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

Alvaro Aguirre
Central Bank of Chile
Santiago
Chile

Matias Tapia
Central Bank of Chile
Santiago
Chile

Lucciano Villacorta
Central Bank of Chile
Santiago
Chile

First draft: November 2020
This version: March 2022

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 the presence of financial frictions, whereas standard methods used to estimate productivity dynamics are biased. We uncover a distribution of investment and wealth accumulation propensities in response to productivity shocks. These propensities are heterogeneous in the stock of wealth and productivity level: (i) investment propensities are larger for high-productivity firms and high-wealth firms, and (ii) wealth accumulation propensities are larger for high-productivity firms with low levels of wealth. We provide evidence of collateral-based and earning-based constraints. Our estimates support the existence of self-financing, but its impact is limited.

JEL classification: C33, E23, O11, L0

Keywords: Investment propensity, wealth dynamics, self-financing, financial frictions, production function, productivity process, proxy variable approach; panel data, nonlinear.

*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 the Conference on econometric methods and empirical analysis of microdata in honor of Manuel Arellano, STEG Workshop on Firms, Frictions and Spillovers, and Industrial Policy 2022, the BSE Summer Forum Workshop on Financial Shocks, Channels, and Macro Outcomes 2022, SED 2022, 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, CEMFI, Central Bank of Chile, and U Chile for their comments, and Diego Huerta and Cristian Valencia for excellent research assistance.

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 to invest according to their productivity levels. This can lead to an inefficient allocation of factors with important aggregate implications such as low aggregate investment and 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 [e.g. Moll, 2014]. This endogenous response, known as the self-financing channel, has the potential to mitigate the aggregate adverse effects of financial frictions. For example, quantitative exercises as in Banerjee and Moll [2010] and Midrigan and Xu [2014] suggest that self-financing is strong enough to rapidly undo the impact of financial frictions on misallocation. Importantly, both financial frictions and the self-financing channel are reflected in the propensities of firms' investment and wealth accumulation to productivity shocks and how these propensities depend on the amount of available collateral. However, a precise empirical assessment of these objects using micro data is currently absent from the literature.

In this paper, we characterize and estimate the firms' relevant policy functions for investment and wealth accumulation decisions under financial fictions. We use these functions to empirically document the response of investment and wealth accumulation to productivity shocks at the micro level and study how these propensities vary along the wealth distribution. These objects provide direct empirical evidence of the presence of financial frictions and the intensity of the self-financing channel at the firm level, and can help discipline quantitative models of firm dynamics with financial frictions.

We implement our methodology using data on manufacturing firms obtained from the census of administrative records of formal firms in Chile from 2006 to 2016. Besides including data on inputs and output at the firm level, our database provides balance sheet information for both public and private firms, which allows us to have a measure of firms' wealth. Our estimates support the existence of self-financing, but suggest that its impact is limited.

Our empirical analysis is guided by the economic models that study the self-financing channel [e.g. Buera and Shin, 2011, Moll, 2014, Midrigan and Xu, 2014]. However, an important aspect of our tractable econometric framework is that it uncovers the firms' productivity process and its effects on firms' investment and wealth accumulation decisions without relying on a structural estimation. In contrast to fully-specified structural approaches, which require the specification of particular functional forms for preferences, financial frictions and productivity, we adopt a non-parametric approach where we recover productivity from the firm production function and estimate nonlinear firm's policy rules that are compatible with a large class of heterogeneous-agent models with financial frictions.[1]

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 (firm wealth), enabling us to reveal

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

the complete distribution of micro-level investments and wealth accumulation propensities in response to productivity shocks. Consequently, our modeling approach can provide a rich picture of the joint relationship between investment decisions, wealth dynamics and productivity shocks drawn directly from the data. In that respect, our empirical framework shares the spirit of the empirical household's consumption-income framework [e.g. Blundell et al., 2008, 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 here we apply it in the context of firms' decisions. A crucial econometric difference of our framework lies in the estimation of the firm 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 firm productivity shocks and their effect on investment and savings, we need to estimate the production function parameters, addressing the fact that the regressors are endogenous and correlated with unobserved productivity.

A first step of our analysis builds on previous literature that estimates production functions and productivity at the firm level. This literature relies on a proxy variable approach to recover productivity using the firm's input decisions [see Ackerberg et al., 2015, for a review]. For instance, Olley and Pakes [1996] recover productivity by inverting an investment demand function, whereas Levinsohn and Petrin [2003] invert the firm demand function for intermediate inputs. We extend this estimation method to allow for financial frictions, as we show that otherwise, proxy methods deliver biased estimates of the production function and the productivity process. Intuitively, financial frictions generate differences in input demands for equally productive firms that the proxy variable method misinterprets as differences in productivity. Additionally, the proxy method is not well-suited to identify and estimate flexible empirical policy functions, as it does not allow for unobservables besides productivity in the policy functions. This is an empirically restrictive assumption since it rules out the possibility of idiosyncratic shocks and measurement error.

Our empirical framework consists of a firm production function, a non-linear investment policy rule, and a non-linear wealth accumulation policy rule. These three equations depend on the latent firm-level productivity process which follows a flexible non-linear Markovian process. In our baseline specification, each policy rule also depends on the stock of wealth, the stock of capital and an unobserved idiosyncratic shock. The inclusion of the stock of wealth as an additional state variable in the investment policy plays an important role to control for financial frictions, in accordance with the insights of theoretical models in which firm wealth determines the degree of the financial constraint. A novelty of our framework is to explicitly model the wealth accumulation policy function and its relationship with productivity. This function is our main focus, as it plays a fundamental role in understanding the scope and implications of the self-financing channel. Allowing for nonlinearities in the empirical model is crucial, since according to the theoretical models the responses of investment and wealth dynamics to productivity shocks vary across firms as a function of their wealth and productivity. In that sense, our non-linear framework uncovers new empirical results that provide evidence on financial fictions and self-financing, complementing the insights and lessons of structural models.

For example, we use the estimated investment policy function 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 and productivity. For example, for all capital levels, the investment propensity to productivity shocks 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. This is consistent with the notion that firms with higher wealth face softer financial constraints, allowing them to adjust capital more rapidly in response to productivity changes. Also, and consistent with Moll [2014] or Midrigan and Xu [2014], we find that the response of investment to changes in wealth is positive but decreasing in the level of wealth, which we also interpret as evidence of collateral constraints. All else equal, firms with low levels of wealth are more constrained and the marginal value of increasing their collateral is larger, as it allows them to increase investment more significantly.

Moreover, in line with earning-based constraints models [as in Lian and Ma, 2020, Drechsel, 2019], the investment propensity in response to productivity shocks is also heterogeneous in the firm productivity level, with a larger propensity for more productive firms. Interestingly, how the propensity change along the wealth distribution also depends on the level of productivity. For instance, for low productivity-firms with low-wealth the investment propensity is low and very sensitive to changes in wealth, suggesting that collateral constraints might play a key role in their investment decisions. By contrast, for very productive firms, the investment propensity is high even for firms with low levels of wealth, as they can probably also rely on earnings. However, even for these productive firms the investment propensity is increasing in wealth, and therefore, they also have strong incentives to save and increase their wealth.

The estimated wealth accumulation policy function 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 wealth and productivity. For highly productive firms at the bottom of the wealth distribution, the elasticity of productivity on wealth accumulation is close to 1. Thus, for very productive, 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 (as in Moll [2014]): 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.

In order to asses the strength of self-financing, we follow Banerjee and Moll [2010], and use our estimated empirical model to compute the speed of convergence of the marginal product of capital (MPK) between two firms that have the same productivity but start with different levels of wealth. On the one hand, results are consistent with the existence of self-financing in the data, as convergence in MPK between firms does occur. On the other hand, the impact of self-financing appears to be limited, as convergence is very slow. 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 our estimates imply that their MPKs only converge after more than 40 years. Still, half of the initial gap in their MPKs disappears after ten years. 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, Midrigan and Xu, 2014].

An overview of our methodology Identification and estimation of our nonlinear model cannot be handled within the standard proxy variable framework, since our nonlinear policy rules are more flexible and include unobservable shocks in addition to the latent productivity process. Combining the insights of the self-financing channel with 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], we propose a sequential identification scheme to nonparametrically identify the production function, the productivity process and the policy functions. From an instrumental variable perspective, both policy functions 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 accumulation policy provides an external instrument for investment (the proxy variable with noise) in the production function regression. Intuitively, due to self-financing, a positive co-movement between investment and wealth accumulation decisions is related to changes in productivity that can be used to identify the production function. Once the production function parameters are identified, the productivity process is identified from the dynamic dependence structure of the firm net income process and the policy rules are identified using non-parametric instrumental variables arguments given the exclusion restrictions provided by our dynamic model.

Regarding estimation, we show that for parsimonious, yet flexible, versions of the policy functions, an instrumental variable estimation strategy within the proxy variable framework delivers consistent estimates of the model, following the arguments in the identification strategy. For more general policies, 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.

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, 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 in a similar fashion to Catherine et al. [2018]. Catherine et al. [2018] uses reduced-form evidence on the sensitivity of investment to collateral as the target to match in a structural estimation of a model with collateral constraints and adjustment costs. Our results provides a set of new empirical moments to match, as the complete distribution of investment and wealth accumulation sensitivities.

Our framework also uncovers new empirical results on the estimates of the firm production function and productivity process as we find significant differences once we control for financial frictions. A consistent estimation of the production function and the productivity process is important as these objects are crucial inputs in structural macro models that quantitatively study the self-financing channel. The estimated capital parameter in the production function increases from 0.35 when using the proxy variable approach to 0.43 when we consider financial frictions in the estimation. By contrast, the estimated labor parameter decreases from 0.67 to 0.44 when controlling for financial frictions. The differences in the estimates of factor elasticities have relevant aggregate implications, as they translate into significant differences in the measure of returns to scale. We also show that not controlling for financial frictions significantly underestimates both the persistence and the dispersion of productivity. The estimated persistence increase in 50% when we account for financial frictions, whereas the estimated standard deviation more than doubles. This results is consistent with the theoretical implications of financial frictions: more productive firms, which all else equal are expected to be more financially constrained according to the canonical model, show larger investment gaps with respect to their optimal levels. This leads the proxy variable estimator to underestimate these firms' productivity relative to unproductive firms and to shrink the estimated productivity distribution. Underestimating these parameters could have aggregate implications as they affect the conclusions of quantitative exercises. In quantitative models a lower persistence reduce the strength of self-financing in mitigating the negative effects of financial frictions, whereas a lower dispersion implies less productivity fluctuation over time reducing the effects of financial frictions in distorting allocations [see Moll, 2014].

