_k from the linear part of $h^{-1}(i_{it}, k_{it})$. Thus, in a second step, OP exploits the Markovian productivity process to estimate β_k by regressing the following model:

$$\hat{\pi}_t(i_{it}, k_{it}) = \beta_k k_{it} + \rho\hat{\pi}_{t-1}(i_{it-1}, k_{it-1}) - \rho\beta_k k_{it-1} + \eta_{it} + \hat{\varepsilon}_{it} \tag{11}$$

where $\hat{\pi}_t(i_{it}, k_{it})$ denotes the estimated fraction of output explained by investment and capital in the first step, $\pi_t(i_{it}, k_{it}) = \beta_k k_{it} + h^{-1}(i_{it}, k_{it})$ [see e.g. Ackerberg et al., 2015] for details.

Intuitively, the OP method interprets observed differences in investment between firms in the data as differences in unobserved productivity between those firms. Hence, by controlling for investment in the production function we can eliminate the endogeneity problem and get consistent estimates of β_l and β_k. However, under borrowing constraints, differences in investment between firms are not only reflecting differences in productivity but also might be driven by differences in borrowing capacity.

In the model with financial frictions described above, the investment function arising from (4) depends not only on productivity and initial capital, but also on net-worth, through its influence on the strength of financial frictions. When we invert the investment policy function in (6) we obtain $z_t = h^{-1}(i_{it}, k_{it}, a_{it})$, with $h_i^{-1} > 0$, $h_k^{-1} > 0$ and $h_a^{-1} \leq 0$. Therefore, for a given level of investment, firms facing more severe constraints due to low levels of net-worth are more productive. The intuition is direct: For a given productivity level, an unconstrained firm will always invest more than a constrained firm. Therefore, for a given level of investment, it must be that the unconstrained firm is less productive. Replacing z_{it} in the production function we have:

$$y_t = \beta_l l_{it} + \beta_k k_{it} + h^{-1}(i_{it}, k_{it}, a_{it}) + \varepsilon_t \tag{12}$$

Hence, when implementing OP's first step, the term that captures the severity of the constraint due to net-worth would go to the error term of the OP regression in equation (10). Thus, if firms operate under borrowing constraints, the OP regression will render biased estimates of β_l and β_k due to the correlation of the regressors with the omitted variable a_{it}. Given that the OP estimation proceeds by two steps, we can analyze the biases separately. Let's focus first on the estimation of β_l. To see the sign of the correlation between l_{it} and a_{it} replace the expression for z_{it} obtained after inverting (6) in the FOC for labor (3):

$$l_{it} = c_l + \frac{1}{1-\beta_l}\left(\beta_k k_{it} + w + h^{-1}(i_{it}, k_{it}, a_{it})\right) \tag{13}$$

Therefore, after controlling for k_{it} and i_{it}, the correlation between l_{it} and the OP residual is positive.[15][16] Because OP cannot control for a fraction of productivity, which goes into the residual term when applying OP to (12), and as labor is increasing in productivity, the coefficient is biased upwards and $\hat{\beta}_l^{OP} > \beta_l$. To see the intuition suppose there are two firms with different productivities but that have the same level of capital and investment due to differences in collateral. OP will tend to equalize estimated productivity between the two, despite differences in output. The productive firm, that is more financially constrained, will choose to hire more workers, since frictions do not directly affect the labor market.[17] Since the OP estimator equalizes productivity between the two firms (given that they have the same investment), it will assign all the difference in output to differences in the amount of labor, leading to an overestimation of β_l.

In the case of the capital elasticity the relevant regression is the one implemented in the second stage (equation (11)). In the OP estimation, the function $\hat{\pi}_{t-1}()$ does not include a_{t-1} and this part of the function goes to the regression's error term. Given that $h_a \geq 0$ in equation (6) and that k_{it} is increasing in i_{it-1}, there is a positive correlation between the stock of capital used in production, k_{it}, and a_{it-1} - the level of collateral at the moment the investment decision is taken-. Therefore, $\tilde{h}_a < 0$ implies a negative correlation between the OP residual in equation (11) and k_{it}, leading to a downward bias: $\hat{\beta}_k^{OP} < \beta_k$. Intuitively, financial constraints generate differences in investment and capital for equally productive firms. The OP framework interprets the observed differences in investment as differences in unobserved productivity, and assigns part of the observed differences in output, which are due to capital, to variations in the productivity proxy, implying a lower estimated marginal effect of capital.

A final observation is that OP will underestimate (overestimate) the dispersion of productivity across firms if more productive firms are more (less) constrained. This depends on the strength of productivity in relaxing constraints both directly, as an argument in κ, and indirectly, through faster wealth accumulation. If these effects are not strong enough to overcome the greater capital needs of productive firms, then, since OP underestimates the productivity of constrained firms, we would expect OP to shrink the estimated productivity distribution relative to its actual value.

The analysis so far suggests that a direct solution to the biases associated to OP in the context of financial constraints is to include net-worth as an additional argument in the investment polynomial. We add to this the fact that models of collateral constraints give us an additional policy function (equation 8) that can be inverted and be used to control for unobserved productivity. In fact, this function might work better than h in the presence of nonconvex adjustment costs or when the upper

[15] For example, if $h^{-1}(i_{it}, k_{it}, a_{it})) = \tilde{h}_i i_{it} + \tilde{h}_k k_{it} + \tilde{h}_a a_{it}$ were linear, then the sign of the biases in $\hat{\beta}_l^{OP}$ and $\hat{\beta}_k^{OP}$ will depend on $\tilde{h}_a E\left[l_{it} a_{it} \mid i_{it}, k_{it}\right] > 0$ and $\tilde{h}_a E\left[k_{it} a_{it-1} \mid \hat{\pi}_{t-1}, k_{it-1}\right] < 0$.

[16] Note that l_{it} depends only on the constants c_l and w, and on state variables, so it is linearly dependent with the rest of the regressors in the production function regression [see Ackerberg et al., 2015]. In our empirical model we allow for the existence of an additional determinant of labor that can capture firm-specific iid shock in wages.

[17] Other models consider that financial constraints can affect the labor input as well. However, we should still expect an upward bias on β_l when the effect of frictions in the labor input are less severe. In our empirical model we will allow the labor input to also depend on the collateral constraint.

limit κ is not influenced by productivity.[18]

2.2 The Effect of Shocks in the Policy Functions

The fact that the estimation of the policy rule has not been an object of interest by itself explains its limited role in this literature. In this paper we pay special attention not only to the estimation of the investment policy rule, but also to the characteristics of the wealth accumulation policy rule. This policy rule provides key elasticities that are informative about the intensity of financial constraints and the self-financing channel, a relevant question in the macro literature in its own terms.

In order to obtain meaningful empirical estimates, we allow for the presence of stochastic components in the policy functions. This is in contrast to the restrictive assumption made by the literature so far, which states that there is a deterministic relationship between investment, capital and productivity.[19] In the presence of stochastic shocks, this strong assumption not only biases the estimates of the policy rule coefficients, a manifestation of the bias that appears in models with classical measurement error, but also those of the production function elasticities obtained under the proxy variable approach. As we discuss in more detail in section 4, the existence of the self-financing channel helps us to overcome this problem.

The model presented in this section can provide possible economic interpretations for the introduction of stochastic shocks in the policy functions.[20] In the case of investment our claim is that there exist an iid shock v_{it} in the FOC for capital (4). Given our focus on financial frictions, the collateral constraint is a natural source for shocks affecting the investment policy function. It may well be the case that firms face temporary idiosyncratic shocks that affect the relationship between debt, productivity and collateral. In this case we can set $\kappa(Z_{it}, A_{it}, v_{it})$ and, after inverting expression (6), we obtain $z_{it} = h^{-1}(i_{it}, k_{it}, a_{it}, v_{it})$, the term that will go into equation (12).[21]

In the case of the wealth accumulation policy function, stochastic shocks can come from unexpected fluctuations in the valuation of firms' financial portfolio or fixed assets. If these occur between the distribution of dividends (when equation 7 is solved) and the time at which the firm uses wealth as collateral to borrow (when equation 6 is solved), then they will appear as unplanned changes in the value of collateral in our framework.

[18] If this were the case investment would not respond to productivity shocks in constrained firms. This invalidates the OP's monotonicity assumption Shenoy [2020], but only for investment since net-worth would still respond to productivity shocks in constrained firms. However, when allowing for shocks in the policy functions, we need both investment and asset accumulation varying with productivity for both constrained and unconstrained firms.

[19] An exception is Hu et al. [2020], who introduce a stochastic component in the investment policy function without financial frictions, although the function itself is not estimated.

[20] Alternatively, the shocks may also reflect measurement error in the respective variables.

[21] An alternative is to extend the model to consider stochastic adjustment costs to capital [See Bachmann et al., 2013].

3 General Empirical Framework

We consider the same Cobb-Douglas production function in equation (9),

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \tag{14}$$

Following the proxy variable literature, we augment equation (1) to consider a nonlinear Markovian process for the Hicks-neutral productivity

$$z_{it} = \varphi\left(z_{it-1}\right) + \eta_{it}, \tag{15}$$

The function $\varphi\left(z_{it-1}\right) = E\left[z_{it} \mid z_{it-1}\right]$ is a non-parametric function of z_{it-1} which is known by the firm. As in the proxy variable approach, η_{it+1} and ε_{it} are not part of the information set for firm's decisions at t. The assumptions about the stochastic processes of ε_{it} and η_{it} are explained in detail at the end of this section. Following the model in section 2, capital k_{it} is a dynamic but predetermined input which was decided in $t-1$ when the firm chose i_{it-1}, and labor l_{it} is a flexible input. So far our empirical model is similar to the empirical models in Olley and Pakes [1996] and Ackerberg et al. [2015]. However, we depart from their setup in the specification of the empirical policy rules, in line with the stylized model discussed in the previous section:

$$i_{it} = h_t\left(z_{it}, k_{it}, a_{it}, v_{it}\right), \tag{16}$$

$$a_{it+1} = g_{t+1}\left(z_{it}, a_{it}, k_{it}, w_{it+1}\right). \tag{17}$$

where h_t and g_{t+1} are the empirical counterparts of the theoretical investment and wealth policy functions that can be derived in a firm-dynamics model with financial frictions as the one discussed in section 2. The specification in (16) adds two new ingredients to the investment functions described in Olley and Pakes [1996] and Ackerberg et al. [2015]. First, it includes a_{it} as a state variable to control for the existence of collateral constraints. Second, and as described earlier, it allows for an additional unobserved shock v_{it} in the policy function. A third novelty from the setup in Olley and Pakes [1996] and Ackerberg et al. [2015] is the inclusion of the self-financing channel in our empirical framework, as captured by equation (17).The function g_{t+1} also includes an additional unobserved shock, w_{it+1}. This can capture unobserved factors, other than z_{it}, that affect the evolution of wealth, like the interest rate shock discussed in section 2.2. Both v_{it} and w_{it+1} are assumed to be i.i.d and independent of the state variables. [22] Also, h_t and g_{t+1} are monotonic in v_{it} and w_{it+1}, respectively. Importantly, the nonlinear functions h_t and g_{t+1} allow for heterogeneous effect of productivity shocks on investment and on wealth accumulation, depending on the amount of collateral and the productivity level of the

[22] In the absence of shocks w_{it+1} to the wealth accumulation policy rule, the fact that the self-financing channel implies that wealth accumulation is increasing in productivity ensures that the policy rule satisfies the monotonicity assumption, a requirement of the proxy variable framework. In our model with shocks, the relationship between wealth accumulation and productivity across the complete distribution of firms will be important for identification.

firm. Particular objects of interest are the following average derivative effects with respect to z_{it}:

$$\Phi_t^h(a,k,z) = E\left[\frac{\partial h_t(z_{it},k_{it},a_{it},v_{it})}{\partial z}\right] \qquad \Phi_{t+1}^g(a,k,z) = E\left[\frac{\partial g_{t+1}(z_{it},k_{it},a_{it},w_{it+1})}{\partial z}\right]$$

In section 6 we document how these new measures of investment and wealth accumulation responses to productivity shocks vary along the wealth distribution and the productivity distribution. The estimated investment response $\hat{\Phi}_t^h(a,k,z)$ provides novel evidence on financial frictions whereas the estimated wealth response $\hat{\Phi}_{t+1}^g(a,k,z)$ novel evidence on the self-financing channel.

Finally, and following, Ackerberg et al. [2015], we model the labor input as a non-dynamic input in the sense that current choices are not affected by past values:

$$l_{it} = n_t(z_{it},a_{it},k_{it},w_{l,it}), \tag{18}$$

where equation (18) is the empirical labor decision. An extension from the stylized model in section 2 is that our empirical specification allows for potential effects of financial frictions over labor decisions, as represented by the inclusion of a_{it} in the policy function. Once again, the term $w_{l,it}$ represents a shock that is independent across periods and independent of the state variables a_{it}, k_{it} and z_{it}. This $w_{l,it}$ can capture exogenous transitory shocks to wages in the model in section 2, or optimization errors as the ones discussed in Ackerberg et al. [2015]. To complete the model details, we formally make the following assumptions, using the notation $x_i^t = (x_{i1},\ldots,x_{it})$ for any variable x_{it}.

Assumption 1. *(Conditional Independence). For all $t \geq 1$:*

(i) ***Output Shock:*** *ε_{it+s} for all $s \geq 0$ is independent over time and independent of $a_i^{t-1}, z_i^{t-1}, i_i^{t-1}, k_i^{t-1}, l_i^t$, y_i^{t-1} and η_{it+s}. Also ε_{i1} is independent of z_{i1}, a_{i1} and k_{i1} and $E[\varepsilon_{it}] = 0$.*

(ii) ***Productivity Shock:*** *η_{it+s} for all $s \geq 0$ is independent over time and independent of $a_i^{t-1}, z_i^{t-1}, i_i^{t-1}, k_i^{t-1}, l_i^{t-1}$, and y_i^{t-1}.*

(iii) ***Policy Functions Shocks:*** *v_{it} and w_{it+1} are mutually independent and also independent of $z_{i1}, (\varepsilon_{is},\eta_{is})$ for all s and of v_{is} and w_{is+1} for all $s \neq t$.*

Assumption 2. *(First Order Markovian). For all $t \geq 1$:*

(i) a_i^{t+1} is independent of $\left(a_i^{t-1},k_i^{t-1},z_i^{t-1}\right)$ conditional on (a_{it},k_{it},z_{it})

(ii) i_i^t is independent of $\left(a_i^{t-1},k_i^{t-1},z_i^{t-1}\right)$ conditional on (a_{it},k_{it},z_{it})

Parts (i) and (ii) of assumption 1 state that current and future productivity and production shocks, which are independent of past productivity and production shocks, are also independent of the current and past wealth and capital stocks, investment, and labor decisions. The initial wealth stock a_{i1}, initial capital stock k_{i1} and initial productivity z_{i1} are arbitrarily dependent. Allowing for a correlation between a_{i1}, k_{i1} and z_{i1} is important, as wealth and capital accumulation upon entry in the sample may be correlated with past persistent productivity shocks. Part (iii) requires investment and wealth shocks to be mutually independent, independent over time and independent of production components. Assumption 1 implies that ε_{it}, v_{it} and w_{it+1} are independent of the state variables $(k_{it},a_{it,}z_{it})$ and mutually independent conditional on $(l_{it},k_{it},a_{it,}z_{it})$. Hence, assumption 1 provides the exclusion restrictions necessary for identification, while assumption 2 is a first order Markov condition on wealth and capital dynamics. Assumption 2-(i) is a natural assumption in macro models with a self-financing channel as the one presented earlier; assumption 2-(ii) is a standard assumption both in the proxy variable framework (see Ackerberg et al. [2015]) and well as in structural models as the one in the previous section.

4 Identification

In this section, we establish identification of the nonlinear dynamic panel model presented in the previous section. The presence of additional shocks in the policy functions makes the identification challenges of our model more demanding than those of firm dynamics models studied in the proxy variable literature [Olley and Pakes, 1996, Levinsohn and Petrin, 2003, Ackerberg et al., 2015]. Therefore, it is important to show that the model we aim to estimate can actually be identified from data. Our model takes the form of nonlinear state-space models. Recently, Hu and Schennach [2008], Hu and Shum [2012], and Arellano et al. [2017] have established conditions under which dynamic nonlinear models with latent variables are non-parametrically identified under conditional independence restrictions. We built on these papers to provide nonparametric identification of the empirical model introduced in section 3. In particular, the goal of this section is to show that β_k, β_l, $\varphi(z_{it-1})$, h_t, g_{t+1} are identified from data on (y_{it}, k_{it}, l_{it}, i_{it}, a_{it},, a_{it+1}) given that (z_{it}, w_{it+1}, v_{it}, ε_{it}) are not observed by the econometrician and z_{it} is correlated with (l_{it}, a_{it}, k_{it}).