Related literature Our paper makes contributions to differents streams of literature. Our paper is motivated by 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 self-financing [e.g. Banerjee and Moll, 2010, 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. Our main contribution is to empirically estimate the wealth accumulation and investment decisions of firms, which in these papers are an endogenous outcome of structural models calibrated with microdata 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 estimates of key elasticities and, unlike these papers, we exploit microeconomic data not only on real variables, but also on financial variables. Ours is the first paper to provide empirical evidence of the self-financing channel studied in this literature.

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 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.[2] In our framework, the investment and wealth 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, and that wealth accumulation decisions are affected by the firm's productivity process.[3]

Finally, this paper 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]. Our paper differs from these papers in several aspects. First, our paper is the first paper to study theoretically and empirically the biases that appear when the proxy method is used to estimate the production function 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.[4] 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 input demand and the wealth accumulation policy rules. Finally, a key difference is the identification and estimation of the policy functions.

The rest of the paper is organized as follows. Section 2 presents a model of firm dynamics with collateral constraints in order to motivates the ingredients of the empirical model and shed light on the biases of the proxy variable approach. Section 3 introduces the empirical model and its assumptions. Section 4 establishes identification of model. Section 5 describes the estimation methods. Section 6 describes the data and presents the main empirical results. Section 7 concludes.

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

[3] 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.

[4] Shenoy [2020] studies how the proxy variable approach fails when any type of market frictions distort the firm's input choices and proposes a dynamic linear panel data approach (Arellano and Bond [1991], Blundell and Bond [1998]) to estimate the production function. Also different from these papers we do not assume a linear model for productivity which is a very restrictive assumption and at odds with the data. Instead we model productivity non linearly.

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. Second, this setup provides insights on the general form of the firm policy rules for investment and wealth accumulation.

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 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. 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}:

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

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.[5] The function $g(\cdot)$ is assumed to be convex, which given the use of linear preferences, rules out corner solutions.[6] 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.

[5] This timing assumption is relevant in Olley and Pakes [1996] 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].

[6] 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.

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.[7] Both 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 Banerjee and Moll [2010] Aguirre [2017], Brooks and Dovis [2020] [see Buera et al., 2015, for a closer examination] and in models with earning-based constraints as in Drechsel [2019], di Giovanni et al. [2022] or Lian and Ma [2020]. 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:

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

[7]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].

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. An important feature of this setup with earning-based constraints is that even for firms with low levels of wealth, investment is still increasing in productivity. We use this property in the discussion of identification of the empirical model in section 4. Equation (4) leads to the following general policy function for investment that we take to the data:

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

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 4 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 proxy variable estimators under financial frictions

An important first step before estimating the policy functions in (6) and (8) is the estimation of the firm productivity and the firm production function. The model described above provides insights into the biases that appear when estimating these objects using standard methods which do not account for financial frictions. In this subsection we provide an intuition on the nature of those biases, and provide a more detailed explanation in Appendix A.1.

We illustrate our general argument in the context of the influential paper by Olley and Pakes [1996], henceforth OP, although the same applies to Levinsohn and Petrin [2003] as long as financial frictions affect the demand of intermediate inputs as in Mendoza and Yue [2012] and Bigio and La'o [2020].

Consider the log of the value-added production function described above:

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

where ε_{it} is measurement error in value added.[8] 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}. By controlling for investment in the production function, OP can eliminate the endogeneity problem and get consistent estimates of β_l and β_k. Intuitively, the OP method implies that observed differences in investment between firms in the data are interpreted as differences in unobserved productivity between those firms.

However, 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. Therefore, under borrowing constraints, equally productive firms might have different levels of investment, capital and output. Even if the implied variation in output is driven by variations in capital due to differences in financial frictions across firms, the OP approach will assign such variation to variations in productivity since OP misinterprets differences in investment as differences in productivity. 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 relatively 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, these biases in the estimates of factor elasticities can translate into significant differences in the measure of returns to scale.

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.

2.2 The Effect of Shocks in the Policy Functions

Another important limitation of the proxy variable approach is the scalar unobservable assumption, which rules out the possibility of other shocks or measurement error in the policy functions.[9] This is

[8]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].

[9]This assumption is required to invert the policy function and express productivity only as function of observable variables and parameters (see Appendix A.1).

a very restrictive assumption, as it is very likely that our empirical policy functions will be affected by other unobservables apart from productivity shocks. For example, in the context of the model presented above, it is natural to think that shocks to collateral constraints can affect 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 (i.e $\kappa(Z_{it}, A_{it}, v_{it})$).

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 in the interim between the distribution of dividends (when equation 7 is solved) and when 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.

3 General Empirical Framework

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

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \tag{10}$$

We augment equation (1) to consider a nonlinear Markovian process for the Hicks-neutral productivity process, since an AR(1) model is too restrictive and is at odds with the literature that estimates firm level production functions

$$z_{it} = \varphi\left(z_{it-1}\right) + \eta_{it}, \tag{11}$$

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 it is standard in the literature, shocks η_{it+1} and ε_{it} are not part of the firm's information set when making decisions at t. The assumptions about the stochastic processes underlying both shocks 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}, while labor l_{it} is a flexible input. The specification of the empirical policy rules follows the stylized model discussed in the previous section:

$$i_{it} = h_t\left(z_{it}, k_{it}, a_{it}, v_{it}\right), \tag{12}$$

$$a_{it+1} = g_{t+1}\left(z_{it}, a_{it}, k_{it}, w_{it+1}\right). \tag{13}$$

where h_t and g_{t+1} are the nonlinear reduced-form policy rules of investment and wealth that can be derived in a firm-dynamics model with financial frictions as the one discussed in section 2. The v_{it} and w_{it+1} capture other unobserved factors besides z_{it} that affect the evolution of investment and wealth, like the shocks discussed in section 2.2.[10] Also, h_t and g_{t+1} are monotonic in v_{it} and w_{it+1},

[10] The inclusion of v_{it} in the investment policy function represents a departure from the scalar unobservable assumption required by the proxy variable framework. It is important to remark that in the proxy variable approach the investment function is not an object of interest by itself, and is used only as an auxiliary equation to recover the production function.

respectively. The specification in (12) nests a number of nonlinear empirical investment functions studied in the literature [e.g. Olley and Pakes, 1996, Cooper and Haltiwanger, 2006, Gala et al., 2020]. The two major innovations of our framework are (i) the inclusion of wealth a as a additional state variable in (12) to control for the existence of collateral constraints and (ii) the explicit modelling of wealth dynamics in (13) and its relationship with productivity (the self-financing channel). An additional important difference to Gala et al. [2020] and Cooper and Haltiwanger [2006] is that we explicitly include z_{it} as a state variable, whereas these papers replace z_{it} for value-added.[11]

The nonlinear functions h_t and g_{t+1} allow for heterogeneous effects of productivity shocks on investment and on wealth accumulation, depending on the amount of collateral and the productivity level of the firm. In that sense, our main objects of interest are the following average derivative effects with respect to z_{it}:

$$\Phi_{it}^{h} = \Phi^{h}\left(a_{it}, k_{it}, z_{it}\right) = E_{v_{it}}\left[\frac{\partial h_t\left(z_{it}, k_{it}, a_{it}, v_{it}\right)}{\partial z}\right] \tag{14}$$

$$\Phi_{it+1}^{g} = \Phi^{g}\left(a_{it}, k_{it}, z_{it}\right) = E_{w_{it+1}}\left[\frac{\partial g_{t+1}\left(z_{it}, k_{it}, a_{it}, w_{it+1}\right)}{\partial z}\right] \tag{15}$$

where the expectations are taken with respect to the idiosyncratic shocks in the policy functions. Therefore, Φ_{it}^{h} and Φ_{it+1}^{g} measure the average propensities of investment and wealth accumulation in response to productivity shocks. Importantly, these propensities are heterogeneous and vary with the firms' stock of wealth and productivity level. Characterizing how these propensities vary along the wealth and productivity distributions allows us to provide novel evidence on financial frictions and the self-financing channel. For example, in collateral constraint models, Φ_{it}^{h} is increasing in a_{it}, whereas in earning-based models [Drechsel, 2019] and forward looking constraint models [Buera et al., 2015] Φ_{it}^{h} is increasing in z_{it}. Moreover, in models with a self-financing channel as in Moll [2014], Φ_{it+1}^{g} is always positive and decreasing in a_{it}. Another important object of interest is the response of investment to wealth shocks:

$$\Xi_{it}^{h} = \Xi^{h}\left(a_{it}, k_{it}, z_{it}\right) = E_{v_{it}}\left[\frac{\partial h_t\left(z_{it}, k_{it}, a_{it}, v_{it}\right)}{\partial a}\right] \tag{16}$$

In models with collateral constraints, Ξ_{it}^{h} is positive but decreasing in a_{it}, as firms with low levels of wealth are more constrained, and therefore the marginal value of of additional collateral is larger.

Finally, as it is standard in both the macro literature and the production function literature, we model the labor input as a non-dynamic input, in the sense that current choices are not affected by past values:

$$l_{it} = n_t\left(z_{it}, a_{it}, k_{it}, w_{l,it}\right), \tag{17}$$

where equation (17) 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 hiring decisions,

[11] As Gala et al. [2020] argue in footnote 10, including z_{it} instead of y_{it} requires the estimation of the production function, which adds a number of econometric problems, most significantly, endogeneity. One of the contributions of our paper is to address this issue and consistently estimate the production function, and the correct investment equation as a function of unobserved productivity.

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 shock 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 description, 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, independent over time and also independent of z_{i1}, a_{i1}, k_{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 macro models as well as in the empirical literature that estimates production functions.

4 Identification

In this section, we establish identification of the nonlinear dynamic panel model presented in the previous section. Identification challenges in our model are more demanding than those of firm dynamics models studied in the proxy variable literature due to the presence of additional shocks in the policy functions. 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 build 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 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 (11), (12) and (13)

$$z_{it} = \rho_z z_{it-1} + \eta_{it}, \tag{18}$$

$$i_{it} = h_z z_{it} + h_a a_{it} + h_k k_{it} + v_{it}, \tag{19}$$

$$a_{it+1} = g_z z_{it} + g_a a_{it} + g_k k_{it} + w_{it+1}, \tag{20}$$

Using (19), z_{it} can be written as a linear separable function of i_{it}, a_{it}, k_{it} and v_{it}.

$$z_{it} = \pi_1 i_{it} + \pi_2 a_{it} + \pi_3 k_{it} + \pi_4 v_{it} \tag{21}$$

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 (21) into the production function, we get:

$$y_{it} = \beta_l l_{it} + (\beta_k + \pi_3) k_{it} + \pi_1 i_{it} + \pi_2 a_{it} + \tilde{\varepsilon}_{it} \tag{22}$$

where $\tilde{\varepsilon}_{it} = \varepsilon_{it} + \pi_4 v_{it}$.

For simplicity, consider the case where h_k=π_3=0. In the absence of investment shocks, a simple OLS regression between y_{it} on l_{it}, k_{it}, i_{it} and a_{it} identifies β_l and β_k, as in the proxy variable approach.[12] The difference with the proxy variable is that the regression presented in (22) controls for a_{it}. However, in the more general case with investment shocks (i.e $v_{it} \neq 0$), z in (21) can not be expressed only as a function of observables and parameters. The latter violates the scalar unobservable assumption required by the proxy variable approach and, therefore, the model in (22) can not be consistently estimated using OLS since $E\left(i_{it}\tilde{\varepsilon}_{it}\right) \neq 0$.[13]

IV identification To solve the endogeneity problem in the proxy variable approach, we notice that the self-financing channel provides a second noisy measure of productivity in a setup with financial

[12] When $\pi_3 \neq 0$then β_k cannot be separately identified from π_3 in the first stage.

[13] Even if the investment shock v_{it} is not directly correlated with inputs of the production function, an OLS estimation of (19) will generate a bias in the estimation of β_l and β_k through the correlation of the inputs and the latent variable z_{it}, as in the classical linear multivariate model with measurement error in one of the regressors.

frictions. Hence, a_{it+1} can be used as an instrument for investment in equation (22,) given the conditional independence assumption - wealth does not have a direct effect in the production function- and the relevance condition implied by the self-financing channel $g_z \neq 0$. Note that even after conditioning on a_{it} and k_{it} the functions h_t and g_{t+1} are correlated via z_{it}. Therefore, we can construct the following IV moment restriction from (22):

$$E\left[y_{it} \mid a_{it+1}, l_{it}, k_{it}, a_{it}\right] = \beta_l l_{it} + (\beta_k + \pi_3) k_{it} + \pi_1 E\left[i_{it} \mid a_{it+1}, k_{it}, l_{it}, a_{it}\right] + \pi_2 a_{it}. \tag{23}$$

A regression between $E\left[y_{it} \mid a_{it+1}, l_{it}, k_{it}, a_{it}\right]$, 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 (23) identifies $\{\beta_l, \pi_1, \pi_2\}$, which in turn can identify $\{h_z, h_a\}$.

The self-financing channel is key for identification. According to the model in section 2, a firm that experiences 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 where the external instrument is justified by the theoretical insights of macro models.

Once β_l and β_k are identified we can define the firm net income process:

$$y_{it} - \beta_l l_{it} - \beta_k k_{it} = \tilde{y}_{it} = z_{it} + \varepsilon_{it} \tag{24}$$

Replacing (24) in (20) and (18) :

$$a_{it+1} = g_z \tilde{y}_{it} + g_a a_{it} + g_k k_{it} + w_{it+1} - g_z \varepsilon_{it} \tag{25}$$

$$\tilde{y}_{it} = \rho_z \tilde{y}_{it-1} + \eta_{it} + \varepsilon_{it} - \rho_z \varepsilon_{it-1}, \tag{26}$$

Notice that an OLS regression of a_{it+1} on $\tilde{y}_{it}$, a_{it} and k_{it} from equation (25) does not identify the policy function, as $E\left(\tilde{y}_{it}\varepsilon_{it}\right) \neq 0$. Likewise, an OLS regression between $\tilde{y}_{it}$ and $\tilde{y}_{it-1}$ does not identify ρ_z, because $\tilde{y}_{it-1}$ is correlated with ε_{it-1}. However, $\tilde{y}_{it-1}$ and $\tilde{y}_{it-2}$ can be used as instruments for $\tilde{y}_{it}$ and $\tilde{y}_{it-1}$ in (25) and (26), respectively. The Markovian assumption for productivity provides the relevance condition, as it ensures that $E\left(\tilde{y}_{is}\tilde{y}_{is-1}\right) \neq 0$, while Assumption 1 provides exogeneity: $E\left(\tilde{y}_{is-1}\varepsilon_{is}\right) = 0$. Once we have identified ρ_z, σ_η^2 and σ_ε^2 can be identified from the following moments:

$$E\left(\tilde{y}_{it}\tilde{y}_{it-1}\right) = \rho_z E\left(\tilde{y}_{it-1}\tilde{y}_{it-1}\right) - \rho_z \sigma_\varepsilon^2 \tag{27}$$

$$E\left(\tilde{y}_{it}\tilde{y}_{it}\right) = \rho_z^2 E\left(\tilde{y}_{it-1}\tilde{y}_{it-1}\right) + \sigma_\eta^2 + \left(1 - \rho_z^2\right)\sigma_\varepsilon^2 \tag{28}$$

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 nonparametrically. 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 nonparametric model 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 (10)-(17) 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.[14] 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 $h_t = E[i_{it} \mid z_{it}, X_{it}]$ is strictly monotonic in z_{it}.[15] Accordingly, the macro model with forward-looking financial constraints sketched in section 2 implies a monotonic relationship between productivity and investment.[16] Condition 1-(iv) is a completeness condition commonly assumed in the literature on nonparametric instrumental variables [Newey and Powell, 2003].[17] 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 (13), 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.[18] 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*

[14] 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].

[15] This condition is weaker than the assumption in the proxy variable approach, where the realization of investment have to be monotonic in z_{it}. Here we require the probability of invest to be different for different values of z_{it}.

[16] 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.

[17] 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.

[18] 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.

productivity and financial frictions as in (10)-(17), 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:

$$f(y_t, i_t \mid a_{t+1}, X_t) = \int f(y_t \mid z_t, k_t, l_t)\, f(i_t \mid z_t, X_t)\, f(z_t \mid a_{t+1}, X_t)\, dz_t \tag{29}$$

We notice that equation (29) 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])[19]

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). [20]

$$\tilde{y}_{it} = z_{it} + \varepsilon_{it} \tag{30}$$

Given that z_{it} is Markovian and ε_{it} is i.i.d over time, equation (30) has a similar structure to the household income process model with non linear Markovian persistent shocks studied in Arellano

[19] 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) for a 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.

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

et al. [2017]. To identify the productivity process we rely on the fact that the net-income process in (30) 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.

$$f\left(a_1, k_1 \mid \tilde{y}^T\right) = E\left[f\left(a_1, k_1 \mid z_1\right) \mid \tilde{y}_i^T = \tilde{y}^T\right] \tag{31}$$

$$f\left(a_{t+1} \mid a^t, k^t, \tilde{y}^T\right) = E\left[f\left(a_{t+1} \mid z_t, a_t, k_t\right) \mid a_i^t = a^t, k_i^t = k^t, \tilde{y}_i^T = \tilde{y}^T\right] \tag{32}$$

$$f\left(k_{t+1} \mid a^t, k^t, \tilde{y}^T\right) = E\left[f\left(k_{t+1} \mid z_t, a_t, k_t\right) \mid a_i^t = a^t, k_i^t = k^t, \tilde{y}_i^T = \tilde{y}^T\right] \tag{33}$$

where the expectation in (31) 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 (32) and (33) 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 (31) 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 (31) 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 (31). Similarly, equations (32) and (33) can be 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 $f\left(a_{t+1} \mid z_t, a_t, k_t\right)$ and $f\left(k_{t+1} \mid z_t, a_t, k_t\right)$ for $t > 2$ are identified recursively from equations (32) and (33). The identification of $f\left(a_{t+1} \mid z_t, a_t, k_t\right)$ and $f\left(k_{t+1} \mid z_t, a_t, k_t\right)$ 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 two approaches to estimate different versions of the empirical model presented in section 3 and discussed in section 4. First, we consider a more parsimonious model 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, following the identification strategy presented in section 4.1. Second, we consider a more flexible model that allows for unrestricted 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 Parsimonious policy functions

Proxy-IV 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. The other policy function as well as the distribution of the shocks are left unrestricted. For example, consider the following wealth accumulation policy function:

$$a_{it+1} = g\left(z_{it}, k_{it}, a_{it}, w_{it}\right) = g_1\left(k_{it}, a_{it}\right) + g_2\left(k_{it}, a_{it}\right) z_{it} + w_{it+1}, \tag{34}$$

It is important to remark that model (34) 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 (34):

$$z_{it} = \pi_1\left(k_{it}, a_{it}\right) + \pi_2\left(k_{it}, a_{it}\right) a_{it+1} + \omega_{it+1} \tag{35}$$

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 (35) in the production function:

$$y_{it} = \beta_l l_{it} + \phi\left(k_{it}, a_{it}\right) + \pi_2\left(k_{it}, a_{it}\right) a_{it+1} + \omega_{it+1} + \varepsilon_{it}, \tag{36}$$

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 (36) 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 (36). Therefore, we propose the following two-stage procedure:

First Stage: Estimate (36) 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 (35) with the markovian model of the productivity process $z_{it} = \rho_z z_{it-1} + \eta_{it}$:

$$z_{it} = \rho_z \pi_1\left(k_{it-1}, a_{it-1}\right) + \rho_z \pi_2\left(k_{it-1}, a_{it-1}\right) a_{it} + \rho_z \omega_{it} + \eta_{it}, \tag{37}$$

Replacing equation (37) into the production function:

$$y_{it} - \beta_l l_{it} = \beta_k k_{it} + \rho_z \pi_1\left(k_{it-1}, a_{it-1}\right) + \rho_z \pi_2\left(k_{it-1}, a_{it-1}\right) a_{it} + \rho_z \omega_{it} + \eta_{it} + \varepsilon_{it}, \tag{38}$$

using *assumption 1* we can define the following moment condition from equation (38)

$$E\left(\omega_{it}+\eta_{it}+\varepsilon_{it} \mid k_{it}, k_{it-1}, a_{it-1}, i_{t-1}\right)=0, \tag{39}$$