4.1 Intuition in a linear model

We first provide an intuition for identification using a version of the model with parametric linear policy functions. Then, we generalize these ideas to establish identification in the case with non-parametric policy functions.

Consider the following linear version of equations (15), (16) and (17)

$$z_{it} = \rho_z z_{it-1} + \eta_{it}, \tag{19}$$

$$i_{it} = h_z z_{it} + h_a a_{it} + h_k k_{it} + v_{it}, \tag{20}$$

$$a_{it+1} = g_z z_{it} + g_a a_{it} + g_k k_{it} + w_{it+1}, \tag{21}$$

Notice that the standard models using the proxy variable approach assume h_a=0 and v_{it}=0 and do not model explicitly equation (21).

Using equation (20), z_{it} can be written as a linear separable function of i_{it}, a_{it}, k_{it} and v_{it}.

$$z_{it} = \pi_1 i_{it} + \pi_2 a_{it} + \pi_3 k_{it} + \pi_4 v_{it} \tag{22}$$

where $\pi_1 = 1/h_z$, $\pi_2 = -h_a/h_z$, $\pi_3 = -h_k/h_z$ and $\pi_4 = -1/h_z$. If we replace equation (22) into the production function, we get:

$$y_{it} = \beta_l l_{it} + (\beta_k + \pi_3)k_{it} + \pi_1 i_{it} + \pi_2 a_{it} + \tilde{\varepsilon}_{it} \tag{23}$$

where $\tilde{\varepsilon}_{it} = \varepsilon_{it} + \pi_4 v_{it}$. If $v_{it} = 0$, then $\tilde{\varepsilon}_{it} = \varepsilon_{it}$ and, given assumption 1-(i), a simple OLS regression between y_{it} on l_{it}, k_{it}, i_{it} and a_{it} identifies β_l, as in the proxy variable approach. The difference with OP is that our regression controls for a_{it}. Note that β_k cannot be separately identified from π_3. As in the proxy variable approach, in a second step (once we have identified β_l), we exploit the Markovian assumption of the productivity process in (19), which combined with (23) leads to the following:

$$y_{it} - \beta_l l_{it} = \beta_k k_{it} + \rho_z \pi_3 k_{it-1} + \rho_z \pi_1 i_{it-1} + \rho_z \pi_2 a_{it-1} + \varepsilon_{it} + \pi_4 v_{it-1} + \eta_{it} \tag{24}$$

Again, if $v_{it-1} = 0$, and given assumption 1-(i), an OLS regression of (24) can identify β_k. The difference with OP would be that our second stage controls for a_{t-1}.

In contrast, in the more general case with investment shocks in equation (20) (i.e $v_{it} \neq 0$), investment i_{it} can be thought as a proxy measure with noise v_{it} for the latent variable z_{it}, conditioned on the observed state variables a_{it} and k_{it}. Therefore, the OLS regressions of (23) and (24) cannot identify β_l and β_k given that $E\left(i_{it}\tilde{\varepsilon}_{it}\right) \neq 0$ and $E\left(k_{it}\tilde{\varepsilon}_{it-1}\right) \neq 0$. Even if the investment shock v_{it} is not correlated with l_{it}, an OLS estimation of (20) will generate a bias in the estimation of β_l through the correlation of l_{it} and the latent variable z_{it} as in the classical linear multivariate model with measurement error in one regressor.

4.1.1 A simple solution: IV identification

Production Function To solve the endogeneity in the proxy variable approach, we notice that the self-financing channel provides a second noisy measure of productivity in a setup with financial frictions. Hence, a_{it+1} can be used as an instrument for investment in equation (23) given the conditional independence assumption in assumption (1) (wealth does not have a direct effect in the production function) and the relevance condition implied by the self-financing channel $\partial g_{t+1}/\partial z \neq 0$. Note that the functions h_t and g_{t+1} are correlated conditional on a_{it} and k_{it} via z_{it}. Therefore, we can construct the following IV moment restriction from (23):

$$E\left[y_{it} \mid a_{it+1}, l_{it}, k_{it}, a_{it}\right] = \beta_l l_{it} + (\beta_k + \pi_3)k_{it} + \pi_1 E\left[i_{it} \mid a_{it+1}, k_{it}, l_{it}, a_{it}\right] + \pi_2 a_{it}. \tag{25}$$

A regression between $E\left[y_{it} \mid a_{it+1}, l_{it}, k_{it}, a_{it}\right]$, which is an object that can be computed from data, and $\left[l_{it}, k_{it}, E\left[i_{it} \mid a_{it+1}, k_{it}, l_{it}, a_{it}\right], a_{it}\right]$ from (25) identifies $\{\beta_l, \pi_1, \pi_2\}$, which in turn can identify $\{h_z, h_a\}$. Then, β_k is identified from (24) using the following moment condition:

$$E\left(\pi_4 v_{it-1} + \eta_{it} + \varepsilon_{it} \mid k_{it-1}, a_{it-1}, a_t\right) = 0$$

The presence of the self-financing channel is key for identification. According to the model in section 2, a firm experiencing a positive persistent productivity shock should increase investment and also accumulate wealth. Therefore, the covariance between i_{it} and a_{it+1} allows us to isolate the variation in i_{it} due to variation in z_{it} from the variation in i_{it} due to variation in v_{it}. The identification sketch that we develop here provides a direct and simple estimation procedure by doing an IV regression to the proxy method. Note that this approach also works for more flexible policies that allow for nonlinearities in the observed state variables and interactions between the observed state variables and productivity like

$$i_{it} = h_{1t}(k_{it}, a_{it}) + h_{2t}(k_{it}, a_{it}) z_{it} + v_{it}, \tag{26}$$

where h_{1t} and h_{2t} are continuous nonlinear functions in a_{it} and k_{it}. The identification of β_l and β_k using the IV-proxy method strategy requires that at least one of the two policy functions is a polynomial of degree one in z_{it} and separable in z_{it} and the policy shock. If we think that this model is a better approximation for the wealth accumulation policy rule, we should invert this policy in the first step and then use the investment equation as the instrument. Note that the other policy function can be left unrestricted.

Remark 1. *In the empirical production model with financial frictions and Hicks-neutral productivity (equation (14)-(18)), if the investment policy rule is a polynomial of degree one in z_{it} and separable in z_{it} and the policy shock (as in equation (26)), assumption 1 holds, and the self-financing channel is active (i.e. $\partial g_{t+1}(z_{it}, a_{it}, k_{it}) / \partial z \neq 0$), the production function parameters β_l and β_k are identified.*

Policy Functions In the linear case, the identification challenge in the policy rules comes from the fact that they depend on unobserved productivity z_{it}. To overcome this, we exploit the Markovian process of z_{it} to construct valid instruments. Once β_l and β_k are identified we can define the net income process:

$$y_{it} - \beta_l l_{it} - \beta_k k_{it} = \tilde{y}_{it} = z_{it} + \varepsilon_{it} \tag{27}$$

Replacing (27) in (21):

$$a_{it+1} = g_z \tilde{y}_{it} + g_a a_{it} + g_k k_{it} + w_{it+1} - g_z \varepsilon_{it} \tag{28}$$

An OLS regression of a_{it+1} on $\tilde{y}_{it}$, a_{it} and k_{it} from equation (28) does not identify the policy function since $E(\tilde{y}_{it}\varepsilon_{it}) \neq 0$. However, it is possible to use $\tilde{y}_{it-1}$ as an instrument for $\tilde{y}_{it}$. The Markovian assumption for productivity provides the relevance condition, as it ensures that $E(\tilde{y}_{it}\tilde{y}_{it-1}) \neq 0$, while Assumption 1 ensures exogeneity $E(\tilde{y}_{it-1}\varepsilon_{it}) = 0$. A similar strategy can identify the investment policy rule in (20). Note that this approach also works for more flexible policies as long as the policy is a polynomial of degree one in z_{it} and separable in z_{it} and the policy shock (like equation (26)). This identification approach provides a direct and simple estimation procedure to estimate the policy rule.

Remark 2. *In the empirical production model with financial frictions and Hicks neutral productivity (equation (14)-(18)), if the policy rules are polynomials of degree one in z_{it} and separable in z_{it} and the policy shocks, assumption 1 holds, z_{it} is Markovian, and β_l and β_k are previously identified, then the policy functions are identified.*

Productivity Process For a linear productivity process, an IV argument exploiting the Markovian assumption identifies the persistence and dispersion parameters. Replacing (27) in (19):

$$\tilde{y}_{it} = \rho_z \tilde{y}_{it-1} + \eta_{it} + \varepsilon_{it} - \rho_z \varepsilon_{it-1}, \tag{29}$$

From equation (29), we can see that an OLS regression between $\tilde{y}_{it}$ and $\tilde{y}_{it-1}$ does not identify ρ_z since $\tilde{y}_{it-1}$ is correlated with ε_{it-1}. However, we can exploit the Markovian assumption on the process for z_{it} and *assumption 1* to use $\tilde{y}_{it-2}$ as an instrument for $\tilde{y}_{it-1}$ in equation (29). The following moment condition identifies ρ_z:

$$E\left(\tilde{y}_{it}\tilde{y}_{it-2}\right) = \rho_z E\left(\tilde{y}_{it-1}\tilde{y}_{it-2}\right),$$

Once we have identified ρ_z, then σ_η^2 and σ_ε^2 are identified from the following moment conditions:

$$E\left(\tilde{y}_{it}\tilde{y}_{it-1}\right) = \rho_z E\left(\tilde{y}_{it-1}\tilde{y}_{it-1}\right) - \rho_z\sigma_\varepsilon^2 \tag{30}$$

$$E\left(\tilde{y}_{it}\tilde{y}_{it}\right) = \rho_z^2 E\left(\tilde{y}_{it-1}\tilde{y}_{it-1}\right) + \sigma_\eta^2 + \left(1-\rho_z^2\right)\sigma_\varepsilon^2 \tag{31}$$

4.2 Nonparametric Identification

In this part we generalize the ideas sketched in the linear version to provide identification of the more general model, where the policy functions and the productivity process are modelled non-parametrically. The use of a general model allows for a richer interaction between productivity shocks and collateral constraints which are particularly important in macroeconomic models with financial frictions. As in the linear case, the sketch of identification is sequential. First, we establish identification of the production function parameters β_k and β_l. We then establish the identification of the productivity process and finally we show identification of the policy functions h_t and g_t. To establish identification of the parameters of the production function we impose the following high-level conditions:

Let $X_{it} = (a_{it,}k_{it}, l_{it})$ be the covariates of the model stated in equations (14)-(18) and let $f(a \mid b)$ be a generic notation for the conditional density $f_{A|B}(a \mid b)$.

Condition 1. *Almost surely in covariate values X_t: (i) the joint density $f(y_t, i_t, a_{t+1}, z_t \mid X_t)$ is bounded, as well as all its joint and marginal densities; (ii) the characteristic function of ε_{it} has no zeros on the real line; (iii) for all $z_{1t} \neq z_{2t}$, $Pr[f(i_{it} \mid z_{1t}, X_t) \neq f(i_{it} \mid z_{2t}, X_t)] > 0$; (iii) $f(a_{t+1} \mid z_t, X_t)$ is complete in z_{it}. (iv) for $\tilde{y}_{it} = y_{it} - \beta_l l_{it} - \beta_k k_{it}$, $f(\tilde{y}_{it} \mid \tilde{y}_{it-1})$, $f(z_{it} \mid \tilde{y}_{it-1})$, $f\left(z_{it} \mid \tilde{y}_i^T\right)$ are complete and the distribution of $f\left(z_{it} \mid a_i^t, k_i^t, \tilde{y}_i^T\right)$ is complete in $\left(a_i^{t-1}, k_i^{t-1}, \tilde{y}_i^T\right)$.*

Condition 1-(i) requires bounded densities. Condition 1-(ii) is a technical assumption previously used in the literature.[23] The normal distribution and many other standard distributions satisfy this condition. Condition 1-(iii) requires that $f(i_{it} \mid z_{it}, X_{it})$ be non-identical at different values of z_{it}. Note that if the investment policy rule is separable in v_{it}, the condition is fulfilled if h_t is strictly monotonic in z_{it}. Accordingly, the macro model with forward-looking financial constraints sketched in section 2 implies a monotonic relationship between productivity an investment.[24] Condition 1-(iv) is a

[23]This condition is used for the i.i.d shock of the household income in Arellano et al. [2017] and for the i.i.d shock in the firm production function in Hu et al. [2020].

[24]Also, macro models with backward-looking constraints (where the financial constraint only depends on collateral) generate an investment rule that is monotonic in investment, as long as the financial constraint is a soft constraint where firms can borrow at much as they want paying a premium in the interest rate that depends on the level of collateral.

completeness condition commonly assumed in the literature on nonparametric instrumental variables [Newey and Powell, 2003].[25] Intuitively, we need enough variation in the densities $f(a_{it+1} \mid z_{it}, a_{it}, k_{it})$ for different values of z_{it}. This requires a statistical dependence between wealth accumulation a_{it+1} and productivity z_{it} conditioned on the observed state variables. This requirement can be met by the self-financing channel in equation (17), which generates a positive relationship between productivity and wealth accumulation for all constrained and unconstrained firms. In the IV terminology, this is a relevance condition, that ensures that a_{it+1} is a valid instrument for z_{it}, similar to the condition discussed in the linear case.[26] Similarly, 1-(v) is a completeness condition that requires that z_{it} and z_{it-1} are statistically dependent, which is ensured by the Markovian assumption.

We then have the following result, which sequentially combines the results in Hu and Schennach [2008] and Arellano et al. [2017].

Theorem 1. *(Sequential identification) In a production function model with Markovian Hicks-neutral productivity and financial frictions as in (14)-(18), if assumption 1, assumption 2 and condition 1 (i)-(v) hold, then β_k, β_l, $\varphi(z_{it-1})$, h_t, g_{t+1} are identified from data on y_{it}, k_{it}, l_{it}, i_{it}, a_{it} for $T \geq 4$.*

As in the linear case discussed above, identification of the production function parameters is based on having two imperfect measures of the unobserved productivity process: the investment and wealth policy functions. Once the production function parameters are identified, the productivity process is non-parametrically identified from the dynamic dependence structure of the firm net-income process, following the ideas discussed in the linear case. Finally, once the productivity process is identified, the policy rules are identified using non-parametric instrumental variables arguments given the first-order Markovian assumption and the exclusion restrictions provided by our dynamic model. Below we discuss the sketch of the sequential identification and we leave the details for Appendix A1.

Production Function From *assumption* 1, ε_{it}, v_{it}, and w_{it+1} are independent conditional on $(l_{it}, k_{it}, a_{it,} z_{it})$, which can be interpreted as the exclusion restrictions in a nonlinear IV setting. Using this conditional independence assumption, we can write the following conditional distribution of the observed variables $f(y_t, i_t \mid a_{t+1}, X_t)$, which is a data object, in terms of some elements of the model that we aim to identify:

$$f(y_t, i_t \mid a_{t+1}, X_t) = \int f(y_t \mid z_t, k_t, l_t)\, f(i_t \mid z_t, X_t)\, f(z_t \mid a_{t+1}, X_t)\, dz_t \tag{32}$$

We notice that equation (32) can be framed into the setup studied in Hu and Schennach [2008]. Given *condition* 1(i)-(iv), Theorem 1 of Hu and Schennach [2008] can be applied to our setting to show that $f(y_t \mid z_t, k_t, l_t)$ is identified from the data, which leads to the identification of the production function parameters [see Hu et al., 2020])[27]

[25] The distribution of $\tilde{y}_{it} \mid \tilde{y}_{it-1}$ is complete if $E[\phi(\tilde{y}_{it}) \mid \tilde{y}_{it-1}] = 0$ implies that $\phi(\tilde{y}_{it}) = 0$ for all ϕ in some space.

[26] For example, if $(a_{it+1}, z_{it}, a_{it}, k_t)$ follows a multivariate normal distribution with zero mean, the completeness condition will require that $E[a_{it+1} z_{it}] \neq 0$ which is ensured by the self-financing channel.

[27] An important difference of our framework from Hu et al. [2020] is that our model with financial frictions provides

Productivity Process Once we have identified β_k, β_l, and given that the productivity is Hicks-neutral, we can write the firm net-income process $\tilde{y}_{it} = y_{it} - \beta_k k_{it} - \beta_l l_{it}$ as an additive model with two independent latent variables (given *assumption* 1). [28]

$$\tilde{y}_{it} = z_{it} + \varepsilon_{it} \tag{33}$$

Given that z_{it} is Markovian and ε_{it} is i.i.d over time, equation (33) has a similar structure to the household income process model with non linear Markovian persistent shocks studied in Arellano et al. [2017]. To identify the productivity process we rely on the fact that the net-income process in (33) has a Hidden-Markov structure (by *assumption* 1) where $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$ are independent given z_{it-1}. The additivity of the net-income process and *condition* 1-(v) allow us to identify the joint distribution of $(\varepsilon_{i2,\cdots,}\varepsilon_{iT-1})$ and the joint distribution of $(z_{i2,\cdots,}z_{iT-1})$ from the autocorrelation structure of $(\tilde{y}_{i1,\cdots,}\tilde{y}_{iT})$ for $T \geq 3$ and identify $\varphi(z_{it-1})$ for $T \geq 4$.