The moment condition in (39) allows us to identify β_k. If we replace β_l, $\pi_1\left(k_{it-1}, a_{it-1}\right)$ and $\pi_2\left(k_{it-1}, a_{it-1}\right)$ by their IV estimates from the first stage, an OLS regression of (38) 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 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 $t=1,\ldots,T$

$$\begin{cases} y_{it}=\beta_l l_{it}+\beta_k k_{it}+z_{it}+\varepsilon_{it} \\ z_{it}=\sum_{r=1}^{R} \alpha_r^{\varphi} \phi_r^{\varphi}\left(z_{it-1}\right)+\eta_{it} \\ i_{it}=\sum_{r=1}^{R} \alpha_r^{h} \phi_r^{h}\left(z_{it}, k_{it}, a_{it}, \delta_t^{h}\right)+v_{it} \\ a_{it+1}=\sum_{r=1}^{R} \alpha_r^{g} \phi_r^{g}\left(z_{it}, k_{it}, a_{it}, \delta_t^{g}\right)+w_{it+1} \\ a_{i1}=\sum_{r=1}^{R} \alpha_r^{g1} \phi_r^{g}\left(z_{i1}, \delta_1^{g1}\right)+w_{i1} \\ l_{it}=\sum_{r=1}^{R} \alpha_r^{n} \phi_r^{n}\left(z_{it}, k_{it}, a_{it}\right)+w_{l,it+1} \end{cases} \tag{40}$$

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 model like the investment and wealth accumulation propensities discussed in section 3. 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. In our baseline specification of the nonlinear model, we specify stationary policy functions and a stationary productivity process with additive errors that are normally distributed to have a more parsimonious model to take to the data but, as we shown in section 4, the model is non-parametrically identified with time-varying functions, non-additive errors and without parametric distributions. The nonlinear mean model in (40) is a nonparametric model for the conditional mean of productivity and the policies and has the flexibility to allow for rich interaction between state variables and to uncover heterogeneous propensities. The normality assumption on the shocks provides a computationally tractable way to estimate the model combining simulation methods (Markov Chain Monte Carlo) and simple OLS regressions. As a robustness we also consider the quantile framework introduced in Arellano et al. [2017] to model productivity and the policies. This is a more general model as it does not assume normality for the shocks in (40) and also does not impose separability in the productivity process leaving the dependence structure of z_{it} unrestricted beyond the Markovian assumption.

Stochastic EM Estimation Algorithm (SEM) To estimate the nonlinear model with latent variables in 40, 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 *assumption 1*, the full model in (40) imply the following integrated moment restrictions:

$$E\left(\int \begin{bmatrix} \sum_{t=2}^{T}\left(a_{it+1}-\sum_{k=1}^{K}\alpha_k^g\phi_k^g\left(z_{it},k_{it},a_{it},\delta_t^g\right)\right)^2 \\ \sum_{t=1}^{T}\left(i_{it}-\sum_{k=1}^{K}\alpha_k^h\phi_k^h\left(z_{it},k_{it},a_{it},\delta_t^h\right)\right)^2 \\ \sum_{t=1}^{T}\left(l_{it}-\sum_{k=1}^{K}\alpha_k^n\phi_k^n\left(z_{it},k_{it},a_{it}\right)\right)^2 \\ \sum_{t=1}^{T}\left(y_{it}-\beta_l l_{it}-\beta_k k_{it}-z_{it}\right)^2 \\ \sum_{t=1}^{T}\left(z_{it}-\sum_{k=1}^{K}\alpha_k^\varphi\phi_k^\varphi\left(z_{it-1}\right)\right)^2 \\ \left(a_{i1}-\sum_{k=1}^{K}\alpha_k^{g1}\phi_k^g\left(z_{i1}\right)\right)^2 \end{bmatrix} f\left(z_i^T \mid X_i^T,\theta\right)dz\right) \tag{41}$$

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 (40), $\theta^y = [\beta_k,\beta_l,\sigma_\epsilon]$, $\theta^h = \left[\alpha_1^h\ldots\alpha_K^h,\sigma_v\right]$, $\theta^g = \left[\alpha_1^g\ldots\alpha_K^g,\sigma_w\right]$, $\theta^\varphi = \left[\alpha_1^\varphi\ldots\alpha_K^\varphi,\sigma_\eta\right]$. Note that (41) are the integrated version of the unfeasible OLS regressions of the equations in (40). 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 (40), 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 OLS estimations of the parameters in θ.[21]

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

[21] For instance, if we specify our nonlinear functions as third-order polynomials, the model in (40) 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.

likelihoods of the equations in (40).

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

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:

$$\begin{cases} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \left(y_{it} - \beta_l l_{it} - \beta_k k_{it} - z_{it}^{(m)}\right)^2 \\ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \left(i_{it} - \sum_{k=1}^{K} \alpha_k^h \phi_k^h \left(z_{it}^{(m)}, k_{it}, a_{it}, \delta_t^h\right)\right)^2 \\ \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{m=1}^{M} \left(a_{it+1} - \sum_{k=1}^{K} \alpha_k^g \phi_k^g \left(z_{it}^{(m)}, k_{it}, a_{it}, \delta_t^g\right)\right)^2 \\ \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{m=1}^{M} \left(z_{it}^{(m)} - \sum_{k=1}^{K} \alpha_k^{\varphi} \phi_k^{\varphi} \left(z_{it-1}^{(m)}\right)\right)^2 \\ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \left(l_{it} - \sum_{k=1}^{K} \alpha_k^n \phi_k^n \left(z_{it}^{(m)}, k_{it}, a_{it}\right)\right)^2 \end{cases} \tag{42}$$

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, LP 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 (41).

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[22] 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

[22] 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.

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 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.

In a nutshell, our results show significant differences in the estimates of the firm production function and productivity when we compare our methodology robust to financial frictions with respect to the proxy variable approach. These differences fall in line with the predictions from the stylized model described in Section 2. Our results show that financial frictions 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 impact of self-financing is limited as convergence in the marginal product of capital between constrained and unconstrained firms is slow.

6.2.1 Production Functions

To highlight the importance of considering financial frictions in the estimation of the firm production function and the firm productive process, we compare the results of our two novel estimators that control for financial frictions (Proxy-IV and SEM) with OP -the proxy variable approach in Olley and Pakes [1996]- which uses investment as an auxiliary equation to recover productivity, and provides our main benchmark to previous literature. To have an alternative benchmark we also compare our results with LP -the proxy variable approach of Levinsohn and Petrin [2003]-, which uses intermediate inputs as an auxiliary equation to recover productivity.

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. By a similar argument, we expect the same type of bias in other methodologies relying on a proxy variable approach, such as LP.

Table 1 presents the results of the full estimation of the production function parameters (β_l, β_k) using the four methodologies. 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. In addition, by relying on the co-movements between wealth accumulation and investment decisions, after controlling for the current stock of net wealth, we can disentangle productivity shocks from transitory shocks that can temporarily affect investment and saving decisions. The estimate of β_l is 0.67 for OP, and, as expected, decreases significantly for the estimates that are robust to financial constraints and allow for shocks to the policy equations, to 0.44 in Proxy-IV and 0.46 in SEM.

Conversely, the opposite pattern holds for the elasticity of capital: the estimate of β_k is 0.35 for OP, and increases to 0.42 for Proxy-IV and 0.43 for SEM. Similar biases appears in the LP estimators which suggest that financial frictions are also present in the demand for intermediate goods as in Mendoza and Yue [2012] and Bigio and La'o [2020].[23]

It is worth noting that the estimates of the production function parameters are very similar with proxy-IV and SEM. Even though we show that the model is non-parametrically identified from data, in order to device tractable estimation methods we impose some restrictions in the empirical model. For instance, for the proxy-IV we assume that one of the policies is a quasi-linear function of degree 1 in productivity but we leave the other policy function and the distribution of productivity and shocks completely unrestricted. In SEM we allow all the policies to be nonlinear in all state variables including productivity but we parametrize the distribution of the shocks. Having robust results with both estimators suggest that neither the parametric assumption on the shocks nor the quasi-linear policy affects the estimation of the production function parameters.

Finally, the differences in the estimated input elasticities have relevant implications for the degree of returns to scale at the firm level, a crucial parameter to understand aggregate dynamics. In particular, OP results are consistent with constant returns to scale, while Proxy-IV and SEM both imply decreasing returns to scale with a span of control around 0.87. This figure lies on 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 relative to OP

[23] Our proxy variables estimates of the production function are similar to ones in Gandhi et al. [2020]. Their proxy variable estimates for a value-added production function with the Chilean data are 0.77 for β_l and 0.33 for β_k.

implies a larger entrepreneurial income share that can be retained by firms, which allows for a faster accumulation of wealth to overcome financial constraints.

To complement our results, we simulate data from an extended version of the stylized model presented in section 2 to confirm the biases of the proxy variable approach and the robustness of our proposed estimators. In line with the empirical estimates with the Chilean data, we set $\beta_k = 0.43$ and $\beta_l = 0.44$ in the calibrated model. See Appendix A3 for model and calibration details.

Table 2 presents the estimates for simulated data. As expected, OP delivers biased estimates, whereas Proxy-IV and SEM recover the true underlying parameters.[24] 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 validation to our novel estimators.

[24] As the model does not include intermediate inputs as required by the LP estimator, we only use the OP, Proxy-IV and SEM estimators.

	OP	LP	Proxy-IV	SEM
β_l	0.67	0.81	0.44	0.46
	0.008	*0.007*	*0.01*	*0.003*
β_k	0.35	0.33	0.42	0.43
	0.05	*0.04*	*0.01*	*0.007*
σ_ϵ	0.68	0.62	0.22	0.20
Observations	13516	13516	13516	13516
Firms	4867	4867	4867	4867

Table 1: Production Function Estimates from Microdata

Note: The table shows the Production function estimates from administrative data for Chile, using alternative methodologies: OP - Olley and Pakes [1996]-, LP- Levinsohn and Petrin [2003]-, and two estimators that control for financial friction, Proxy-IV and SEM.

	OP	Proxy-IV	SEM
β_l	0.505	0.443	0.442
β_k	0.397	0.424	0.431

Table 2: Production Function Estimates Using Simulated Data

Note: Production function estimates from simulated data using alternative methodologies: OP - Olley and Pakes [1996]-, and two estimators that control for financial friction, Proxy-IV and SEM. The model used to generate data is described in Appendix A.3.

6.2.2 Productivity Process

Figure 1 depicts the productivity distribution across firms for the proxy variable approach and our more general model (SEM). There are relevant differences in the dispersion of the estimated productivity distributions across both methodologies. In OP, the standard deviation of productivity is 0.18, significantly lower than 0.40 under SEM (see Table 3). We also find that the gap between ours and OP productivity estimates, i.e. the fraction by which true productivity is underestimated, is increasing in the productivity level of the firm. For instance, the coefficient of a linear regression between $z_{it}^{SEM} - z_{it}^{OP}$ and z_{it}^{SEM} is 0.7.

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 distribution relative to the methods that are robust to frictions.

As a robustness we also display the distribution of the estimated productivity when we do not assume normality in our nonlinear model (SEM-Quantile). Recall that in model (40) the productivity process is a nonlinear mean model in past productivity but it is separable in a shock that is normally distributed. We also implement a more non-parametric model where productivity is non-separable in the shock: $z_{it} = Q_z(z_{t-1}, \eta_{it})$ with $(\eta_{it} \mid z_{t-1})$ uniform distributed. This quantile model is a direct non-parametric model for the distribution of productivity. From figure 1 we can see that using the non-parametric quantile model (SEM-Quantile) delivers a very similar distribution than the one estimated using the nonlinear model with normal errors, so assuming separability and normality for the shocks seems to be inconsequential for our empirical results.

Table 3 also presents results for productivity persistence when we fit a linear model to the estimated productivity. The first row displays the autocorrelation of the estimated productivity ρ_z.[25] We can see that estimated autocorrelation is considerably lower under OP. The estimated value of ρ_z raises from 0.53 under OP to 0.87 in proxy-IV and to 0.85 under SEM, respectively. Also, as we can see in figure 7 in appendix A.4, the relationship between z_{it} and z_{it-1}, is nonlinear with a persistence that change depending on the value of initial productivity. Hence, assuming a linear productivity process as in Blundell and Bond [1998] or Shenoy [2020] may be at odds with the data.

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.

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

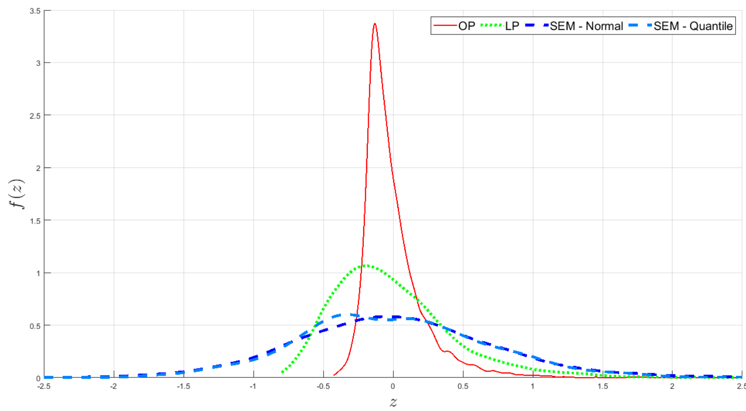

FIGURE 1: Estimated distribution of productivities

Note: The figure shows the estimated distribution of firm-level productivities using administrative microdata for Chile, under alternative methodologies: OP - Olley and Pakes [1996]-, LP- Levinsohn and Petrin [2003]-, and the SEM algorithm using Normal shocks and the SEM algorithm using a quantile model .

	OP	LP	Proxy-IV	SEM
ρ_z	0.53	0.89	0.87	0.85
	0.01	*0.01*	*0.01*	*0.01*
σ_η	0.18	0.15	0.30	0.39
Observations	13516	13516	13516	13516
Firms	4867	4867	4867	4867
R^2	0.37	0.53	-	0.70

TABLE 3: Estimated Parameters of the Productivity Process

Note: The table shows the estimated parameters for the firm-level productivity process from administrative microdata for Chile, using alternative methodologies: OP - Olley and Pakes [1996]-, LP- Levinsohn and Petrin [2003]-, and the two estimators that control for financial frictions, Proxy-IV and SEM.

6.2.3 Policy Functions

We now present the estimated policy functions, one of the main goals of our empirical exercise. 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 Proxy-IV approach, as presented in Figure 2. This method assumes that the effect of productivity on investment is linear, but allows the investment propensity to vary across firms with different levels of net wealth and capital. Panel (a) of Figure 2 depicts the estimated propensity as a function of wealth a_t. The relationship between the propensity and wealth is presented for two levels of capital k_t, associated to the 10th and the 90th percentiles of the capital distribution in the sample.

As expected, the response of investment to productivity 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.[26]

For all capital levels, the investment propensity is monotonically increasing in wealth. In all cases, propensities at low levels of wealth are significantly smaller than those for firms with the highest wealth. Additionally, for a given level of wealth, propensities are larger in firms with less capital.

For instance, for firms at the 90th percentile of the capital distribution, moving from the bottom to the top of the wealth distribution almost doubles the response of investment to income shocks, which jumps from 0.08 to 0.15. For firms with capital at the 10th percentile, investment propensities are significantly larger, although the impact of wealth is smaller, with responses ranging from 0.27 to 0.36 across the wealth distribution (which represent an increase in 30%). 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. Also, propensities change more with wealth at higher levels of capital as these firms are more leveraged.

As mentioned, for a given level of wealth, propensities are decreasing in capital. Differences are large: propensities for firms in the 10th capital percentile are three times larger than those of firms in the 90th percentile. For a synthetic firm with wealth equal to the median, the propensity decreases from 0.34 to 0.11 as we move up the capital distribution. The smaller investment response of high-capital firms to productivity shocks could reflect that they are close to their optimal capital levels, or that they face tighter financial constraints, as they have higher leverage given their wealth.

Panel (b) displays the marginal effects of wealth shocks on investment. The response of investment to wealth is decreasing in 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

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

in Section 2. All else equal, firms with low levels of wealth 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 4 displays the estimated average derivative effect of productivity on investment $\hat{\Phi}_t^h(a,k,z)$ (the investment propensity) in the full non-linear model. In this model the investment policy function is allowed to be non-linear on productivity z. Therefore, the three-dimensional graphs show how the response of investment to productivity changes for different combinations of a_t and z_t. Panels (a) and (b) present the results for different values of k.[27]

The main contribution of the non-linear method is that it can be used to analyze how responses vary as we change z. In general, the sensitivity of investment to productivity shocks increases with z. This is, for given values of wealth and capital, investment responses to productivity shocks are larger in ex-ante more productive firms. This general pattern is consistent with the implications of models of financial constraints in which firm productivity can affect firm lending contracts and the amount of borrowing available to the firm, as is the case in the models with earning-based constraints as in Drechsel [2019], Lian and Ma [2020] and di Giovanni et al. [2022] or forward-looking constraints as in Aguirre [2017], DeMarzo and Fishman [2007] and Brooks and Dovis [2020], in which firms can use their future cash-flows as collateral. This allows more productive firms to take more debt for a given level of net wealth, and hence expand investment more.

Figure 4 shows that the estimated investment response to productivity has a large degree of heterogeneity, with values ranging from 0.08 to 0.6. Propensities are smallest in low productivity firms with low levels of wealth, suggesting that these firms are less able to adjust investment in response to a positive productivity shock as they are financially constrained and can not rely on earnings.

Interestingly, the change in propensity along the wealth distribution also depends on the levels of productivity and capital. For low productivity firms, investment propensities are very sensitive to collateral in the form of net wealth. For instance, for a firm at the bottom of the productivity distribution the investment propensity doubles (from 0.08 to 0.16) as we move along the wealth distribution. This result is consistent, again, with the idea that collateral constraints play a crucial role in the investment response of low-productive firms. The story differs for high productivity firms, and depends on the stock of capital. For instance, for a synthetic firm at the top of the productivity distribution, but low capital (panel a), investment propensity is at its highest, and is insensitive to the level of collateral (wealth). This result suggests that wealth-based collateral constraint might be less relevant for very productive firms with low levels of capital, as these firms might rely on earnings as predicted by earning-based constraints models. In contrast, in firms with high capital the response of investment is increasing in wealth across all productivity levels. In fact, propensities increase 30% as we move along the wealth distribution, even in the case of high-productivity firms (from 0.3 to 0.4). A potential explanation is high-capital firms are more leveraged, so that future cash flows cannot

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

overcome financial frictions, and collateral plays a key role.

To have a taste of how propensities behaves using the actual combinations of state variables that we see in the data, we compute the propensity of each of the firms in our sample, and we plot it against the wealth-to-capital ratio $\frac{A}{K}$ in figure 5 panels (a)-(c). To analyze heterogeneity across different productivity levels, we use our estimated productivity variable to cluster firms in three different "productivity groups": (i) low-productivity firms with productivity below the 50 percentile of the productivity distribution, (ii) median-productivity firms with productivity between the 50 and 75 percentile and (iii) high-productivity firms with productivity above the 75 percentile. The data replicates the patterns suggested by the estimated policy functions. Investment propensities are increasing in $\frac{A}{K}$ and in z. As we can see, there is a positive relationship between the investment propensity and $\frac{A}{K}$ for all productivity levels, although the marginal impact of $\frac{A}{K}$ is decreasing in $\frac{A}{K}$. Moreover, propensities are larger for more productive firms.

For example, the investment propensity of low-productivity firms with little $\frac{A}{K}$ (panel (a)) is close to 0.1 on average. We can also see that for some firms with low-productivity and very low wealth to capital ratios, the propensity is close to zero, indicating that the investment of those firms does not react much to productivity shocks. However, the propensity increases up to 0.3 as we move along the distribution of $\frac{A}{K}$. This positive relationship between the investment propensity and $\frac{A}{K}$ is present across all three productivity groups. Panels (b) and (c) show that the propensities for median- and high- productivity firms start at 0.25 and 0.45, respectively. These propensities are much higher than the propensities of low-productivity firms with a similar level of $\frac{A}{K}$. As discussed earlier, a potential explanation is that these firms are more capable of adjusting investment because they can rely on current and future earnings. However, collateral constraints are also crucial for these firms, as propensities increase for firms with higher levels of $\frac{A}{K}$. For median- and high-productivity firms, the propensity increases up to 0.4 and 0.6 (on average) and stabilizes around these numbers. The positive relation between investment propensities and $\frac{A}{K}$ for high-productivity firms suggests that earnings are not sufficient to self-finance all the investment, and a combination of earnings and wealth are important for all firms with low levels of wealth as in di Giovanni et al. [2022]. For high levels of $\frac{A}{K}$, propensities are roughly constant, as these firms are probably not constrained, and investment responses are close to optimal.

These graphic results are tested more formally in Table 4, which shows a quadratic regression between investment propensities and the wealth-to -capital ratio, for each productivity group. The estimates confirm that: (i) there are significant differences in the intercepts of each productivity groups, with a higher intercept for more productive firms, (ii) for each productivity level the slope is positive, indicating that propensities are increasing in $\frac{A}{K}$, and (iii) the quadratic term is negative, indicating that the marginal effect of $\frac{A}{K}$ is dampened as firms become less leveraged.