Policy Functions Once $\left(z_{i1} \mid \tilde{y}_i^T\right)$ is identified, we use *assumptions* 1 and *assumption* 2 to construct the following IV moment restriction, which allows us to relate the conditional distribution of observable variables $f\left(a_1, k_1 \mid \tilde{y}^T\right)$, $f\left(a_{t+1} \mid a^t, k^t, \tilde{y}^T\right)$, and $f\left(i_t \mid a^t, k^t, \tilde{y}^T\right)$ which are data objects, to the distribution of the policy rules we want to identify.

$$f\left(a_1, k_1 \mid \tilde{y}^T\right) = E\left[f\left(a_1, k_1 \mid z_1\right) \mid \tilde{y}_i^T = \tilde{y}^T\right] \tag{34}$$

$$f\left(a_{t+1} \mid a^t, k^t, \tilde{y}^T\right) = E\left[f\left(a_{t+1} \mid z_t, a_t, k_t\right) \mid a_i^t = a^t, k_i^t = k^t, \tilde{y}_i^T = \tilde{y}^T\right] \tag{35}$$

$$f\left(k_{t+1} \mid a^t, k^t, \tilde{y}^T\right) = E\left[f\left(k_{t+1} \mid z_t, a_t, k_t\right) \mid a_i^t = a^t, k_i^t = k^t, \tilde{y}_i^T = \tilde{y}^T\right] \tag{36}$$

where the expectation in (34) is taken with respect to the density of z_{i1} given $\tilde{y}_i^T$ for fixed values of a_1 and k_1 and the expectation in (35) and (36) are taken with respect to the density of z_{it} given $\tilde{y}_i^T$, k_i^t, and a_i^t for a fixed value of a_{t+1} and k_{t+1}, respectively. Equation (34) is analogous to a nonlinear IV problem where z_{i1} is the endogenous regressor and $\tilde{y}_i^T$ is the vector of instruments. The difference with a standard nonlinear IV is that the "endogenous regressor" in the moment condition in (34) is a latent variable. However, this is not a problem since we have identified $\left(z_{i1} \mid \tilde{y}_i^T\right)$ using the production function. Provided that the distribution of $\left(z_{i1} \mid \tilde{y}_i^T\right)$ is complete (*condition* 1(v)), the unknown density $f\left(a_1, k_1 \mid z_1\right)$ is identified from (34). Similarly, equations (35) and (36) can be

a policy rule (the self-financing channel) for an observed variable that is not directly linked to the production function regression (i.e a_{it+1} is not an input in the production function). Hence, we do not have the collinearity problem between inputs that leads Hu et al. [2020] to include k_{t+1} as a covariate in X_t. Our covariates in X_t allow us to have a standard law of motion for capital as in Olley and Pakes [1996] and Ackerberg et al. [2015] without the need of an unobserved component affecting the law of motion of capital. The latter is particularly important in applied work because in most cases researchers do not have data on both capital and investment separately and use the perpetual inventory method to recover the capital series from investment or vice-versa.

[28] For identification and estimation of production functions with non-neutral productivity see Doraszelski and Jaumandreu [2018] and Villacorta [2018].

interpreted as nonlinear IV restrictions where a_{it} and k_{it} are the controls (they are arguments in the wealth function and investment functions), and the vector $\tilde{y}_i^T$ contains the excluded instruments. Given *condition* 1(v) and *assumption* (2), the distributions $f(a_{t+1} \mid z_t, a_t, k_t)$ and $f(k_{t+1} \mid z_t, a_t, k_t)$ for $t > 2$ are identified recursively from equations (35) and (36). The identification of $f(a_{t+1} \mid z_t, a_t, k_t)$ and $f(k_{t+1} \mid z_t, a_t, k_t)$ allows us to recover the policy functions $g_{t+1}()$ and $h_t()$. As in the linear case we are using the autocorrelation structure of $\tilde{y}_i^T$ to construct instruments to identify the policy functions. In the linear example we use lagged values whereas here we use lagged and lead values of the firm's net income process.

5 Empirical Strategy

In this section we discuss three approaches to estimate different versions of the empirical model presented in section 3 and discussed in section 4. First, we consider a model without shocks in the policy functions, as in the proxy variable approach. For this model we propose two new proxy variables to estimate the model by GMM . Second, we consider a model that includes shocks in the policy functions, but where at least one of the policies is a quasi linear function in productivity and separable in productivity and the policy shock. For this model, we propose a novel procedure that consists of an IV regression within the proxy variable framework of Olley and Pakes [1996] and Ackerberg et al. [2015], following the identification strategy presented in section 4.1. Finally, we consider a more flexible model that allows for shocks in the policy functions and nonlinear effects of productivity. For this model we introduce a flexible estimation method well suited for nonlinear panel data models with latent variables.

5.1 Policy functions without shocks: proxy variable approach

Augmented OP: In a model where the investment equation is a deterministic function of the state variables of the model (z_{it},k_{it},a_{it}), it is possible to identify and estimate the model with financial frictions using the proxy variable approach by a simple modification of the moment conditions used by OP to control for collateral constraints. Under the assumption that the function h_t in (16) is monotonic in z_{it}, it is possible to invert the investment policy function to recover z_{it} as a function of i_{it}, k_{it} and a_{it} and follow the two-step approach discussed in Olley and Pakes [1996] and Ackerberg et al. [2015] (see Appendix A.4 for the details)

Wealth accumulation policy rule as the proxy variable Note that in the absence of shocks in the wealth accumulation policy rule we can also invert (17) and use the wealth accumulation as the proxy variable. Under the assumption that the function g_{t+1} in (17) is monotonic in z_{it}, it is possible to invert the wealth policy function and express z_{it} as a function of a_{it+1}, k_{it} and a_{it}. This approach is novel, since we are the first paper to use the self-financing channel as the proxy variable for the production function estimation. We refer to this novel estimator that use the wealth accumulation policy function to construct the proxy variable as *Proxy-Wealth*. Since z_{it} is perfectly recover, estimation of the productivity process and the policy functions are straightforward.

5.2 Policy functions with shocks

Our main specification allows for unobservable i.i.d shocks in the policy functions to capture unanticipated interest rate shocks, optimization error, modeling error or measurement error in the policies.

5.2.1 Parsimonious policy functions

Proxy-IV As discussed in section 4 for a policy functions that is a polynomial of degree one in productivity and separable in productivity and the policy shock we propose an IV estimator within the proxy variable approach. For example, consider the following wealth accumulation policy function:

$$a_{it+1} = g\left(z_{it}, k_{it}, a_{it}, w_{it}\right) = g_1\left(k_{it}, a_{it}\right) + g_2\left(k_{it}, a_{it}\right) z_{it} + w_{it+1}, \tag{37}$$

It is important to remark that model 37 is flexible enough to capture heterogeneous effects of productivity on wealth accumulation depending on the level of collateral. The investment policy is left unrestricted. As in the proxy variable approach we can invert equation (37):

$$z_{it} = \pi_1\left(k_{it}, a_{it}\right) + \pi_2\left(k_{it}, a_{it}\right) a_{it+1} + \omega_{it+1} \tag{38}$$

where $\pi_1\left(k_{it}, a_{it}\right) = -g_1\left(k_{it}, a_{it}\right)/g_2\left(k_{it}, a_{it}\right)$, $\pi_2\left(k_{it}, a_{it}\right) = 1/g_2\left(k_{it}, a_{it}\right)$ and $\omega_{it+1} = -w_{it+1}/g_2\left(k_{it}, a_{it}\right)$. Replacing (38) in the production function:

$$y_{it} = \beta_l l_{it} + \phi\left(k_{it}, a_{it}\right) + \pi_2\left(k_{it}, a_{it}\right) a_{it+1} + \omega_{it+1} + \varepsilon_{it}, \tag{39}$$

where $\phi(k_{it}, a_{it}) = \beta_k k_{it} + \pi_1\left(k_{it}, a_{it}\right)$. As we emphasize in section 4, an OLS regression of (39) does not deliver a consistent estimator of β_l since $E\left(\omega_{it+1} \mid a_{it+1}\right) \neq 0$. However, given *assumption 1*, i_{it} can be use as an instrument for a_{it+1} in equation (39). Therefore, we propose the following two-stage procedure:

First Stage: Estimate (39) with an IV estimator using $\pi_2\left(k_{it}, a_{it}\right) i_{it}$ as the instrument for $\pi_2\left(k_{it}, a_{it}\right) a_{it+1}$. The IV regression delivers a consistent estimator of β_l, $\phi\left(k_{it}, a_{it}\right)$ and $\pi_2\left(k_{it}, a_{it}\right) a_{it+1}$. For instance, in the linear case where $g_2\left(k_{it}, a_{it}\right) = 1$, i_{it} will be the instrument for a_{t+1}.

Second Stage: Combining equation (38) with the markovian model of the productivity process $z_{it} = \rho_z z_{it-1} + \eta_{it}$:

$$z_{it} = \rho_z \pi_1\left(k_{it-1}, a_{it-1}\right) + \rho_z \pi_2\left(k_{it-1}, a_{it-1}\right) a_{it} + \rho_z \omega_{it} + \eta_{it}, \tag{40}$$

Replacing equation (40) into the production function:

$$y_{it} - \beta_l l_{it} = \beta_k k_{it} + \rho_z \pi_1\left(k_{it-1}, a_{it-1}\right) + \rho_z \pi_2\left(k_{it-1}, a_{it-1}\right) a_{it} + \rho_z \omega_{it} + \eta_{it} + \varepsilon_{it}, \tag{41}$$

using *assumption 1* we can define the following moment condition from equation (41)

$$E\left(\omega_{it} + \eta_{it} + \varepsilon_{it} \mid k_{it}, k_{it-1}, a_{it-1}, i_{t-1}\right) = 0, \tag{42}$$

The moment condition in (42) allows us to identify β_k. If we replace β_l, $\pi_1(k_{it-1}, a_{it-1})$ and $\pi_2(k_{it-1}, a_{it-1})$ by their IV estimates from the first stage, an OLS regression of (41) delivers a consistent estimate of β_k. We refer to this novel estimator as *Proxy-IV*. Once β_l and β_k are estimated we can estimate the productivity process and the policy functions following the IV strategy discussed in section 4.1.

5.2.2 Flexible policy functions

To estimate more flexible policy functions that allow for nonlinear interactions between z_{it} and observed state variables we bring to the data the following nonlinear specifications. For $t = 1, \ldots, T$

$$\begin{cases} y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it} \\ z_{it} = \sum_{r=1}^{R} \alpha_r^{\varphi} \phi_r^{\varphi}(z_{it-1}) + \eta_{it} \\ i_{it} = \sum_{r=1}^{R} \alpha_r^{h} \phi_r^{h}\left(z_{it}, k_{it}, a_{it}, \delta_t^{h}\right) + v_{it} \\ a_{it+1} = \sum_{r=1}^{R} \alpha_r^{g} \phi_r^{g}\left(z_{it}, k_{it}, a_{it}, \delta_t^{g}\right) + w_{it+1} \\ a_{i1} = \sum_{r=1}^{R} \alpha_r^{g1} \phi_r^{g}\left(z_{i1}, \delta_1^{g1}\right) + w_{i1} \\ l_{it} = \sum_{r=1}^{R} \alpha_r^{n} \phi_r^{n}\left(z_{it}, k_{it}, a_{it}\right) + w_{l,it+1} \end{cases} \tag{43}$$

where ϕ_r^h, ϕ_r^g, ϕ_r^n and ϕ_r^{φ} are dictionary of functions and α_r^h, α_r^g, α_r^n, and α_r^{φ} are the parameters associated. Note that ϕ_r^h, ϕ_r^g, ϕ_r^n and ϕ_r^{φ} are anonymous functions without an economic interpretation. They are just building blocks of flexible models. Objects of interest will be summary measures of derivative effects constructed from the models. We follow the proxy variable literature and model the functions as high-order polynomials to allow for flexible interactions between productivity and observed state variables. We model stationary policy functions with time-invaring coefficients and additive errors to have a more parsimonious model to take to the data but, as we shown in section 4, the model is identified with time-varying functions and non-additive errors.[29] To control for unobserved aggregate shocks in the policy rules we include time-specific fixed effects δ_t^h and δ_t^g. Both δ_t^h and δ_t^g are left unrestricted, so we allow for potential correlation between them. This is important since for instance, an aggregate financial shock (like the financial crisis) might affect both policy rules. Finally, in our empirical specification we assume that v_{it}, w_{it}, η_{it} and ε_{it} are normally distributed.

Stochastic EM Estimation Algorithm (SEM) To estimate our nonlinear model with latent variables, we adapt a stochastic EM algorithm to our production function framework. Let $X_i^T = \left(y_i^T, k_i^T, l_i^T, a_i^T,\right)$ and z_i^T the history of observables and productivity for firm i, respectively. Given

[29] We can also model the empirical productivity process and the policy functions using quantile regressions to allow for nonlinear persistence in productivity -like then household income process in Arellano et al. [2017]- and richer interactions between state variables and the shocks in the policies. However, this would lead to a less parsimonious specification, and this extra richness is not needed for the purpose of this paper.

assumption 1, the full model in (43) imply the following integrated moment restrictions:

$$E\left(\int\begin{bmatrix}\sum_{t=2}^{T}\left(a_{it+1}-\sum_{k=1}^{K}\alpha_k^g\phi_k^g\left(z_{it},k_{it},a_{it},\delta_t^g\right)\right)^2\\ \sum_{t=1}^{T}\left(i_{it}-\sum_{k=1}^{K}\alpha_k^h\phi_k^h\left(z_{it},k_{it},a_{it},\delta_t^h\right)\right)^2\\ \sum_{t=1}^{T}\left(l_{it}-\sum_{k=1}^{K}\alpha_k^n\phi_k^n\left(z_{it},k_{it},a_{it}\right)\right)^2\\ \sum_{t=1}^{T}\left(y_{it}-\beta_l l_{it}-\beta_k k_{it}-z_{it}\right)^2\\ \sum_{t=1}^{T}\left(z_{it}-\sum_{k=1}^{K}\alpha_k^\varphi\phi_k^\varphi\left(z_{it-1}\right)\right)^2\\ \left(a_{i1}-\sum_{k=1}^{K}\alpha_k^{g1}\phi_k^g\left(z_{i1}\right)\right)^2\end{bmatrix} f\left(z_i^T \mid X_i^T,\theta\right)dz\right) \tag{44}$$

where $f\left(z_i^T \mid X_i^T,\theta\right)$ is the posterior density of the vector z_i^T given the data. The vector $\theta = \left[\theta^y,\theta^h,\theta^g,\theta^{g1},\theta^n,\theta^\varphi\right]$ contains all the parameters of the model in (43), $\theta^y = [\beta_k,\beta_l,\sigma_\epsilon]$, $\theta^h = \left[\alpha_1^h\ldots\alpha_K^h,\sigma_v\right]$, $\theta^g = \left[\alpha_1^g\ldots\alpha_K^g,\sigma_w\right]$, $\theta^\varphi = \left[\alpha_1^\varphi\ldots\alpha_K^\varphi,\sigma_\eta\right]$. Note that (44) are the integrated version of the unfeasible OLS regressions of the equations in (43). The OLS are unfeasible because we do not observe z_{it}.

The stochastic EM algorithm possesses computational advantages with respect to a maximum likelihood estimation of the model in (43), given that each policy function depends on a considerable number of parameters. Therefore, rather than maximize the likelihood with respect to a lot of parameters, our stochastic EM estimator iterates between simulating draws from the posterior distribution of latent productivity given the data $f\left(z_i^T \mid X_i^T,\theta\right)$ and simple OLS estimation of the parameters in θ.[30] Arellano et al. [2017] use a similar approach in a nonlinear panel model with latent variables to estimate an income process and nonlinear consumption and assets policy rules from household data.

The two following steps describe our procedure. Starting with a parameter vector θ^0, we iterate the following two steps on $s = 0,1,2,\ldots$ until convergence of the θ^s process to a stationary distribution:

1. *Stochastic E-step:* For each firm i, draw $\left\{z_{i1}^{(m)}\ldots z_T^{(m)}\right\}$ M realizations of z_i^T from $f\left(z_i^T \mid X_i^T,\theta\right)$. Using *assumptions 1 and 2* we can express the posterior distribution of z_{it} as a function of the likelihoods of the equations in (43).

$$\begin{aligned} f\left(z_i^T \mid X_i^T,\theta\right) = &\prod_{t=1}^{T} f\left(y_{it} \mid k_{it}, l_{it,} z_{it},\theta^y\right)\times\\ &\prod_{t=1}^{T} f\left(i_{it} \mid k_{it}, z_{it}, a_{it},\theta^h\right) f\left(l_{it} \mid k_{it}, z_{it}, a_{it},\theta^n\right)\times\\ &\prod_{t=2}^{T} f\left(a_{it} \mid z_{it}, k_{it}, a_{it},\theta^g\right) f\left(a_{i1} \mid z_{i1},\theta^{g1}\right)\times\\ &\prod_{t=1}^{T} f\left(z_{it} \mid z_{it-1},\theta^\varphi\right) f\left(z_{i1}\right)\end{aligned}$$

[30] For instance, if we specify our nonlinear functions as third-order polynomials, the model in (43) would contain more than 200 parameters to be estimated. If in addition we want to estimate policy functions that include firm fixed effects that maximum likelihood estimation would be computationally infeasible.

where $f\left(y_{it} \mid k_{it}, l_{it,} z_{it}, \theta^{y}\right)$ is the likelihood of the production function, $f\left(i_{it} \mid k_{it}, z_{it}, a_{it}, \theta^{h}\right)$ is the likelihood of the investment policy rule, $f\left(a_{it+1} \mid z_{it}, k_{it}, a_{it}, \theta^{g}\right)$ is the likelihood of the wealth policy rule and $f\left(z_{it} \mid z_{it-1}, \theta^{\varphi}\right)$ is the likelihood of the productivity process. To simulate $f\left(z_{i}^{T} \mid X_{i}^{T}, \theta\right)$, we use a random-walk Metropolis-Hastings sampler, targeting an acceptance rate of 0.3.

2. *M-step:* compute the integrated-OLS estimator of the parameters:

$$\begin{cases} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(y_{it}-\beta_{l}l_{it}-\beta_{k}k_{it}-z_{it}^{(m)}\right)^{2} \\ \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(i_{it}-\sum_{k=1}^{K}\alpha_{k}^{h}\phi_{k}^{h}\left(z_{it}^{(m)},k_{it},a_{it},\delta_{t}^{h}\right)\right)^{2} \\ \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{m=1}^{M}\left(a_{it+1}-\sum_{k=1}^{K}\alpha_{k}^{g}\phi_{k}^{g}\left(z_{it}^{(m)},k_{it},a_{it},\delta_{t}^{g}\right)\right)^{2} \\ \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{m=1}^{M}\left(z_{it}^{(m)}-\sum_{k=1}^{K}\alpha_{k}^{\varphi}\phi_{k}^{\varphi}\left(z_{it-1}^{(m)}\right)\right)^{2} \\ \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(l_{it}-\sum_{k=1}^{K}\alpha_{k}^{n}\phi_{k}^{n}\left(z_{it}^{(m)},k_{it},a_{it}\right)\right)^{2} \end{cases} \tag{45}$$

In practice, we stop the iterative procedure after S=500 iterations and check the convergence of the estimates. In each iteration of the chain we simulate 100 draws from step 1 (i.e M=100). We start the algorithm from different initial values (OP, OPA or Proxy-IV) and we get similar results. The statistical properties of a similar stochastic algorithm has been studied in Nielsen et al. [2000] in a likelihood context and in Arellano and Bonhomme [2016] in a GMM context where the M-step consists of quantile-based regressions. Arellano and Bonhomme [2016] show that the estimates of the stochastic EM algorithm for parametric models (where R does not grow with the sample size) are asymptotically normally distributed as M and N tend to infinity (for fixed R) with an asymptotic variance that is the asymptotic variance of the method-of-moments estimator of the integrated moment restrictions. Our M-step, which consist of a set of OLS regressions can be framed in the GMM framework studied in Arellano and Bonhomme [2016]. Therefore, θ has the following distribution as N and M go to infinity:

$$\sqrt{N\left(\hat{\theta}-\theta\right)} \xrightarrow{d} N\left(0, \Sigma\right)$$

where Σ is the asymptotic variance of the GMM estimator of (44).

6 Data and Empirical Results

6.1 Data

Our database comes from administrative records generated by Chile's tax collection agency (*Servicio de Impuestos Internos* - SII). The data covers all firms that operate in the formal sector and all formal wage employment in Chile. Each firm in this administrative dataset is assigned a unique identifier by SII, so they can be tracked across time while at the same time preserving anonymity and the confidentiality of the data. We use information contained in income tax form F22, which is submitted annually by firms. The data set contains information on *firms* (as opposed to *plants*) of all ages, sizes and sectors, although we focus on firms operating in the manufacturing sector. Firms are defined as productive units that generate revenue, utilize production factors and operate under a unique tax ID that allows us to track them across time. Data is present on annual frequency.

Form F22 has firm level information on annual sales, expenditures on intermediate materials, a proxy of the capital stock ("immobile assets") and the firm's wage bill, as well as the firm's economic sector. We combine this information with tax form 1887, which reports monthly information on individual workers that were employed on the firm, and therefore allows us to calculate a measure of annual employment adjusted by the number of months per worker.

Crucially, form F22 also provides information on the firms' balance sheets. In particular, we can build a measure of net worth, defined as the difference between reported total assets and total liabilities. This allows us to combine the information on the production side traditionally used in the literature on production functions and TFP estimates with information on the firm's self-reported wealth, and its evolution across time.

To clean up the raw data and have a consistent dataset with our empirical strategy, we follow several steps. First, we drop observations with zero or missing information for our proxy of capital, sales, expenditures on intermediate inputs, employment or net worth. Second, we focus on firms that have at least 5 workers. Third, we build a measure of annual investment by using the annual change in the capital stock, and assuming a 10% depreciation rate[31] The final dataset has 4867 firms in the manufacturing sector between 2005 and 2016.

The fact that the data provides information on balance sheets is an advantage relative to most databases used in the literature on production function estimations, either from surveys or administrative records, which typically provide detailed information on the production side of firms but do not account for assets or wealth. As clearly stated in the previous sections, access to joint information on the production process of firms and the evolution of wealth is an absolute necessity given our framework. Of course, the combination of financial statements and information on production activity is not exclusive to our dataset, and is also available, with long and detailed information for a large number of countries in datasets such as Compustat, Amadeus and Orbis. Relative to those sources, our dataset has the advantage of including firms of all ages and sizes in the context of a developing country. In that sense, this might be a better setup to study the effects of financial frictions, that are likely to be less relevant in the developed world, in particular for relatively large firms. Other datasets, such as the Enterprise Surveys conducted by the World Bank, are similar to ours in that they also include firms of all sizes in developing countries, although by their nature they are less suited to follow a specific firm across several consecutive years, as we do here.

A relevant reference point for the dataset used in this paper is ENIA, the manufacturing sector survey for Chilean firms that has been widely used in the literature (see, for example, Gandhi et al. [2020], among many others). Similarly to this dataset, ENIA has rich information on production, investment and employment, but is silent regarding the firm's financial position, so it cannot be used to implement our framework. Interestingly, the OP estimates of the production function parameters

[31] As an alternative, we also use the information on tax form F29, which has monthly data on investment in machinery and equipment. The behavior of both investment series is very similar.

using our administrative set reported in the next section are similar to those that can be obtained using similar methodologies with ENIA. This provides a form of external validation to this dataset in the sense that it is associated to estimates for Chile that are quantitatively consistent with those obtained in the large literature that has used ENIA.

6.2 Empirical Results

We now use the data presented in the previous section to implement the empirical methodology discussed in Section 5. Following on our previous discussion, the goals of this section are twofold. First, to estimate firm level production functions, correctly accounting for the presence of financial frictions. Second, to provide an empirical characterization of investment and wealth accumulation policy functions at the firm level.

As presented earlier, our empirical strategy considers five alternative specifications:

(i) OP: a standard approach following the methodology in Olley-Pakes. which uses investment as an auxiliary equation to recover productivity, and provides a benchmark to previous literature. As argued earlier, this methodology, as well as similar methodologies that use alternative auxiliary equation such as intermediate inputs, are in general not robust to financial frictions.

(ii) OPA-Inv: augments the Olley-Pakes approach by including firm wealth in the auxiliary investment equation.

(iii) Proxy-Wealth: a proxy variable approach that uses the wealth accumulation equation as an auxiliary equation.

As discussed, methods (i), (ii), and (iii) assume absence of shocks in the policy functions, and rely on the ability of the auxiliary equations to perfectly recuperate the unobserved productivity (under a scalar unobservable assumption and a monotonicity assumption). Estimated parameters between these methods will differ depending on the presence and severity of financial constraints, and the extent to which the self financing channel that relates wealth accumulation decisions to productivity is active in the data. In contrast, the next 2 methodologies allow for shocks to the auxiliary equations:

(iv) Proxy-IV: uses both the investment equation and the wealth accumulation equation through an IV regression.

(v) SEM: a non-linear approach that uses the full information of both investment and wealth accumulation equations.

In a nutshell, our results show significant differences in the estimates across the different approaches, which fall in line with the predictions from the stylized model described in Section 2. Our results show that financial frictions in the data are relevant, and that their severity depends on available collateral in the form of the firm's net wealth. Estimated policy functions are consistent with theory, showing, for example, a weaker investment response to productivity innovations among firms that have less collateral and are therefore likely to be more constrained. Estimated policy functions also show a significant response of net wealth accumulation to productivity, in line with the predictions of the self financing channel. However, the persistence of significant differences in estimated production function parameters across methodologies suggest that self financing is not strong enough to eliminate the importance of frictions in investment and, through it, on production function estimations.

Finally, our novel empirical methodology also highlights the importance of a more thorough assessment of the policy functions. The estimates from OPA-IV and SEM support the notion that iid shocks to the policy functions are relevant, and confirm that the joint estimation of the policy functions improves the results.

6.2.1 Production Functions

As discussed in Section 2, we expect OP to underestimate the capital elasticity, and to overestimate the labor elasticity, as it incorrectly interprets differences in value added due to financial constraints as generated by differences in productivity, and not in capital. This bias in estimated productivity makes the co-movement of output and labor stronger than expected, generating a larger estimated elasticity of labor.

Table 1 presents the results of the first stage regressions of the production function estimation for methods (i) - (iv), in a simple case in which all specifications use a linear proxy for productivity.[32] Specifically, Table 1 displays the results of regressions y_{it} on l_{it}, k_{it}, and the productivity proxy in each method. The goal of this table is to give intuition on the information provided by each of the different productivity proxies used by the alternative methodologies.

The OP estimate in column 2, which controls for (i_{it}, k_{it}) as a proxy for productivity, delivers an estimate of β_l of 0.65. The OPA-Inv estimate in column 3, which adds a_{it} as a control, delivers a significantly lower β_l of 0.53. This indicates that firm net wealth is relevant, playing an important role in the investment decision, and therefore should be included when constructing the proxy variable for productivity. The estimate of Proxy-Wealth in Column 4, which controls for (a_{it+1}, k_{it}, a_{it}), delivers a β_l of 0.50. Therefore, a_{it+1} is significant, implying that the wealth accumulation policy function contains important information about productivity. While the estimates of β_l are similar between OPA-Inv and Proxy-Wealth, they are still statistically different. This suggests misspecification in the auxiliary equations, likely related to the exclusion of shocks.[33] Finally, column 5 displays the estimates of Proxy-IV, which exploits both policy rules and allows for shocks to these auxiliary equations. In this case, there is a relevant change in the estimate of β_l , which falls to 0.43.

Building on the insights provided by the initial exercise, Table 2 presents the results of the full estimation of the production function parameters (β_l, β_k) using methods (i)-(v). The estimates of β_l are different than the ones in table 1 because Table 2 uses more general non-linear policy functions.

[32] Recall that, in this stage, the coefficient on labor delivers the estimate of β_l, while the coefficient associated to β_k is recovered in the second stage.

[33] In fact, the difference in estimates persists even when we model nonlinear policies as in Table 2. Note that in the absence of shocks the two methods should deliver identical estimates.

	OP	OPA	Proxy-Wealth	Proxy-IV
l_{it}	0.67***	0.53***	0.50***	0.43***
i_{it}	0.07***	0.04***	-	-
a_{it+1}	-	-	0.24***	0.67***
a_{it}	-	0.36***	0.21***	-0.06***
k_{it}	0.03***	0.04***	0.07***	0.07***
Observations	13516	13516	13516	13516
Firms	4867	4867	4867	4867

TABLE 1: First Stage: Estimation of the Labor Elasticity in the Production Function

Similarly to the linear case, there are significant differences across estimators, with a general pattern that is consistent with the presence of financial constraints, and with the theoretical predictions derived earlier. Controlling for wealth in the policy functions allows us to discriminate between of productivity and the effects of collateral constraints. Labor elasticities exhibit the same pattern described in Table 1. The estimate of β_l is 0.65 for OP, and decreases for all the estimates that are robust to financial constraints: to 0.51 in OPA-Inv, to 0.48 in Proxy-Wealth, and to 0.43 in Proxy-IV and 0.44 in SME. Conversely, the opposite pattern holds for the elasticity of capital: the estimate of β_k is 0.35 for OP, and increases to 0.41 for OPA-Inv, to 0.43 for Proxy-Wealth, and to 0.45 for Proxy-IV and 0.43 to SME.

In addition, by relying on the co-movements between wealth accumulation and investment decisions, controlling for the current stock of net wealth allows us to disentangle productivity shocks from transitory shocks that can temporarily affect investment and saving decisions. The differences between the estimates of OPA-Inv and Proxy-Wealth from Proxy-IV and SME confirm that transitory shocks are relevant for the behavior of the policy functions.

Finally, the differences in the estimated input elasticities have relevant implications for the degree of returns to scale at the firm level, a crucial quantitative issue to understand aggregate dynamics. In particular, OP results are consistent with constant returns to scale, while OPA-Inv, Proxy-Wealth, Proxy-IV and SME all imply decreasing returns to scale. The estimate of Proxy-IV or SEM leads to a span of control around 0.87. This figure lies on the the upper-end of the range used in the related literature (for instance, Buera and Shin [2013a] use 0.79 while Restuccia and Rogerson [2008] and Midrigan and Xu [2014] use 0.85). This lower span of control implies a larger entrepreneurial income share that can be retained by firms, which allows for a faster accumulation of wealth to overcome financial constraints.

	OP	OPA	Proxy-Wealth	Proxy-IV	SEM
β_l	0.65	0.51	0.48	0.43	0.44
	0.008	*0.007*	*0.006*	*0.01*	*0.002*
β_k	0.35	0.41	0.43	0.45	0.43
	0.05	*0.04*	*0.04*	*0.01*	*0.001*
σ_ϵ	0.68	0.57	0.52	0.22	0.20
Observations	13516	13516	13516	13516	13516
Firms	4867	4867	4867	4867	4867

Table 2: Production Function

6.2.2 Productivity Process

Figure 1 depicts the productivity distribution across firms for OP, OPA-Inv, and SEM.[34] There are relevant differences in the dispersion of the estimated productivity distributions across methodologies. In OP, the standard deviation of productivity is 0.16, significantly lower than the 0.31 and 0.42 under OPA-Inv and SME, respectively (see Table 3). In particular, OP appears to significantly underestimate firm productivity in the upper half of the distribution. We find that the gap between ours and OP producticity estimates, i.e. the fraction by which true productivity is underestimated, is increasing in the productivity level of the firm (measured with either method). Specifically a firm that is 1% more productive under OP is on average 1.2% and 1.3% more productive under OPA and Proxy-Wealth, respectively. The fact that OP dampens productivity differentials across firms is once again consistent with the presence of financial frictions: as their actual investment is relatively low, OP underestimates the productivity of constrained high productivity firms. Conversely, the productivity of unconstrained but low productivity firms, which can invest comparatively more, is overestimated. Hence, by ignoring firm wealth, OP estimates a more compressed productivity distribution relative to the methods that are robust to frictions.

The compression of the productivity distribution under OP occurs despite the existence of the self-financing channel, since more productive firms actually finance a larger fraction of capital through their own assets. We explore more in detail this when we analyze the policy function, but for now we observe that the correlation between our estimates of productivity and the ratio between net assets and capital is roughly 0.3, which is consistent with the accumulation of profits by productive firms in order to minimize the adverse effects of frictions. However this endogenous reaction by firms is not strong enough to make the differences between the estimation methods insignificant.

Table 3 also presents results for productivity persistence. The first row displays the autocorrelation

[34] The estimated distributions of productivity for Proxy-Wealth and OPA-Inv are very similar, so we omit Proxy-Wealth from the table for the sake of brevity.

of the estimated productivity ρ_z.[35] We can see that estimated persistence is considerably lower under OP. The estimated value of ρ_z raises from 0.53 under OP to 0.7 in OPA and to 0.82 under SME, respectively. The persistence of the productivity process is a crucial parameter in quantitative models that assess the strength of the self-financing channel and the importance of financial frictions on aggregate productivity and misallocation. For instance, Moll [2014] shows that low persistence in productivity leads to large effects of financial frictions on aggregate TFP, as the self-financing channel is less powerful. This is the result of weaker incentives to wealth accumulation when positive productivity shocks are not expected to last for long.