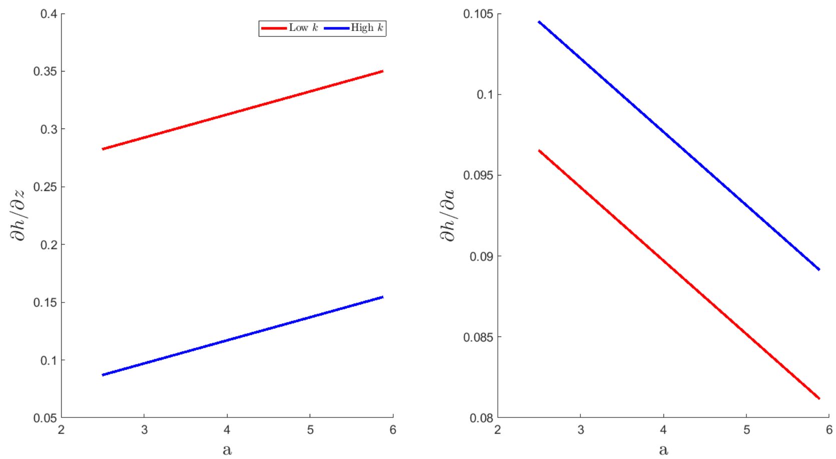

FIGURE 2: Investment propensity in response to productivity and wealth

Notes: Panel (a) exhibits the estimated derivative effect of productivity in the investment policy function using the Proxy-IV approach. The figure shows how the effect changes as wealth varies, and is evaluated at two different capital levels. Panel (b) exhibits the estimated derivative effect of the stock of wealth (lagged wealth) in the investment policy function using the Proxy-IV approach. The figure shows how the effect changes as wealth varies, and is evaluated at three different capital levels.

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 3 depicts the marginal effects of productivity and wealth shocks on wealth accumulation under the Proxy-IV approach. Panel (a) displays the estimated marginal effect of productivity on wealth accumulation (the wealth accumulation propensity) as a function of wealth a for two different values of capital k. The propensity is positive and significant for almost all of the possible combinations of state variables.[28] In contrast to the investment policy function, the effect of productivity on wealth accumulation is now strictly decreasing on wealth, across all capital levels: the elasticity falls in 10 percentage points as we move along the wealth distribution, a reduction of around 25%.

As before, this result is qualitatively consistent with the self-financing channel. Firms that have less collateral, and therefore are more likely to be constrained, have a stronger incentive to boost up their savings when they experiment 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 wealth persistence is nonlinear and increasing in wealth: wealth is more persistent at higher levels of current wealth.

Nonlinear effects Panels (c) and (d) of Figure 4 display the estimated average derivative effect of productivity on wealth accumulation (the nonlinear propensity) $\hat{\Phi}^{g}_{t+1}(a, k, z)$ using SEM. As before, this method allows the wealth accumulation policy function to be non-linear in productivity z. Hence, the three-dimensional graph presents how savings propensities change for different combinations of a and z. We can see that the general patterns from figure 3 hold. In almost all cases, the average derivative effect of productivity on savings decreases as a grows. Similarly, for a given combination of wealth and productivity, in most cases propensities are increasing in capital, consistent with the theoretical impact of larger leverage.

Regarding non-linearities, for a given level of capital, propensities are largest in firms that are highly productive but hold little wealth. In fact, the savings propensity to productivity shocks in firms on the upper end of the productivity distribution and the lower end of the wealth distribution is close to 1. This is, earnings shocks for highly productive but severely constrained firms are almost completely saved, as the value of alleviating the constraint is comparatively large. As predicted by theoretical models with the self-financing channel, the propensity decreases as we move along the wealth distribution (up to half of the value) since high-wealth firms are less constrained and have less incentives to save.

The savings propensity is also heterogeneous in productivity, as it is significantly lower for low productivity firms, which are probably less constrained and has less incentives to save. However, at

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

low wealth levels, even low-productivity firms save a considerable share of the earnings associated to a productivity shock (the propensity is between 0.3-0.5) when wealth is low. This propensity decreases to 0.1 as wealth increases.

We see similar patterns when we characterize saving propensities using the actual combination of all state variables that we see in the data (including estimated productivity) in figure 5 panels (d)-(f). Propensities are positive for all firms in the data and are increasing in productivity and decreasing in wealth. Again, the propensity is higher for high-productivity firms with low levels of wealth. As we discussed above, even for high-productivity firms that can also rely on earnings, the investment propensity increases with wealth (see figure 5-(c)), so these firms also have strong motives to save and accumulate wealth (see figure 5-(f)). The higher wealth accumulation propensity for very productive firms is consistent with the insights of a model where collateral and earning-based constraints interact. As emphasized in di Giovanni et al. [2022], even with earnings-based constraints, more productive firms are more financially constrained (in terms of the distance to their optimal size) for a given level of wealth (net worth) and are the ones that benefit the most from an additional unit of wealth. In the benchmark parametrization of di Giovanni et al. [2022], a high-productivity firm has better access to credit than a low-productivity firm, for the same level of wealth, through the earning-based constraint. However, the high-productivity firm is still "more constrained" since it is further away from its optimal capital level.

Moreover, in models with collateral and earning-based constraints, the marginal effect of wealth on investment is increasing in productivity: an increase in wealth reduces borrowing constraints directly through the standard collateral constraint channel, generating an increase in investment and production, which in turn reduces borrowing constraints through the earnings-based constraint channel. This indirect channel is more potent for high-productivity firms than for low-productivity firms since their earnings increase more with the initial increase in wealth. The latter creates a higher incentive for high-productivity firms with low levels of wealth to increase savings and accumulate wealth in response to a positive productivity shock.

These graphic results are confirmed the estimates in Table 5, which runs the quadratic regression of savings propensities on the wealth-to-capital ratio. As expected, (i) intercepts are increasing in productivity, indicating that more productive firms have a higher marginal savings propensity; (ii) the marginal effect of $\frac{A}{K}$ on savings propensity is negative, and (iii) the marginal effect becomes more muted as $\frac{A}{K}$ grows.

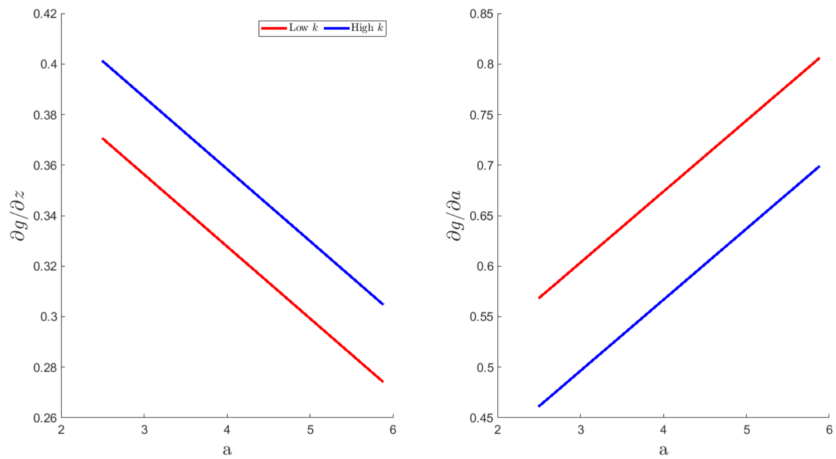

Figure 3: Wealth accumulation propensity in response to productivity and wealth

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 shows how the effect changes as wealth varies, and is evaluated at two different capital levels. Panel (b) of the figure exhibits the estimated derivative effect of the stock of wealth (lagged wealth) in the wealth accumulation policy function using the Proxy-IV approach. The figure shows how the effect changes as wealth varies, and is evaluated at two different capital levels.

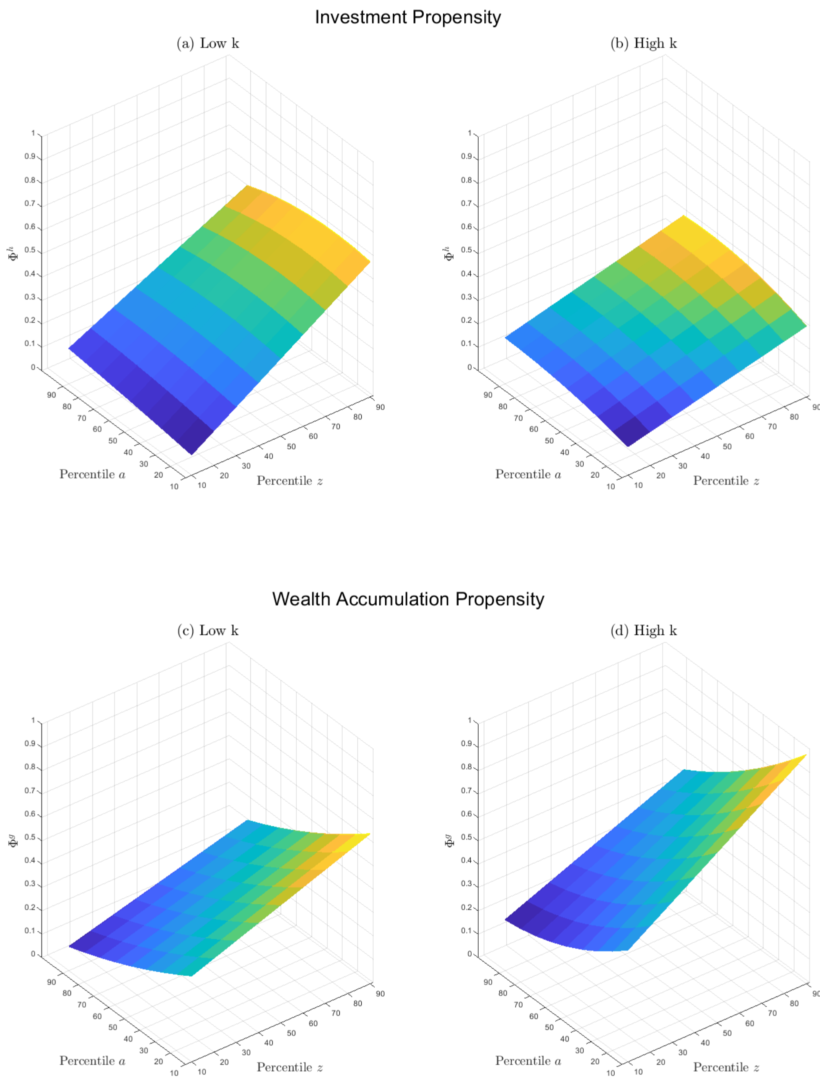

Figure 4: Nonlinear model: Investment and Wealth accumulation propensities to productivity

Notes: The figure exhibits the estimated derivative effect of productivity in the investment policy function (panels a and b) and the estimated derivative effects of productivity in the wealth accumulation policy (panels c and d) 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 wealth for two different values of capital.