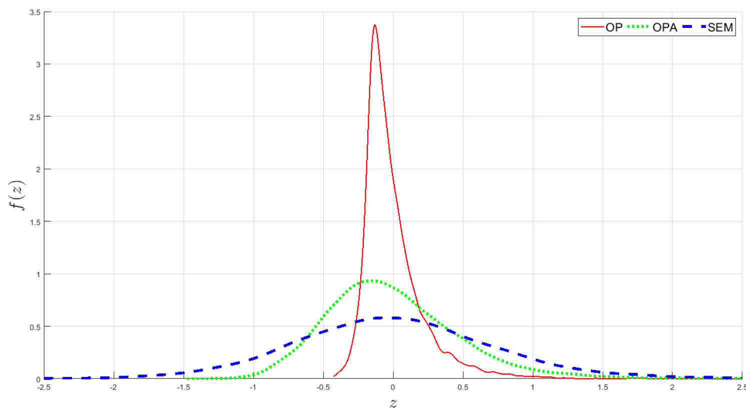

FIGURE 1: Estimated distribution of productivities

Notes: The figure exhibits the distribution of productivities estimated by methods: (i) OP, (ii) OPA-Inv, and (v) SEM.

	OP	OPA	Proxy-Wealth	Proxy-IV	SEM
ρ_z	0.53	0.70	0.83	0.87	0.82
	0.01	*0.01*	*0.01*	*0.01*	*0.01*
σ_η	0.16	0.31	0.26	0.28	0.42
Observations	13516	13516	13516	13516	13516
Firms	4867	4867	4867	4867	4867
R^2	0.37	0.53	0.74	-	0.67

TABLE 3: Productivity Process

[35]The estimation procedure does not assume an AR(1) process for productivity.

6.2.3 Policy Functions

We now present the estimated policy functions, one of the main goals of our empirical exercise. As discussed earlier, the literature on production function estimations has typically used the policy rules as auxiliary equations to control for unobserved productivity, but has not analyzed them as objects of interest by themselves, as their focus lies on the production function parameters and the productivity process.

Given our interest in understanding the role of financial frictions and the self financing channel, we pay special interest to the estimation of policy functions, and the analysis of the economic forces that underlie them.

6.2.4 Investment Policy Function

Linear effects We get a first glimpse through the estimates of the OPA-IV approach, as presented in Figure 8. This method assumes that the effect of productivity on investment is linear, but allows this elasticity to vary across firms with different levels of net wealth and capital. Panel (a) of Figure 8 depicts the estimated marginal effect of productivity on investment as a function of the wealth level a_t. The relationship between the marginal effect and wealth is presented for three levels of capital k_t, associated to the 10th, 50th, and 90th percentiles of the capital distribution. As expected, the marginal effect is positive and significant for almost all of the possible combinations of the state variables: higher productivity boosts investment, as the optimal level of capital grows.[36]

For all capital levels, the marginal effect of productivity on investment is monotonically increasing in wealth. In all cases, elasticities at the lowest levels of wealth are significantly lower than those for firms with the highest wealth. For instance, for firms at the 90th percentile of capital, the elasticity of investment to income shocks almost doubles when we move from firms at the bottom to the top of the wealth distribution. The increase in the elasticity for firms at the 50th, and 10th percentiles of the capital distribution are 32% and 25%, respectively. This is consistent with the notion that firms with higher wealth face softer financial constraints, allowing them to adjust capital more in response to productivity changes. On the other hand, for a given level of wealth, the marginal effect is decreasing in capital. Differences are large: elasticities for firms in the 10th capital percentile are three times larger than elasticities in the 90th percentile. This could simply reflect the existence of adjustment costs in capital, which might be relatively more important for firms that are already close to their optimal capital level, or could also be a manifestation of financial constraints, as for a given level of wealth firms with more capital have higher leverage. Panel (b) repeats the exercise for the marginal effect of wealth shocks. The response of investment to wealth is decreasing in the level of wealth and increasing in firms' capital, although differences along both dimensions are more modest than in the case of productivity.

These results are once again consistent with the notion of financial constraints, and the implications of models in the spirit of the one described in Section 2. All else equal, firms with low levels of wealth

[36] Estimates with confidence bands are presented in Appendix A.5

are more constrained. Therefore, the marginal value of increasing their collateral is larger, as it allows them to increase investment more significantly. Similarly, for a given level of wealth, firms with more capital are more leveraged, so an increase in wealth also has a stronger effect on alleviating their financial constraints.

Nonlinear effects Figure 3 displays the estimated average derivative effect of productivity on investment $\hat{\Phi}_t^h(a, k, z)$ using method (v) SEM. This method allows the investment policy function to be non-linear on productivity z_t. Therefore, the three-dimensional graphs show how the elasticity changes for different combinations of a_t and z_t. Each of the four panels, a) to d), in Figure 3 present the results for different percentiles of the k_t distribution: 25th, 50th, 75th and 90th.[37] In general, the same patterns seen in Figure 8 hold. First, for a given z_t, effects are typically increasing in a_t. Second, for a given combination of a_t, and z_t, the derivative effects are strictly decreasing in capital k_t.

The main innovation of the non-linear method is that it can be used to analyze how the pattern varies as we change z_t. In general, the sensitivity of investment to productivity shocks increases with z_t. This is, for given values of wealth and capital, investment responses to productivity shocks are larger for ex-ante more productive firms.

The general pattern is coherent with the implications of models of financial constraints in which firm productivity can affect firm lending contracts and the amount of borrowing, as it is the case in the models of Aguirre [2017], DeMarzo and Fishman [2007], Brooks and Dovis [2020] or Lian and Ma [2020], in which firms can use their future cash-flows as collateral. Figure 3 suggests that financial frictions might be very nonlinear in a, k and z. For instance, for very productive firms at the top of the productivity distribution, but with capital below the median (panel a and b), the marginal effect is at its highest, and is not sensitive to the level of collateral. However, the response of firms with the same level of capital but productivity below the median of the productivity distribution, is sensitive (and increasing) to collateral. On the other hand, for all productivity levels, the response of investment in firms that have capital above the median is increasing in wealth. Interestingly, for firms in the 90th percentile of the capital distribution, the marginal effect of productivity shocks is close to zero at low levels of wealth, regardless of their initial level of productivity. A potential explanation is that future cash flows cannot overcome financial frictions in highly-leveraged firms, so the collateral constraint plays a key role. As wealth increases, the elasticity becomes more significant. This increase is more pronounced for more productive firms, for which the elasticity increases up to almost 0.20.

6.2.5 Wealth Accumulation Policy Function

Linear effects We now turn our attention to the estimates of the wealth accumulation function following the same strategy as in the previous case. Figure 4 depicts the marginal effects of productivity and wealth shocks on wealth accumulation under the OPA-IV approach. Panel (a) displays the estimated marginal effect of productivity on wealth accumulation as a function of wealth a_t, and for the 25th, 50th, and 90th percentiles of capital k_t. The marginal effect is positive and significant

[37] Confidence intervals are presented in Appendix A.5.

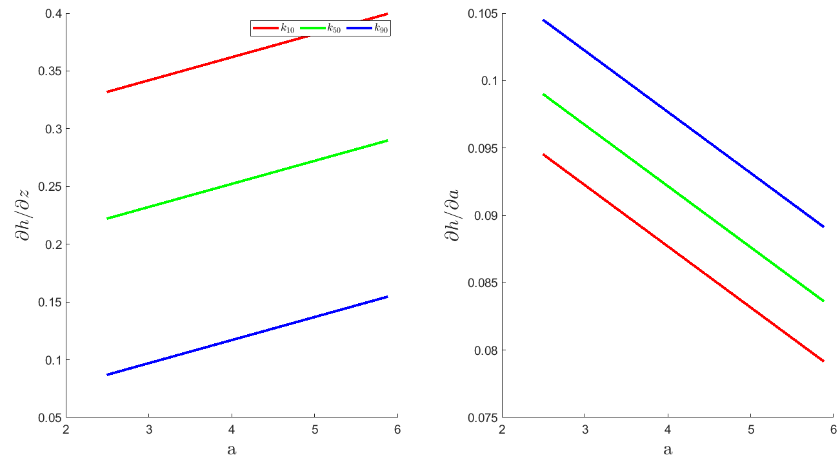

FIGURE 2: Marginal effect of productivity and wealth on investment

Notes: Panel (a) of the figure exhibits the estimated derivative effect of productivity in the investment policy function using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital. Panel (b) of the figure exhibits the estimated derivative effect of the stock of wealth (previous wealth) in the investment policy function using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital.

for almost all of the possible combinations of state variables.[38] In contrast to the investment policy function, the effect of productivity on the wealth accumulation policy is now strictly decreasing on wealth, for all levels of capital: the elasticity falls in 10 percentage points as we move along the wealth distribution. This decrease represents a reduction of around 25% of the elasticity of productivity on savings. Once again, this result is qualitatively consistent with insights of models with financial constraints, and in particular the self-financing channel. Firms that have less collateral, and therefore more likely to be constrained, have an stronger incentive to boost up their savings in the presence of persistent productivity shocks, in order to finance future investments. The figure also shows that, for a given level of wealth, the marginal effect is increasing in capital, although differences across capital percentiles are relatively modest. The interpretation once again relates to the self-financing channel, and the additional incentives to save for firms with higher leverage.

Panel (b) displays the marginal effect of wealth a_t on the next period wealth stock a_{t+1} - the conditional persistence in the wealth accumulation equation-. We can see that the elasticity is increasing in wealth (therefore, the elasticity is non-linear: wealth is more persistent at higher levels of current wealth), but decreasing in capital.

Nonlinear effects Figure 10 displays the estimated average derivative effect of productivity on wealth accumulation $\hat{\Phi}^{g}_{t+1}(a, k, z)$ using method (v) SEM. As before, this method allows the wealth

[38]Once again, confidence intervals are presented in Appendix A.5

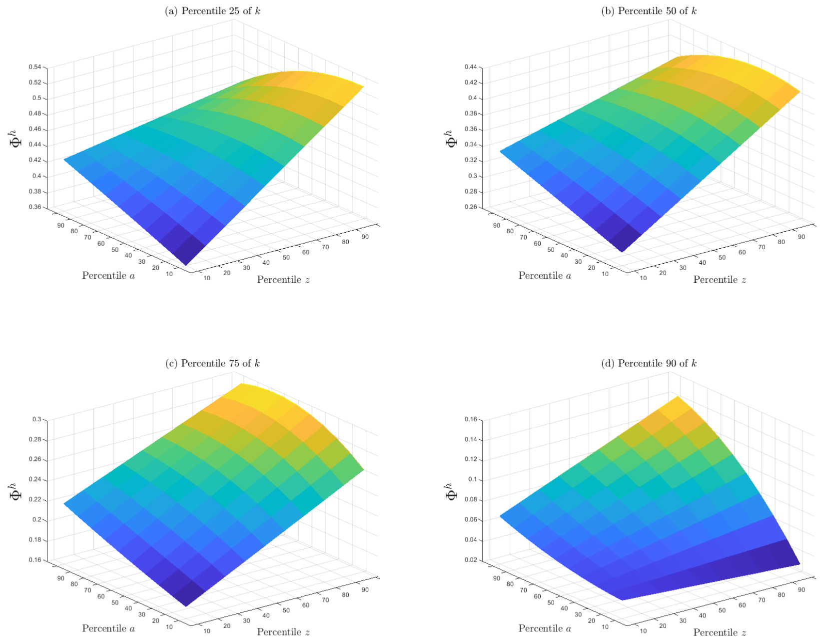

FIGURE 3: Marginal effect of productivity on investment

Notes: The figure exhibits the estimated derivative effect of productivity over investment using the SEM method. The estimated model is highly non-linear, so the figure displays the marginal effect for different values of productivity and stock of wealth, keeping the stock of capital at its 25th percentile (a), 50th percentile (b), 75th percentile (c), 90th percentile (d).

accumulation policy function to be non-linear on productivity z_t. Hence, the three-dimensional graph presents how this elasticity changes for different combinations of a_t and z_t, for four different levels of capital in panels a) to d). We can see that the general patterns from figure 4 hold. In almost all cases, the average derivative effect of productivity on savings decreases with a_t. Similarly, for a given combination of wealth and productivity, in most cases elasticities are increasing in capital, consistent with the theoretical impact of larger leverage.

Regarding non-linearities, for a given level of capital, elasticities are the largest in firms that are highly productive but hold little wealth. This is once again consistent with the insights of our stylized model, as those firms are likely to be the ones that are most constrained, and where the value of an additional unit of savings in response to a positive productivity shock is the largest. In fact, the elasticity to productivity of firms on the upper end of the productivity distribution and the lower end of the wealth distribution is close to 1 across all capital percentiles. This is, the savings propensity to income shocks of highly productive but severely constrained firms is roughly 1, as the value of alleviating the constraint is comparatively large. In contrast, this propensity falls significantly for firms that, while also having low-wealth, have low productivity, and therefore probably less constrained. In general, as wealth increases, the propensity to save decreases, across all productivity levels.

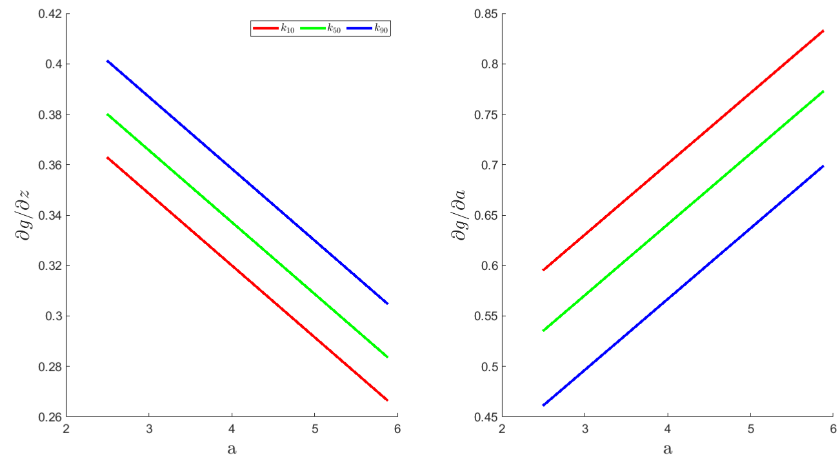

Figure 4: Marginal effect of productivity on wealth accumulation

Notes: Panel (a) of the figure exhibits the estimated derivative effect of productivity in the wealth accumulation policy function using the proxy-IV approach. The figure displays how the effects changes along different values of the stock of wealth and is evaluated at three different level of stock of capital. Panel (b) of the figure exhibits the estimated derivative effect of the stock of wealth (previous wealth) in the wealth accumulation policy function using the proxy-IV approach. The figure displays how the effects changes along different values of the stock of wealth and is evaluated at three different level of stock of capital.

6.2.6 An application to the self-financing channel

To get a more direct appraisal of the implications of our estimated policy functions for the self-financing channel, we use our data and estimates to look at the convergence of the marginal product of capital (MPK) between constrained and unconstrained firms in the spirit of the exercise in Banerjee and Moll [2010].

To do so, we use the data to calculate capital in firms that share the same level of initial productivity but have different levels of initial wealth. We then use the estimated policy functions to simulate the evolution of their capital, labor, and wealth across time, assuming that productivity is constant and there are no additional shocks. Using the estimated production function parameters, we calculate the evolution of the MPK associated with the simulated capital path.

Results are presented in Figure 6. For each row, the graphs plot the evolution across time of the marginal product of capital for a firm that starts on the lower end of the wealth distribution (10th percentile) vis-a-vis firms with the same constant level of productivity z, but larger levels of initial wealth (50th percentile in the first column, 75th percentile in the second, 90th in the third). We report the convergence in MPKs between a constrained and unconstrained firm for three different productivity scenarios. The first row depicts firms in the 10th percentile of the productivity distribution, while the 50th and 90th productivity deciles are presented in the second and third rows.