Figure 5: Investment and Wealth accumulation propensities in response to productivity

Notes: The figure exhibits how the investment propensity and the wealth accumulation propensity vary along the distribution of $\frac{A}{K}$ in the micro data, for different values of productivity. Each point represents the propensity of each particular firm evaluated at its actual value of a, k and z. Figures (a), (b) and (c) are the investment propensities for low-, median- and high-productivity firms. Figures (d), (e) and (f) are the wealth accumulation propensities for low-, median- and high-productivity firms.

	Low Productivity	Medium Productivity	High Productivity
$\hat{\beta}_0$	0.143	0.263	0.507
	(0.006)	(0.004)	(0.011)
$\hat{\beta}_1$	0.067	0.139	0.134
	(0.017)	(0.009)	(0.017)
$\hat{\beta}_2$	-0.014	-0.038	-0.027
	(0.007)	(0.004)	(0.005)

Table 4: Effects of $\frac{A}{K}$ on Investment Propensities: Quadratic Regressions

Note: The table shows the estimated parameters of a quadratic regression of investment propensities at the firm level and the net wealth-to-capital ratio, for three productivity groups.

	Low Productivity	Medium Productivity	High Productivity
$\hat{\beta}_0$	0.295	0.432	0.631
	(0.004)	(0.003)	(0.006)
$\hat{\beta}_1$	-0.093	-0.182	-0.235
	(0.012)	(0.006)	(0.009)
$\hat{\beta}_2$	0.013	0.037	0.041
	(0.006)	(0.002)	(0.003)

Table 5: Effects of $\frac{A}{K}$ on Wealth Accumulation Propensities: Quadratic Regressions

Note: The table shows the estimated parameters of a quadratic regression of wealth accumulation propensities at the firm level and the net wealth-to-capital ratio, for three productivity groups.

6.2.6 Quantitative Implications: MPKs convergence

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 and our estimates of firm productivity and the production function to calculate the initial MPK of two firms that share the same level of initial productivity but have different levels of initial wealth and capital. 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 and labor 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.

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 twice as large as 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.

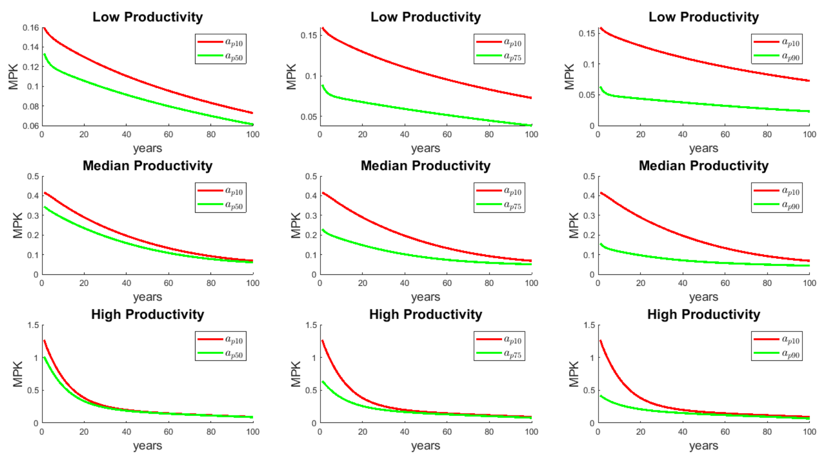

Figure 6: Convergence in the marginal product of capital across firms

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.

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 0.67 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.

8 Acknowledgements

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. 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.

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 Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

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.

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.

Saki Bigio and Jennifer La'o. Distortions in production networks. *The Quarterly Journal of Economics*, 135(4):2187–2253, 2020.

Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143, 1998.

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.

Sylvain Catherine, Thomas Chaney, Zongbo Huang, David Alexandre Sraer, and David Thesmar. Quantifying reduced-form evidence on collateral constraints. *forthcoming Journal of Finance*, 2018.

Russell W Cooper and John C Haltiwanger. On the nature of capital adjustment costs. *The Review of Economic Studies*, 73(3):611–633, 2006.

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.

Julian di Giovanni, Manuel García-Santana, Priit Jeenas, Enrique Moral-Benito, and Josep Pijoan-Mas. Government procurement and access to credit: Firm dynamics and aggregate implications. 2022.

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.

Thomas Drechsel. Earnings-based borrowing constraints and macroeconomic fluctuations. *manuscript, London School of Economics*, 2019.

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

Vito D Gala, Joao F Gomes, and Tong Liu. Investment without q. *Journal of Monetary Economics*, 116:266–282, 2020.

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.

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.

Enrique G Mendoza and Vivian Z Yue. A general equilibrium model of sovereign default and business cycles. *The Quarterly Journal of Economics*, 127(2):889–946, 2012.

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.

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.

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

Appendix A.1: The bias in the OP estimator under financial frictions

Here we provide a more formal assessment of the types of biases emerging under OP in the context of financial frictions, and, by extension, on other methodologies relying on the proxy variable approach. As mentioned, 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}:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \tag{43}$$

where ε_{it} is measurement error in value added. 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 that h is monotonic in z_{it}, we can invert the policy function to recover productivity as $z_{it} = h^{-1}\left(i_{it}, k_{it}\right)$ and construct valid moment conditions. For instance, we can rewrite (43) as:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + h^{-1}\left(i_{it}, k_{it}\right) + \varepsilon_{it}. \tag{44}$$

Since ε_{it} is assumed to be uncorrelated with the inputs, OP propose to approximate $h^{-1}\left(i_{it}, k_{it}\right)$ 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:

$$\hat{\pi}_t\left(i_{it}, k_{it}\right) = \beta_k k_{it} + \rho\hat{\pi}_{t-1}\left(i_{it-1}, k_{it-1}\right) - \rho\beta_k k_{it-1} + \eta_{it} + \hat{\varepsilon}_{it} \tag{45}$$

where $\hat{\pi}_t\left(i_{it}, k_{it}\right)$ 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.

As discussed, observed differences in investment between firms in the data are intepreted as productivity differentials under OP. Hence, by controlling for investment in the production function, OP can eliminate the endogeneity problem and get consistent estimates of β_l and β_k. However, under borrowing constraints, differences in investment between firms might not only reflect differences in productivity, but also differences in borrowing capacity.

In the model with financial frictions described in Section 2, the investment function arising from (4) depends not only on productivity and initial capital, but also on net-worth, through its direct 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:

$$y_t \quad = \quad \beta_l l_{it} + \beta_k k_{it} + h^{-1}(i_{it}, k_{it}, a_{it}) + \varepsilon_t \tag{46}$$

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 (44). 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):

$$l_{it} = c_l + \frac{1}{1-\beta_l}\left(\beta_k k_{it} + w + h^{-1}(i_{it}, k_{it}, a_{it})\right) \tag{47}$$

Therefore, after controlling for k_{it} and i_{it}, the correlation between l_{it} and the OP residual is positive.[29][30] Because OP cannot control for a fraction of productivity, which goes into the residual term when applying OP to (46), 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.[31] 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 (45)). 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 (45) 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

[29]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 $\tilde{h}_a E[l_{it}a_{it} \mid i_{it}, k_{it}] > 0$ and $\tilde{h}_a E[k_{it}a_{it-1} \mid \hat{\pi}_{t-1}, k_{it-1}] < 0$.

[30]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.

[31]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.

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.

Appendix A.2

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:

$$f(y_{it}, i_{it} \mid a_{it+1}, X_{it}) = \int f(y_{it} \mid z_{it}, i_{it}, a_{it+1}, X_{it}) f(i_{it} \mid z_{it}, a_{it+1}, X_{it}) f(z_{it} \mid a_{it+1}, X_{it}) dz_{it}, \quad (48)$$

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 (48) as

$$f(y_{it}, i_{it} \mid a_{it+1}, X_{it}) = \int f(y_{it} \mid z_{it}, k_{it}, l_{it}) f(i_{it} \mid z_{it}, X_{it}) f(z_{it} \mid a_{it+1}, X_{it}) dz_{it} \quad (49)$$

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 (49) can be expressed in terms of integral operators:

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

Integrating both sides of (49) with respect to I:

$$L_{y|a,X} = L_{y|z,k,l} L_{z|a,X} \tag{51}$$

From (51), 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 (51) in (50) we get:

$$L_{y;I|a,X} L_{y|a,X}^{-1} = L_{y|z,k,l} D_{I;z|X} L_{y|z,k,l}^{-1} \tag{52}$$

Note that the observed quantity $L_{y;I|a,X} L_{y|a,X}^{-1}$ in (52) 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 (52). 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 (52), the inverse of $L_{y|a,X}$ has to exist. Looking at (51) 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 (13) 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 (52) 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 (30) 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 (10)-(11), 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 (30), by deconvolution, we can identify the distribution of z_{it}.

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

$$\tilde{y}_{it} = \psi(\tilde{y}_{it-1}) + \zeta_{it} \tag{53}$$

$$\tilde{y}_{it-1}\tilde{y}_{it} = \phi(\tilde{y}_{it-1}) + \upsilon_{it} \tag{54}$$

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 (53) and (54): $E[\tilde{y}_{it} - \psi(\tilde{y}_{it-1}) \mid \tilde{y}_{it-2}] = 0$ 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:

$$E(\tilde{y}_{it} \mid z_{it-1}) = E(\psi(\tilde{y}_{it-1}) \mid z_{it-1}), \tag{55}$$

$$z_{it-1}E(\tilde{y}_{it} \mid z_{it-1}) = E(\phi(\tilde{y}_{it-1}) \mid z_{it-1}). \tag{56}$$

Equation (55) 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 (56) 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 (53) and (54) and data on $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$, we can use equation (55) and (56) to identify the distribution of ε_{it-1} for a fixed value of z:

$$E_{\varepsilon_{it-1}}\left[z\psi\left(z+\varepsilon_{it-1}\right)\right] = E_{\varepsilon_{it-1}}\left[\phi\left(z+\varepsilon_{it-1}\right)\right] \tag{57}$$

By deconvolution we can recover the density of ε_{it-1} from (57). 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 $t = \{2, \dots T-1\}$. 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 (10)-(11), 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

$$f(a_1, k_1 \mid \tilde{y}) = \int f(a_1, k_1 \mid z_1, \tilde{y}) f(z_1 \mid \tilde{y}) dz_1, \tag{58}$$

by *assumption 1*, $f(a_1, k_1 \mid z_1, \tilde{y}) = f(a_1, k_1 \mid z_1)$ equation (58) can be expressed as:

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

Equation (59) can be rewritten as the following moment restriction:

$$f(a_1, k_1 \mid \tilde{y}) = E\left[f(a_1, k_1 \mid z_1) \mid \tilde{y}_i = \tilde{y}\right] \tag{60}$$

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 $f(a_1, k_1 \mid z_1)$ is identified from (60). 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:

$$f(a_2 \mid a_1, k_1, \tilde{y}) = \int f(a_2 \mid z_1, a_1, k_1) f(z_1 \mid a_1, k_1, \tilde{y}) dz_1 \tag{61}$$

where $f(a_2 \mid a_1, k_1, \tilde{y}) = f(a_2 \mid z_1, a_1, k_1)$. Equation (61) can be rewritten in terms of the following moment restriction:

$$f(a_2 \mid a_1, k_1, \tilde{y}) = E\left[f(a_2 \mid z_1, a_1, k_1) \mid a_{i1} = a_1, k_{i1} = k_1, y_i = y\right] \tag{62}$$

Equation (62) 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

$$f(k_2 \mid a_1, k_1, \tilde{y}) = E\left[f(k_2 \mid z_1, a_1, k_1) \mid a_{i1} = a_1, k_{i1} = k_1, y_i = y\right] \tag{63}$$

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:

$$f(a_3 \mid a_2, k_2, a_1, k_1, \tilde{y}) = \int f(a_3 \mid z_2, a_2, k_2) f(z_2 \mid a_2, k_2, a_1, k_1, \tilde{y}) dz_1 \tag{64}$$

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 64. 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.3: Estimations using Simulated Data

We use 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.[32] Second, we introduce adjustment costs to capital. We choose a standard quadratic function with a parameter η determining its size.[33] 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.

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.[34]

Model simulations can also be used to explore how the biases of the production function estimates vary with the intensity of financial frictions, i.e. with different values of the parameters governing the collateral constraint, λ y λ_z. For instance, when λ decreases from 2.5 -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. 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.

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

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

[34] 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.

Appendix A.4

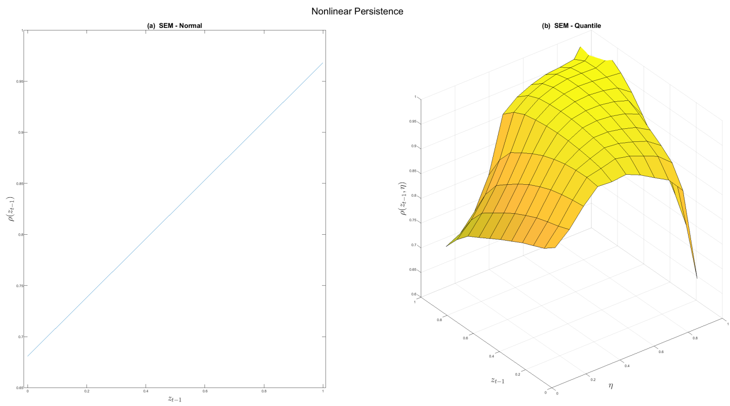

Figure 7: Estimated persistence of productivity

Note: The figure shows the estimated nonlinear persistence of firm-level productivity using administrative microdata for Chile. The first plot displays the persistence of estimated productivity using the model with normal errors along the distribution of initial productivity, whereas the second plot displays the persistence of estimated productivity using the quantile model where the size and the sign of the shock might affect the persistence depending on the initial value of productivity.

Appendix A.5

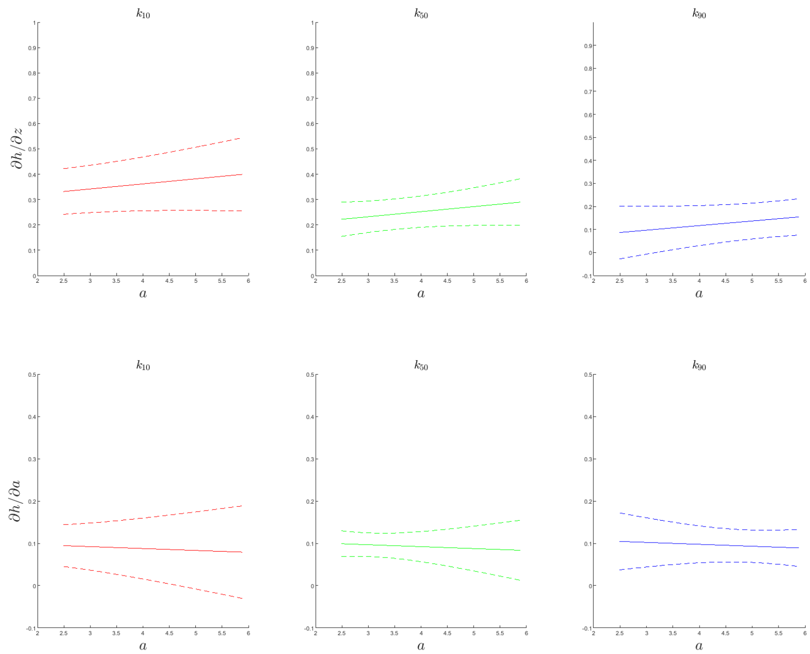

Figure 8: Confidence Intervals: Marginal effect of productivity and wealth on investment

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).

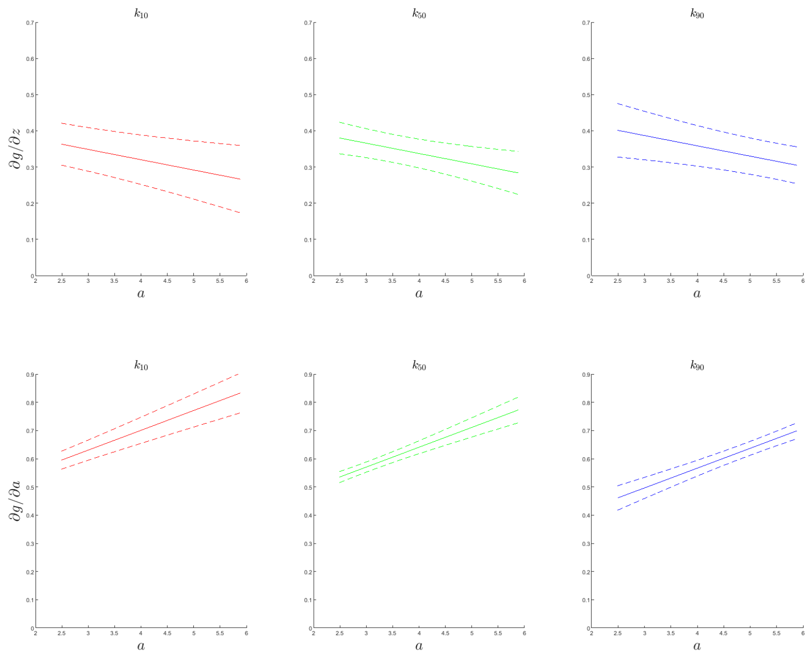

FIGURE 9: Confidence Intervals: Marginal effect of productivity and wealth on wealth accumulation

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).

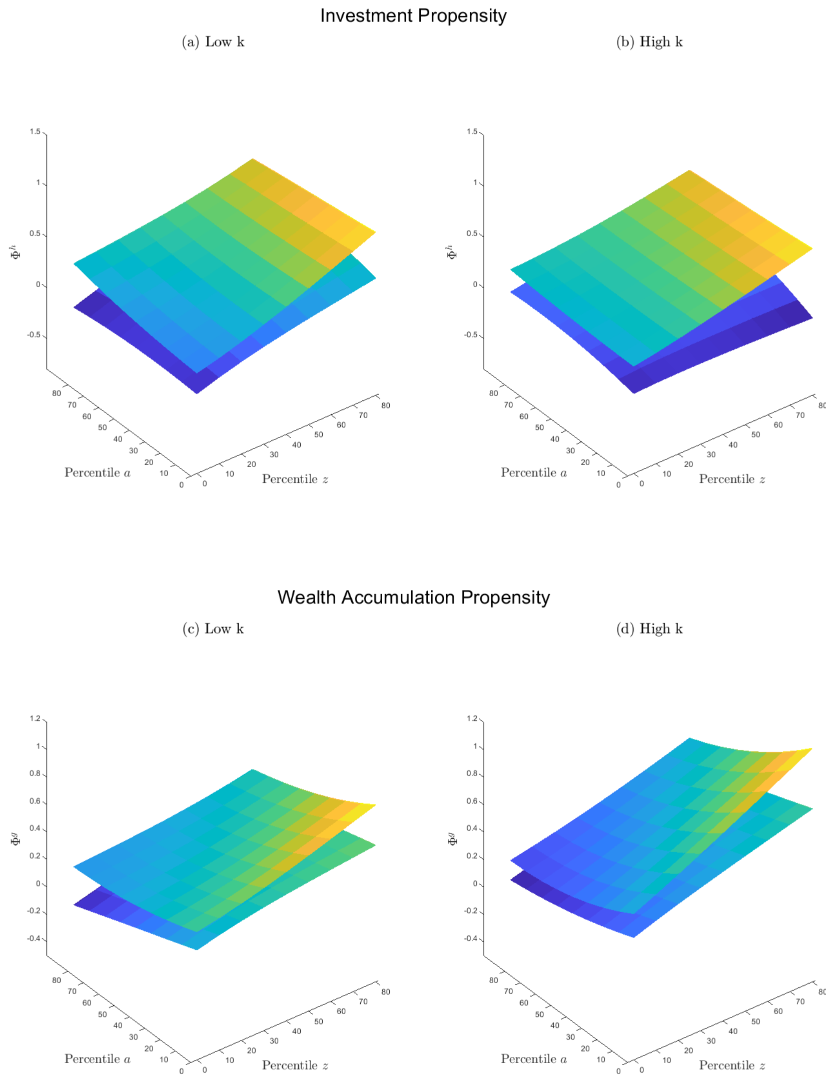

Figure 10: Confidence Intervals: Investment and Wealth Accumulation propensities

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