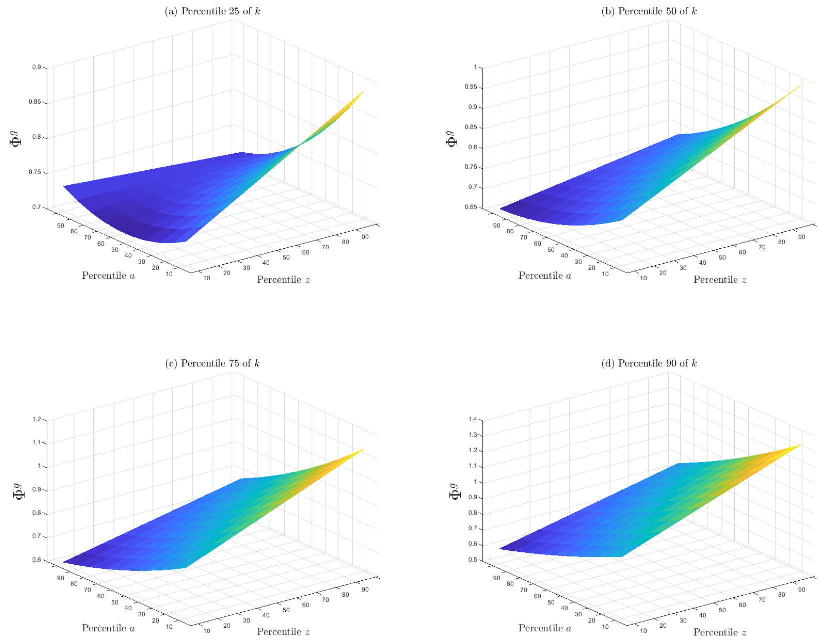

FIGURE 5: Marginal effect of productivity on wealth accumulation

Notes: The figure exhibits the estimated derivative effect of productivity in the wealth accumulation policy function using the SEM method. The estimated model is highly non-linear, so the figure displays the marginal effect for different values of productivity and stock of wealth, keeping the stock of capital at its 25th percentile (a), 50th percentile (b), 75th percentile (c), 90th percentile (d).

Consistent with the self-financing channel, low-wealth, constrained firms are able to increase their capital stock across time, such that the marginal product of capital converges towards that of firms with similar firm productivity z but higher levels of initial wealth a_0. Convergence, however, is relatively slow, and marginal productivity gaps persist for decades. For example, across all three productivity levels, the marginal product of capital in a firm with initial wealth in the 10th percentile of the wealth distribution is close to three times larger than in a firm in the 90th wealth percentile. While this gap closes steadily across the years, marginal products in low wealth firms are still at least double those of high wealth firms after one decade. The speed of convergence in our data is much slower than in Banerjee and Moll [2010], where, for a similar initial gap, differences in marginal product between constrained and unconstrained firms vanish in less than a decade. For example, among firms in the 90th percentile of the productivity distribution, convergence in the marginal product of capital between firms in the 10th and 90th wealth percentiles takes more than 40 years, although half of the initial gap disappears after ten years.

Therefore, our results indicate that while the self-financing channel plays an important role in reducing productivity gaps and the extent of misallocation in this context, it cannot offset the persistence of significant differentials in marginal productivity over the medium term.

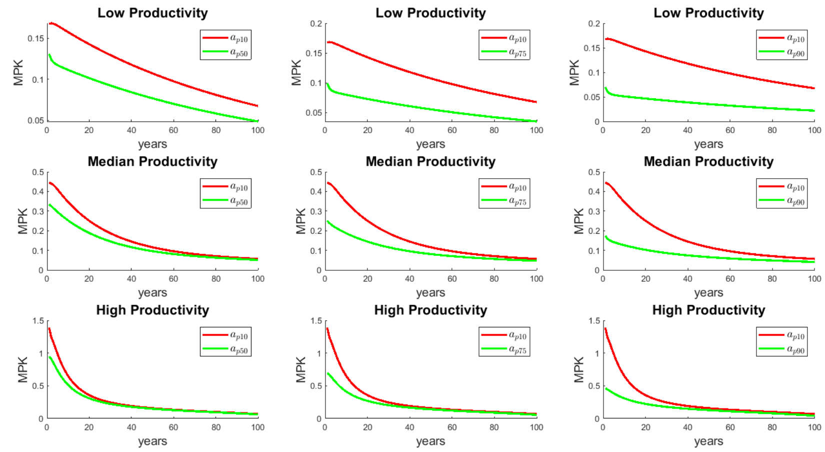

Figure 6: Convergence in the marginal product of capital across firms

Notes: The figure exhibits the simulated evolution of the marginal product of capital for firms with different levels of initial productivity and wealth. Low wealth firms (10th percentile) are depicted in red, while high wealth firms (50th percentile in column 1, 75th in column 2, 90th in column 3) are depicted in green. The first row presents firms in the 10th percentile of the productivity distribution, while the second and third row presents figures in the 50th and 90th productivity deciles. The simulation uses the estimated production function and investment and wealth accumulation policy functions, holding firm productivity constant.

6.3 Estimations using Simulated Data

We conclude this section by using an extended version of the stylized model presented in section 2 to generate data that is consistent with the theoretical framework that explicitly accounts for collateral constraints. We use this data to provides a validation of our proposed empirical specification.

The spirit of the model and its theoretical implications are very similar to those of the model presented in Section 2, although we generalize it in two dimensions. First, we no longer impose linearity in preferences and assume a CRRA utility function with risk aversion coefficient σ.[39] Second, we introduce adjustment costs to capital. We choose a standard quadratic function with a parameter η determining its size.[40] Note that the introduction of adjustment costs implies that capital is a state variable, as in our empirical estimations.

We assume a specific functional form for the general collateral constraint described in Section 2:

$$\kappa(A_{it}, Z_{it}) = (\lambda + \lambda_z(z_{it} - \bar{z}))A_{it}$$

where λ and λ_z are constants, z_{it} is the log of Z_{it} and $\bar{z}$ its mean, and we impose $\lambda+\lambda_z(\min(z_{it})-\bar{z}) \geq 1$. Thus, for a given level of collateral, the capital to assets ratio is strictly increasing in productivity.

[39]We remove the convex function $g(\cdot)$ included in Section 2, as it is no longer needed to have an interior solution.

[40]Specifically we use $\eta(I_{it}/K_{it})^2 K_{it}$

In line with the estimates in Section 6.2, we set $\beta_k = 0.43$ and $\beta_l = 0.44$ in the calibrated model. This implies a span of control parameter of 0.87. In the case of the productivity process we impose a linear Markov process, $z_{t+1} = \rho z_t + \mu_t$ and, consistent with our estimations, set $\rho = 0.82$ and $\sigma_\mu = 0.42$. We calibrate three key parameters to match certain moments of the sample. These parameters are the ones defining the strength of the collateral constraint (λ and λ_z) and the one determining the relevance of adjustment costs η. The moments we use to calibrate them are the mean capital to output ratio, which is 1.69, the net assets to output ratio, which is 0.89, and the correlation between productivity and the net assets to capital ratio, which is 0.3. For the rest of the parameters we use standard values: discount factor $\beta = 0.8$, risk aversion coefficient $\sigma = 0.2$, depreciation rate $\delta = 0.1$ and interest rate $r = 4\%$.

We use the calibrated model to generate simulated data and use that data to replicate the empirical estimations of the previous section.[41] Table 4 presents the estimates, using the data generated by the simulated model, in the case in which there are no shocks to the investment and wealth accumulation decisions. We can see that, consistent with the theoretical predictions in Sections 2 and 3, the existence of financial frictions to capital accumulation implies that OP fails to deliver the true underlying parameters for the firm's production function. As expected, OP overestimates the elasticity of labor, and underestimates the elasticity of capital. Estimates for both OPA-Inv and Proxy Wealth, on the other hand, are consistent with the true values of β_K and β_L.

Table 5 presents the estimates for simulated data allowing for shocks on investment and wealth accumulation. As expected, in the presence of shocks to the policy functions, both OPA- and Proxy Wealth fail, and deliver biased estimates just like OP. However, estimators Proxy IV and SEM, which explicitly allow for policy function shocks, recover the true underlying parameters.

Therefore, data generated from a quantitative model, which explicitly includes financial frictions and the theoretical mechanisms described in Section 2, provides validation to our insights regarding the biases of traditional methodologies in the presence of financial constraints, as well as for the validity of using methodologies that account for firm wealth and allow for shocks in the policy functions.

	OP	OPA	Proxy-Wealth
β_l	0.505	0.428	0.446
β_k	0.397	0.416	0.424

Table 4: Estimates Using Simulated Data: Policy Functions without Shocks in the Policies

Model simulations can also be used to explore how the biases of the production function estimates

[41] Following Ackerberg et al. [2015] we introduce iid shocks to wages. This generates extra variability on labor that is not due to variation in the state variables, allowing us to identify β_l in the first stage.

	OP	OPA	Proxy-Wealth	Proxy-IV	SEM
β_l	0.543	0.540	0.514	0.443	0.442
β_k	0.503	0.500	0.553	0.424	0.431

Table 5: Estimates Using Simulated Data: Policy Functions with Shocks in the Policies

vary with the intensity of financial frictions, i.e. with different values of the parameters governing the collateral constraint, λ y λ_z. When λ decreases from 2.5, which is the value found in the calibrated version of the model, to 2, the OP bias in β_l grows from 0.07 to 0.11, and the OP bias in β_k goes from -0.04 to -0.06. Interestingly, neither the OPA nor Proxy-Wealth coefficients are affected. This confirms that these estimators are robust to financial frictions in the absence of stochastic shocks in the policies.

When we make collateral constraints more severe through a change in λ_z, the effects on OP are similar. When λ_z goes from 0.5, the value found in the calibration, to 0, the biases for β_l y β_k increase to 0.17 and -0.07, respectively. Unlike the case of a change in λ, now the OPA estimators also move away from the real coefficients. The bias in β_l is 0.08 and the one in β_k is -0.05. This happens because OPA requires that capital varies for constrained firms for a given level of assets. Note, however, that this estimator still performs better than OP. Finally, as expected, the Proxy-Wealth estimates are not sensitive to λ_z, the same result we observed for λ.

7 Conclusions

We provide an empirical analysis of wealth accumulation and investment dynamics in firms that operate under financial frictions, and how these decisions relate to the unobservable firm's productivity process. We argue that standard approaches to recover productivity process from production function estimations fail under the presence of financial frictions that limit the firm's ability to hire inputs, as the auxiliary equations used to characterize input decisions do not hold. For instance, in the case of the OP estimator, the auxiliary investment equation does not account for wealth, a relevant variable for capital decisions in macro models of with financial constraint such as Moll [2014] and Buera and Shin [2013b]. We argue that this renders a considerable bias in the estimation of the parameters of the firm's production function and, therefore, in the estimation of the characteristics of the productivity process.

As an alternative, we extend the OP approach to account for financial frictions, introducing wealth and unobservable firm-specific shocks in the investment demand function. This flexible framework allows us to jointly model and estimate the firm wealth accumulation dynamics, its investment decisions and the unobservable productivity process.

Our results, using Chilean manufacturing data, show that the estimated capital elasticity in the production function increases from 0.35 when using OP to 0.43 when we estimate a model that allows for financial frictions. In contrast, the labor elasticity in the production function decreases from 0.65 in OP to 0.44 when we use our estimator that is robust to financial frictions. We replicate these patterns using simulated data generated by a quantitative macro model that explicitly includes collateral constraints. We also show that OP underestimates the dispersion in productivities significantly relative to our method.

We use our setup to provide a detailed analysis of the firm´s policy functions, with a particular interest in understanding the mechanics of the self-financing channel. We show that, consistent with theoretical predictions in the presence of financial frictions, the marginal effect of productivity on investment is increasing in wealth and decreasing in capital. We also find a positive and significant marginal effect of productivity on wealth accumulation, stronger for more constrained firms, which provides support to the existence of an active self-financing channel. We also use our estimated empirical model to measure the power of self-financing on reducing misallocation by studying the convergence of MPKs of two firms with the same productivity but with different levels of financial frictions. We show that the MPKs of these firms converge over time, although the convergence is not fast and takes time. For instance, when we compare firms at the 10th-percentile with firms at the 90th-percentile of the wealth distribution, the MPK of poor firms is around three times the MPK of wealthy firms at the initial period, and it takes more than 40 years to see convergence in their MPKs. Still, half of the initial gap in their MPKs disappears after ten years.

References

Daniel A Ackerberg, Kevin Caves, and Garth Frazer. Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451, 2015.

Alvaro Aguirre. Contracting institutions and economic growth. *Review of Economic Dynamics*, 24: 192–217, 2017.

Heitor Almeida, Murillo Campello, and Michael Weisbach. The cash flow sensitivity of cash. *Journal of Finance*, 59:1777–1804, 2004.

Isaiah Andrews, Matthew Gentzkow, and Jesse M Shapiro. Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592, 2017.

Isaiah Andrews, Matthew Gentzkow, and Jesse M Shapiro. Transparency in structural research. *Journal of Business & Economic Statistics*, 38(4):711–722, 2020.

Manuel Arellano. Uncertainty, persistence, and heterogeneity: A panel data perspective. *Journal of the European Economic Association*, 12(5):1127–1153, 2014.

Manuel Arellano and Stéphane Bonhomme. Nonlinear panel data estimation via quantile regressions. *Econometrics Journal*, 19(3):C61–C94, 2016.

Manuel Arellano and Stéphane Bonhomme. Nonlinear panel data methods for dynamic heterogeneous agent models. *Annual Review of Economics*, 9:471–496, 2017.

Manuel Arellano, Richard Blundell, and Stéphane Bonhomme. Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica*, 85(3):693–734, 2017.

Rüdiger Bachmann, Ricardo J Caballero, and Eduardo MRA Engel. Aggregate implications of lumpy investment: new evidence and a dsge model. *American Economic Journal: Macroeconomics*, 5(4): 29–67, 2013.

Abhijit Banerjee and Benjamin Moll. Why does misallocation persist? *American Economic Journal: Macroeconomics*, 2:189–206, 2010.

Ben S Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393, 1999.

Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921, 2008.

Steve Bond, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 2021.

Stéphane Bonhomme. Discussion of "transparency in structural research" by isaiah andrews, matthew gentzkow, and jesse shapiro. *Journal of Business & Economic Statistics*, 38(4):723–725, 2020.

Wyatt Brooks and Alessandro Dovis. Credit market frictions and trade liberalizations. *Journal of Monetary Economics*, 111:32–47, 2020.

Francisco J Buera and Yongseok Shin. Self-insurance vs. self-financing: A welfare analysis of the persistence of shocks. *Journal of Economic Theory*, 146(3):845–862, 2011.

Francisco J Buera and Yongseok Shin. Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, 121(2):221–272, 2013a.

Francisco J Buera and Yongseok Shin. Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, 121(2):221–272, 2013b.

Francisco J Buera, Joseph P Kaboski, and Yongseok Shin. Finance and development: A tale of two sectors. *The American Economic Review*, 101(5):1964–2002, 2011.

Francisco J Buera, Joseph P Kaboski, and Yongseok Shin. Entrepreneurship and financial frictions: A macro-development perspective. *Annual Review of Economics*, 2015.

Francisco J Buera, Joseph P Kaboski, and Robert M Townsend. From micro to macro development. 2021.

Andrea Caggese and Vicente Cuñat. Financing constraints, firm dynamics, export decisions, and aggregate productivity. *Review of Economic Dynamics*, 16(1):177–193, 2013.

Jan De Loecker. Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451, 2011a.

Jan De Loecker. Recovering markups from production data. *International Journal of Industrial Organization*, 29(3):350–355, 2011b.

Peter M DeMarzo and Michael J Fishman. Optimal long-term financial contracting. *The Review of Financial Studies*, 20(6):2079–2128, 2007.

Ulrich Doraszelski and Jordi Jaumandreu. R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies*, 80(4):1338–1383, 2013.

Ulrich Doraszelski and Jordi Jaumandreu. Measuring the bias of technological change. *Journal of Political Economy*, 126(3):1027–1084, 2018.

Steven Fazzari, R Glenn Hubbard, and Bruce C Petersen. Financing constraints and corporate investment. Technical report, National Bureau of Economic Research, 1987.

Amit Gandhi, Salvador Navarro, and David A Rivers. On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016, 2020.

Neus Herranz, Stefan Krasa, and Anne P Villamil. Entrepreneurs, risk aversion, and dynamic firms. *Journal of Political Economy*, 123(5):1133–1176, 2015.

Hugo A Hopenhayn. Firms, misallocation, and aggregate productivity: A review. *Annu. Rev. Econ.*, 6(1):735–770, 2014.

Yingyao Hu and Susanne M Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008.

Yingyao Hu and Matthew Shum. Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1):32–44, 2012.

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

Greg Kaplan and Giovanni L Violante. How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics*, 2(4):53–87, 2010.

Aubhik Khan and Julia K Thomas. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, 121(6):1055–1107, 2013.

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

Chen Lian and Yueran Ma. Anatomy of corporate borrowing constraints. *The Quarterly Journal of Economics*, 136(1):229–291, 2020.

Kalina Manova. Credit constraints, heterogeneous firms, and international trade. *Review of Economic Studies*, 80(2):711–744, 2013.

Virgiliu Midrigan and Daniel Yi Xu. Finance and misallocation: Evidence from plant-level data. *The American Economic Review*, 104(2):422–458, 2014.

Benjamin Moll. Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221, 2014.

Emi Nakamura and Jón Steinsson. Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86, 2018.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Søren Feodor Nielsen et al. The stochastic em algorithm: estimation and asymptotic results. *Bernoulli*, 6(3):457–489, 2000.

Steven Olley and Ariel Pakes. The dynamics of productivity in the telecomunications equipment industry. *Econometrica*, 64:1263–97, 1996.

Tim Opler, Lee Pinkowitz, Rene Stultz, and Rohan Williamson. The determinants and implications of corporate cash holdings. *Journal of Financial Economics*, 52:3–46, 1999.

Vincenzo Quadrini. Entrepreneurship, saving, and social mobility. *Review of economic dynamics*, 3 (1):1–40, 2000.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720, 2008.

Ajay Shenoy. Estimating the production function under input market frictions. *Review of Economics and Statistics*, pages 1–45, 2020.

Zheng Song, Kjetil Storesletten, and Fabrizio Zilibotti. Growing like china. *American economic review*, 101(1):196–233, 2011.

Ludwig Straub. Consumption, savings, and the distribution of permanent income. *Unpublished manuscript, Harvard University*, 2019.

Chad Syverson. What drives productivity differences? *Journal of Economic Literature*, 49(2):326–365, 2011.

Lucciano Villacorta. Estimating country heterogeneity in capital-labor substitution using panel data. Technical report, Mimeo, 2018.

Appendix A.1

Here we discuss theorem 1 and show that β_k, β_l, $\varphi(z_{it-1})$, h_t, g_{t+1} are identified from data on y_{it}, k_{it}, l_{it}, i_{it}, a_{it} for $T \geq 4$ in a sequential way. First, we establish identification of the parameters of the production function. Second once β_k and β_l are identified, we show that the joint and marginal distributions of the productivity process are identify from the time series dependence structure of the net income process. Finally, once the conditional distribution of the productivity process given the firm net income process is identified, we show that h_t, g_{t+1} are identified.

Step 1: Production function Using the conditional independence assumption in *assumption 1* we can write the following conditional distribution of the observed variables $f(y_{it}, i_{it} \mid a_{it+1}, X_{it})$ in terms of some pieces of the model:

$$f(y_{it}, i_{it} \mid a_{it+1}, X_{it}) = \int f(y_{it} \mid z_{it}, i_{it}, a_{it+1}, X_{it})\, f(i_{it} \mid z_{it}, a_{it+1}, X_{it})\, f(z_{it} \mid a_{it+1}, X_{it})\, dz_{it}, \tag{46}$$

where $f(y_{it} \mid z_{it}, k_{it}, l_{it})$ is the conditional distribution of the production function. From assumption 1, ε_{it}, v_{it}, and w_{it+1} are independent conditional on $(l_{it}, k_{it}, a_{it,} z_{it})$, which can be interpreted as the exclusion restrictions in a nonlinear IV setting. Thus, we have that $f(y_{it} \mid z_{it}, i_{it}, a_{it+1}, X_{it}) = f(y_{it} \mid z_{it}, k_{it}, l_{it})$ and $f(i_{it} \mid z_{it}, a_{it+1}, X_{it}) = f(i_{it} \mid z_{it}, X_{it})$, and we can re-write (46) as

$$f(y_{it}, i_{it} \mid a_{it+1}, X_{it}) = \int f(y_{it} \mid z_{it}, k_{it}, l_{it})\, f(i_{it} \mid z_{it}, X_{it})\, f(z_{it} \mid a_{it+1}, X_{it})\, dz_{it} \tag{47}$$

Now, the identification challenge is to recover the latent conditional density of the production function $f(y_{it} \mid z_{it}, k_{it}, l_{it})$ given the observed conditional density $f(y_{it}, i_{it} \mid a_{it+1}, X_{it})$. We notice that given assumption 1 and the structure of our dynamic model, our setup can be framed into the setup studied in Hu and Schennach [2008] and Hu et al. [2020]. Hence, Theorem 1 of Hu and Schennach [2008] can be applied to our setting to show that $f(y_{it} \mid z_{it}, k_{it}, l_{it})$ is identified from the data. Once we identify $f(y_{it} \mid z_{it}, k_{it}, l_{it})$ we can construct $E[y_{it} \mid z_{it} = 0, k_{it}, l_{it}] = \beta_l l_{it} + \beta_k k_{it}$ and identify β_k, β_l with a regression between $E[y_{it} \mid z_{it} = 0, k_{it}, l_{it}]$ and (l_{it}, k_{it}) as in theorem 1 in Hu et al. [2020].

Discussion: An important difference of our framework from Hu et al. [2020] is that our model with financial frictions provides a policy rule (the self-financing channel) that connects the latent productivity with an observed variable a_{it+1} that is not directly linked to the production function regression (i.e a_{it+1} is not an input in the production function regression). Hence, we do not have to use the policy rule in $t+1$ to avoid collinearity between inputs and therefore k_{t+1} is not part of the covariates in X_t. This allow us to have a standard law of motions for capital as in Olley and Pakes [1996] and Ackerberg et al. [2015] without the need of an unobserved component affecting the law of motion of capital. The latter is particularly important in applied work because most of the cases the researcher do not have data on both capital and investment separately and use the perpetual inventory method to recover the capital series from investment or vice-versa.

We then have the following result, which is a direct application of theorem 1 in Hu and Schennach [2008] and theorem 1 in Hu et al. [2020].

Proposition 1. *Under the conditional independence assumption in assumption 1, the high-level conditions in condition (1) (i)-(iv), and the assumption that the ε_{it} has mean zero, β_l and β_k are identified from the observed density $f(y_{it}, i_{it} \mid a_{it+1}, X_{it})$*

To show how theorem 1 of Hu and Schennach [2008] can be applied to our setup, we will follow their paper and define the integral operators and show that it admits an eigenvalue-eigenvector decomposition that can be learned from data. Then, to build intuition and remark the importance of the wealth accumulation equation, we will make a connection with the IV setup discussed in the linear model. Lets define $L_{y;I|a,X}$ as the integral operator such that $L_{y;I|a,X} = \int f(y_t, i_t \mid a_{t+1}, X_t) \, p(a_{t+1} \mid X_t) \, da$ and $D_{I;z|X}$ is a "diagonal" matrix operator mapping the function $g(z \mid X)$ to the function to the function $f(i_t \mid z_t, X_t) \, g(z \mid X)$ for a given value of investment i. Analogously, $L_{y|z,k,l}$ and $L_{a|z,X}$ are the integral operators associated with the conditional densities $f(y_t \mid z_t, k_t, l_t)$ and $f(a_{t+1} \mid z_t, X_t)$, respectively. Equation (47) can be expressed in terms of integral operators:

$$L_{y;I|a,X} = L_{y|z,k,l} D_{I;z|X} L_{z|a,X} \tag{48}$$

Integrating both sides of (47) with respect to I:

$$L_{y|a,X} = L_{y|z,k,l} L_{z|a,X} \tag{49}$$

From (49), we can see that the identification of $L_{y|z,k,l} = L_{y|a,X} L_{z|a,X}^{-1}$, our object of interest, has the form of an IV regression where a_{it} is the instrument for the endogenous variable z_{it} after controlling for covariates in X_{it}. This type of IV approach is unfeasible because z_{it} is unobservable. However, replacing (49) in (48) we get:

$$L_{y;I|a,X} L_{y|a,X}^{-1} = L_{y|z,k,l} D_{I;z|X} L_{y|z,k,l}^{-1} \tag{50}$$

Note that the observed quantity $L_{y;I|a,X} L_{y|a,X}^{-1}$ in (50) admits an eigenvector-eigenvalue decomposition $L_{y|z,k,l} D_{I;z|X} L_{y|z,k,l}^{-1}$. Therefore, $L_{y|z,k,l}$ is identify as the eigenvector of $L_{y;I|a,X} L_{y|a,X}^{-1}$ of (50). If $L_{y|z,k,l}$ is identify, then $f(y_t \mid z_t, k_t, l_t)$ is identify.

Rank Condition (Injectivity) To identify $L_{y|z,k,l}$ from (50), the inverse of $L_{y|a,X}$ has to exist. Looking at (49) we can show that $L_{y|a,X}$ has an inverse if $L_{y|z,k,l}$ and $L_{a|z,X}$ are invertible. Given the linearity and additivity of the Cobb Douglas production function in logs, the assumption that the characteristic function of ε_{it} has no zeros on the real line will ensure inyectivity (and invertibility) of $L_{y|z,k,l}$. The operator $L_{a|z,X}$ is injective (and invertible) if there is sufficient variation in the densities $f(a_{t+1} \mid z_t, a_t, k_t)$ for different values of z_{it}. The condition for $f(a_{t+1} \mid z_t, a_t, k_t)$ requires an statistical dependence between wealth accumulation a_{it+1} and productivity z_{it} conditioned on the observed state variables. This requirement can be met by the self-financing channel in equation (17) which implies a positive relationship between productivity and wealth accumulation for all constrained and non-constrained firms. In the IV terminology, the later is a relevance condition, that ensures that a_{it+1} is valid instrument for z_{it}, similar to the condition discussed in the linear case. Note that the expression $L_{y;I|a,X} L_{y|a,X}^{-1}$ in (50) looks like and IV regression using i_{it} as the proxy measure with error of z_{it} and a_{it+1} as the instrument for the proxy measure once we control for X_{it}.

Step 2: Productivity Process Given that our production function is Cobb-Douglas with Hicks neutral productivity, the net income process (after netting out the firm production function from the endogenous inputs) in (33) is linear and additive in the two unobserved components. The linearity and the stochastic assumptions on z_{it} and ε_{it} allow us to frame our model into the Nonlinear Markov model studied in Arellano et al. [2017]. Hence the identification of the productivity process follows the same arguments in Appendix A.1 and the supplemental material S.4 in Arellano et al. [2017].

Proposition 2. ***Identification of the Productivity Process.*** *In a Cobb-Douglas production function with Markovian Hicks neutral productivity as in equations (14)-(15), if assumption (1) and condition (1)(v) hold and β_l and β_k are previously identified, then the joint distribution of $(\varepsilon_{i2,\cdots}, \varepsilon_{iT-1})$ and the joint distribution of $(z_{i2,\cdots}, z_{iT-1})$ are identified from i.i.d observations of $(\tilde{y}_{i1,\cdots}, \tilde{y}_{iT})$ where $\tilde{y}_{it}$ is the net-income process for $T \geq 3$. With $T \geq 4$ the Markov probability $f_{z_t|z_{t-1}}(z_{it} \mid z_{it-1}))$ and $\phi = E[z_{it} \mid z_{it-1}]$ are identified.*

To provide some intuition on how identification works in our production function model we follow the discussion in Arellano [2014] and Arellano et al. [2017] and applied to our firm net income process. Following Arellano [2014] and Arellano et al. [2017], we first discuss the non-parametric identification of the distribution of ε_{it} for all t. Then using the linear structure of equation (33), by deconvolution, we can identify the distribution of z_{it}.

Given assumption 1 (i) and (ii) we can write the following nonlinear IV equation:

$$\tilde{y}_{it} = \psi(\tilde{y}_{it-1}) + \zeta_{it} \tag{51}$$

$$\tilde{y}_{it-1}\tilde{y}_{it} = \phi(\tilde{y}_{it-1}) + \upsilon_{it} \tag{52}$$

where $E[\zeta_{it} \mid \tilde{y}_{it-2}] = 0$ and $E[\upsilon_{it} \mid \tilde{y}_{it-2}] = 0$, and $\psi(.)$ and $\phi(.)$ are the solutions of an IV regression where $\tilde{y}_{it-2}$ is the instrument of $\tilde{y}_{it-1}$ in (51) and (52): $E[\tilde{y}_{it} - \psi(\tilde{y}_{it-1}) \mid \tilde{y}_{it-2}] = 0$ and $E[\tilde{y}_{it-1}\tilde{y}_{it} - \phi(\tilde{y}_{it-1}) \mid \tilde{y}_{it-2}] = 0$. The solutions $\psi(.)$ and $\phi(.)$ exist and are unique if both the conditional distributions of $\tilde{y}_{it} \mid \tilde{y}_{it-1}$ and $\tilde{y}_{it-1} \mid \tilde{y}_{it}$ are complete. This is a nonlinear relevance assumption that is ensured by the markovian condition of z_{it}. The distribution of $\tilde{y}_{it} \mid \tilde{y}_{it-1}$ is complete if $E[\phi(\tilde{y}_{it}) \mid \tilde{y}_{it-1}] = 0$ implies that $\phi(\tilde{y}_{it}) = 0$ for all ϕ in some space of functions (Newey and Powell 2003).

Identification of $\psi(.)$ and $\phi(.)$ relies on the autocorrelation structure in the data $(\tilde{y}_{i1,\cdots}, \tilde{y}_{iT})$. Note that both $\psi(.)$ and $\phi(.)$ are data objects that can be estimated with data on $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$.

Given assumption 1 (parts (i) and (ii)), $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$ are independent given z_{it-1}. Provided that the conditional distribution of z_{it-1} given $\tilde{y}_{it-2}$ is complete we have:

$$E(\tilde{y}_{it} \mid z_{it-1}) = E(\psi(\tilde{y}_{it-1}) \mid z_{it-1}), \tag{53}$$

$$z_{it-1}E(\tilde{y}_{it} \mid z_{it-1}) = E(\phi(\tilde{y}_{it-1}) \mid z_{it-1}). \tag{54}$$

Equation (53) uses the condition that $E(\psi(\tilde{y}_{it-1}) \mid z_{it-1}, \tilde{y}_{it-2}) = E(\psi(\tilde{y}_{it-1}) \mid z_{it-1})$ and $E(\tilde{y}_{it} \mid z_{it-1}, \tilde{y}_{it-2}) = E(\tilde{y}_{it} \mid z_{it-1})$, while equation (54) uses also the condition that $E(\varepsilon_{it-1} \mid z_{it-1}) = 0$ and $E(\phi(\tilde{y}_{it-1}) \mid z_{it-1}, \tilde{y}_{it-2}) = E(\phi(\tilde{y}_{it-1}) \mid z_{it-1})$.

Since $\psi(.)$ and $\phi(.)$ are identified from (51) and (52) and data on $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$, we can use equation (53) and (54) to identify the distribution of ε_{it-1} for a fixed value of z:

$$E_{\varepsilon_{it-1}}\left[z\psi\left(z+\varepsilon_{it-1}\right)\right] = E_{\varepsilon_{it-1}}\left[\phi\left(z+\varepsilon_{it-1}\right)\right] \tag{55}$$

By deconvolution we can recover the density of ε_{it-1} from (55). Using the same argument we can recover the density of ε_{it} using $\{\tilde{y}_{it-1}, \tilde{y}_{it}, \tilde{y}_{it+1}\}$, for all $t = \{2, \ldots T-1\}$. By the separability of $\tilde{y}_{it} = z_{it} + \varepsilon_{it}$, once we identify the distribution of $(\varepsilon_{i2,\cdots,}\varepsilon_{iT-1})$, we can identify the distribution of $(z_{i2,\cdots,}z_{iT-1})$ given the observed data on $(\tilde{y}_{i2,\cdots,}\tilde{y}_{iT-1})$, assuming that the characteristic functions of ε_{it-1} do not vanish on the real line. Note that we need a panel with $T \geq 4$ for identifying the Markovian process of productivity . With $T \geq 4$ we can identify the join distribution of (z_{i2}, z_{i3}) which in turn identify the conditional distribution of z_{i3} given z_{i2}. If we assume that the productivity process is stationary we have identified the conditional distribution of z_{it} given z_{it-1} for all t.

Step 3: Policy Functions Once we have identified β_k, β_l and $f(z_1 \mid \tilde{y})$ we can identify $f(a_1, k_1 \mid z_1)$ and $f(a_{t+1} \mid z_t, a_t, k_t)$ and $f(k_{t+1} \mid z_t, a_t, k_t)$ for all $t > 1$ in a sequential way starting with period 1 in a similar way as in Arellano et al. [2017].

Proposition 3. ***Identification of the Policy Functions***. *In a Cobb-Douglas production function with Markovian Hicks neutral productivity as in equations (14)-(15), if assumption (1), (2) and condition (1)(v) hold and $f(z_1 \mid \tilde{y})$ is previously identified, then $f(a_1, k_1 \mid z_1)$, $f(a_{t+1} \mid z_t, a_t, k_t)$ and $f(k_{t+1} \mid z_t, a_t, k_t)$ are identified for all $t > 1$.*

Period 1

$$f(a_1, k_1 \mid \tilde{y}) = \int f(a_1, k_1 \mid z_1, \tilde{y})\, f(z_1 \mid \tilde{y})\, dz_1, \tag{56}$$

by *assumption 1*, $f(a_1, k_1 \mid z_1, \tilde{y}) = f(a_1, k_1 \mid z_1)$ equation (56) can be expressed as:

$$f(a_1, k_1 \mid \tilde{y}) = \int f(a_1, k_1 \mid z_1)\, f(z_1 \mid \tilde{y})\, dz_1. \tag{57}$$

Equation (57) can be rewritten as the following moment restriction:

$$f(a_1, k_1 \mid \tilde{y}) = E\left[f(a_1, k_1 \mid z_1) \mid \tilde{y}_i = \tilde{y}\right] \tag{58}$$

where the expectation is taken with respect to the density of z_{i1} given $\tilde{y}_i$ and for a fixed values of a_1 and k_1. Provided that the distribution of $(z_{i1} \mid \tilde{y}_i)$, which is identified from the production function structure is complete, the unknown density $f(a_1 \mid z_1)$ is identified from (58). The density $f(a_1, k_1, z_1 \mid \tilde{y}) = f(a_1, k_1 \mid z_1)\, f(z_1 \mid \tilde{y})$ is also identified.

Using Bayesian rule, we can identify the following density:

$$f(z_1 \mid a_1, k_1, \tilde{y}) = \frac{f(a_1, k_1, z_1 \mid \tilde{y})}{f(a_1, k_1 \mid \tilde{y})}$$

Period 2 Like the analysis in period 1, we can use *assumption 1* to express $f(a_2 \mid a_1, k_1, \tilde{y})$ as:

$$f(a_2 \mid a_1, k_1, \tilde{y}) = \int f(a_2 \mid z_1, a_1, k_1) f(z_1 \mid a_1, k_1, \tilde{y}) \, dz_1 \tag{59}$$

where $f(a_2 \mid a_1, k_1, \tilde{y}) = f(a_2 \mid z_1, a_1, k_1)$. Equation (59) can be rewritten in terms of the following moment restriction:

$$f(a_2 \mid a_1, k_1, \tilde{y}) = E\left[f(a_2 \mid z_1, a_1, k_1) \mid a_{i1} = a_1, k_{i1} = k_1, y_i = y\right] \tag{60}$$

Equation (60) provides identification for $f(a_2 \mid z_1, a_1, k_1)$ as long as $f(z_1 \mid a_1, k_1, \tilde{y})$, which is identified in period 1, is complete in $\tilde{y}_i$). Note that $f(a_2, z_1 \mid a_1, k_1, \tilde{y})$ is also identified. Similarly, $f(k_2 \mid z_1, a_1, k_1)$ (and consequently $f(k_2, z_1 \mid a_1, k_1, \tilde{y})$) is identified from

$$f(k_2 \mid a_1, k_1, \tilde{y}) = E\left[f(k_2 \mid z_1, a_1, k_1) \mid a_{i1} = a_1, k_{i1} = k_1, y_i = y\right] \tag{61}$$

Given *assumption* 1 $f(a_2, k_2 \mid z_1, a_1, k_1) = f(k_2 \mid z_1, a_1, k_1) f(a_2 \mid z_1, a_1, k_1)$. Using Bayesian rule and *assumption* 1 we recover $f(z_1 \mid a_2, k_2, a_1, k_1)$ from:

$$f(a_2, k_2 \mid z_1, a_1, k_1) = \frac{f(z_1 \mid a_2, k_2, a_1, k_1) f(a_2, k_2 \mid a_1, k_1)}{f(z_1 \mid a_1, k_1)}$$

Given that $f(z_1 \mid a_2, k_2, a_1, k_1)$ is identified from above, $f(z_2 \mid z_1)$ is identified from the net-income process, and given *assumption* 1 we can identified: $f(z_2, z_1 \mid a_2, k_2, a_1, k_1) = f(z_1 \mid a_2, k_2, a_1, k_1) f(z_2 \mid z_1)$, which in turns allow us to identified $f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y})$ using Bayesian rule and given *assumption* 1:

$$f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y}) = \int \frac{f(\tilde{y} \mid z_2, z_1) f(z_2, z_1 \mid a_2, k_2, a_1, k_1)}{f(\tilde{y} \mid a_2, k_2, a_1, k_1)} dz_1$$

Period 3 Using *assumption* 1 and *assumption* 2 we have:

$$f(a_3 \mid a_2, k_2, a_1, k_1, \tilde{y}) = \int f(a_3 \mid z_2, a_2, k_2) f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y}) \, dz_1 \tag{62}$$

Provided that $f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y})$, which is identified from above, is complete in $(a_{i1}, k_{i1}, \tilde{y})$, $f(a_3 \mid z_2, a_2, k_2)$ is identified from 62. Analogously, $f(k_3 \mid z_2, a_2, k_2)$ is identified from:

$$f(k_3 \mid a_2, k_2, a_1, k_1, \tilde{y}) = \int f(k_3 \mid z_2, a_2, k_2) f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y}) \, dz_1$$

Given assumption *assumption* 2 $f(a_{t+1} \mid z_t, a_t, k_t)$ and $f(k_{t+1} \mid z_t, a_t, k_t)$ are identified provided that for all $t > 1$, the distribution of $\left(z_{it} \mid a_i^t, k_i^t, \tilde{y}_i\right)$ is complete in $\left(a_i^{t-1}, k_i^{t-1}, \tilde{y}_i\right)$.

Appendix A.4

Augmented OP In a model where the investment equation is a deterministic function of the state variables of the model (z_{it},k_{it},a_{it}), it is possible to identify and estimate the model with financial frictions using the proxy variable approach by a simple modification of the moment conditions used by OP to control for collateral constraints. Consider the following investment policy function without shocks:

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

First Stage: Under the assumption that the function h_t in monotonic in z_{it}, it is possible to invert the investment policy function to recover z_{it} as a function of observable variables:

$$z_{it} = \pi_t\left(i_{it}, k_{it}, a_{it}\right) \tag{63}$$

where $\pi_t = h_t^{-1}$. Then, we can replace z_{it} into the production function:

$$y_{it} = \beta_l l_{it} + \phi\left(i_{it}, k_{it}, a_{it}\right) + \varepsilon_{it}, \tag{64}$$

where $\phi\left(i_{it}, k_{it}, a_{it}\right) = \beta_k k_{it} + \pi_t\left(i_{it}, k_{it}, a_{it}\right)$. Using *assumption 1*, we can define the following moment condition from (64):

$$E\left(\varepsilon_{it} \mid l_{it}, k_{it}, a_{it}, i_{it}\right) = 0, \tag{65}$$

The moment condition in (65) allows us to identify β_l and the function $\phi_t\left(i_{it}, k_{it}, a_{it}\right)$, but not β_k and $\pi_t\left(i_{it}, k_{it}, a_{it}\right)$ separately. For instance, if we use a polynomial to approximate $\pi\left(i_{it}, k_{it}, a_{it}\right)$ as in OP, an OLS estimation of (64) delivers a consistent estimator of β_l and $\phi_t\left(i_{it}, k_{it}, a_{it}\right)$.

Second Stage: Combining equation (63) with the markovian model of the productivity process $z_{it} = \varphi\left(z_{it-1}\right) + \eta_{it}$:

$$\phi_t\left(i_{it}, k_{it}, a_{it}\right) = \beta_k k_{it} + \varphi\left(\phi_t\left(i_{it-1}, k_{it-1}, a_{it-1}\right) - \beta_k k_{it-1}\right) + \eta_{it} + \varepsilon_{it}, \tag{66}$$

using *assumption 1*, we can define the following moment condition from equation (66):

$$E\left(\eta_{it} + \varepsilon_{it} \mid k_{it}, k_{it-1}, a_{it-1}, i_{it-1}\right) = 0 \tag{67}$$

The moment condition in (67) allows us to identify β_k. An OLS regression of (66) is unfeasible since $\phi_t\left(i_{it}, k_{it}, a_{it}\right)$ is an unobserved variable. However, if we replace $\phi_t\left(i_{it}, k_{it}, a_{it}\right)$ by its OLS estimate from the first stage $\hat{\phi}_t\left(i_{it}, k_{it}, a_{it}\right)$, an OLS regression of (66) delivers a consistent estimate of β_k. We refer to this estimator that augments the OP estimator to control for the stock of wealth when constructing the proxy variable as *OPA-Inv.*

Wealth accumulation policy rule as the proxy variable Note that in the absence of shocks in the wealth accumulation policy rule we can also invert (17) and use the wealth accumulation as the proxy variable:

$$z_{it} = \pi_t\left(a_{it+1}, k_{it}, a_{it}\right), \tag{68}$$

where $\pi_t = g_t^{-1}$. Then, we can follow the two-stage procedure describe above but using a_{it+1} instead of i_{it} in the first stage (equations (64) and (65)) and using a_{it} instead of i_{it-1} in the second stage (equations (66) and (67)). This approach is novel, since we are the first paper to use the self-financing channel as the proxy variable for the production function estimation. We refer to this novel estimator that use the wealth accumulation policy function to construct the proxy variable as *Proxy-Wealth*. Since z_{it} is perfectly recover, estimation of the productivity process and the policy functions are straightforward.

Appendix A.5

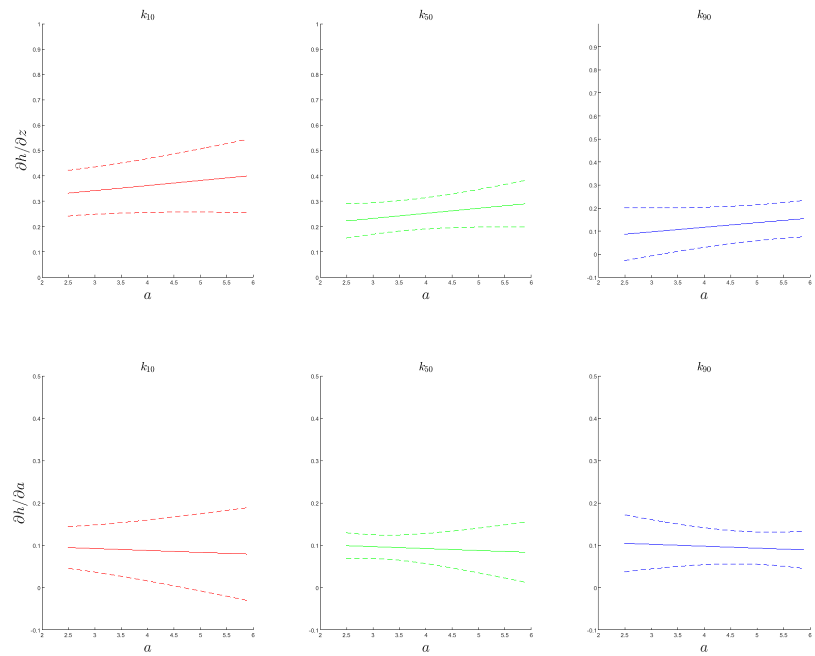

FIGURE 7: Confidence Intervals: Marginal effect of productivity and wealth on investment

Notes: The top panel exhibits the estimated derivative effect of productivity in the investment policy function and its 95% confidence intervals using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital (10th, 50th and 90th percentile of the capital distribution). The bottom panel of the figure exhibits the estimated derivative effect of the stock of wealth (previous wealth) in the investment policy function and its 95% confidence intervals using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital (10th, 50th and 90th percentile of the capital distribution).

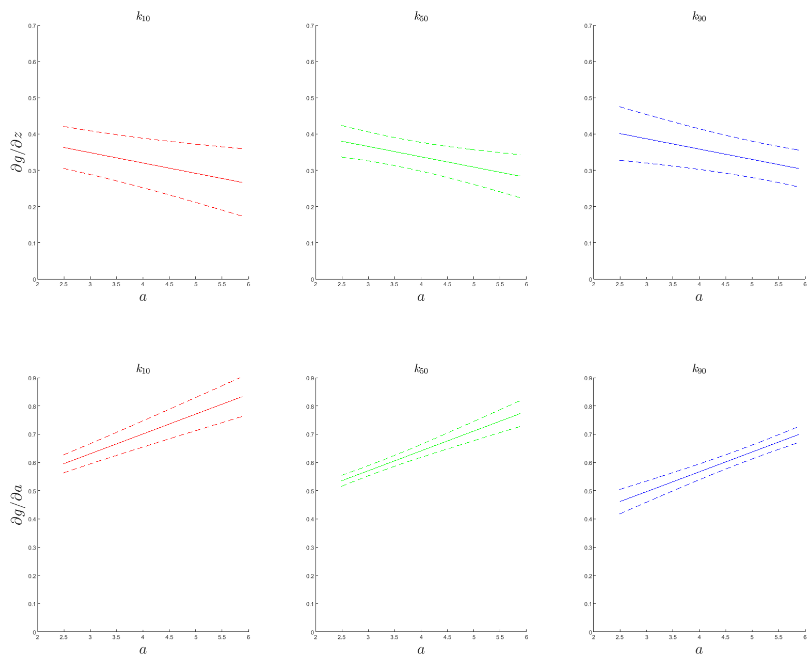

Figure 8: Confidence Intervals: Marginal effect of productivity and wealth on wealth accumulation

Notes: The top panel exhibits the estimated derivative effect of productivity in the wealth policy function and its 95% confidence intervals using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital (10th, 50th and 90th percentile of the capital distribution). The bottom panel of the figure exhibits the estimated derivative effect of the stock of wealth (previous wealth) in the wealth policy function and its 95% confidence intervals using the proxy-IV approach. The figure displays how the effect changes along different values of the stock of wealth and is evaluated at three different level of stock of capital (10th, 50th and 90th percentile of the capital distribution).

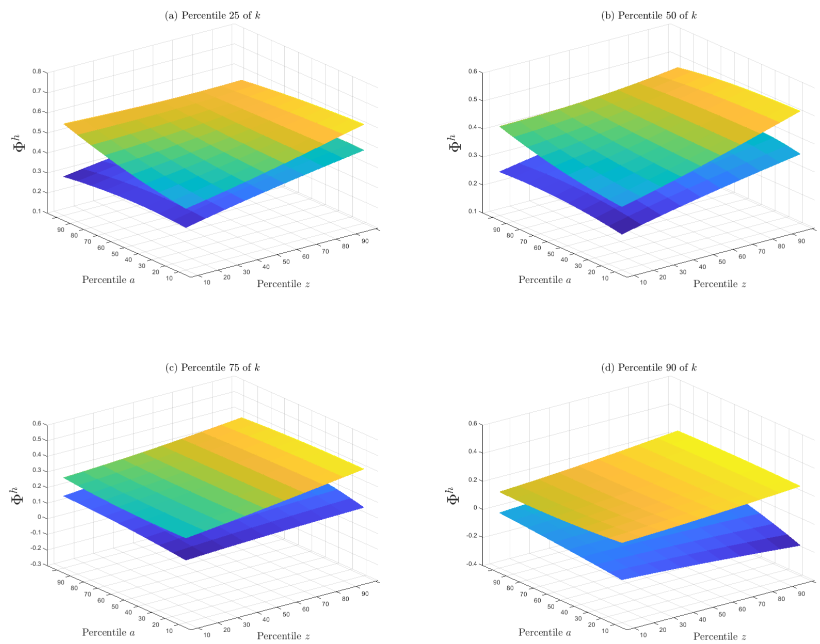

Figure 9: Confidence Intervals: Marginal effect of productivity on investment

Notes: The figure exhibits the 95% confidence intervals of the estimated derivative effect of productivity in the investment policy function using the SEM method.

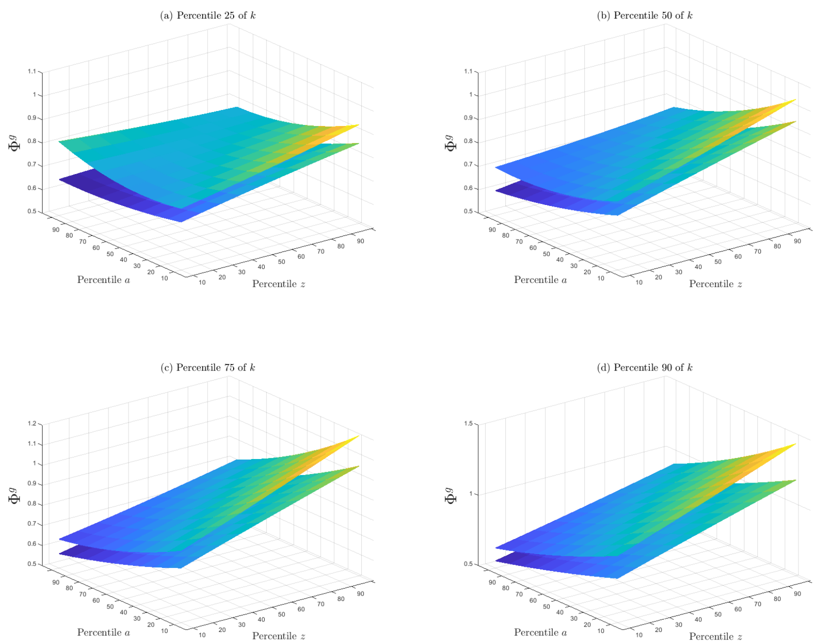

FIGURE 10: Confidence Intervals: Marginal effect of productivity on wealth accumulation

Notes: The figure exhibits the 95%confidence intervals of the estimated derivative effect of productivity in the wealth accumulation policy function using the SEM method.