

DOCUMENTOS DE TRABAJO

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

Alvaro Aguirre
Matias Tapia
Lucciano Villacorta

N° 904 Marzo 2021

BANCO CENTRAL DE CHILE





BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

La serie Documentos de Trabajo es una publicación del Banco Central de Chile que divulga los trabajos de investigación económica realizados por profesionales de esta institución o encargados por ella a terceros. El objetivo de la serie es aportar al debate temas relevantes y presentar nuevos enfoques en el análisis de los mismos. La difusión de los Documentos de Trabajo sólo intenta facilitar el intercambio de ideas y dar a conocer investigaciones, con carácter preliminar, para su discusión y comentarios.

La publicación de los Documentos de Trabajo no está sujeta a la aprobación previa de los miembros del Consejo del Banco Central de Chile. Tanto el contenido de los Documentos de Trabajo como también los análisis y conclusiones que de ellos se deriven, son de exclusiva responsabilidad de su o sus autores y no reflejan necesariamente la opinión del Banco Central de Chile o de sus Consejeros.

The Working Papers series of the Central Bank of Chile disseminates economic research conducted by Central Bank staff or third parties under the sponsorship of the Bank. The purpose of the series is to contribute to the discussion of relevant issues and develop new analytical or empirical approaches in their analyses. The only aim of the Working Papers is to disseminate preliminary research for its discussion and comments.

Publication of Working Papers is not subject to previous approval by the members of the Board of the Central Bank. The views and conclusions presented in the papers are exclusively those of the author(s) and do not necessarily reflect the position of the Central Bank of Chile or of the Board members.

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

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

Alvaro Aguirre
Central Bank of Chile

Matias Tapia
Central Bank of Chile

Lucciano Villacorta
Central Bank of Chile

Abstract

The ability of firms to accumulate wealth and build collateral is key to overcome financial frictions. The strength of this self-financing channel depends on the productivity process faced by firms and the parameters associated with the production function, and may be quantified by the elasticity of wealth accumulation to productivity shocks. We propose a framework to jointly estimate the production function, the productivity process, and the wealth accumulation process that is robust to financial frictions. We show that standard methods (e.g. Olley-Pakes) fail under financial frictions: they overestimate the labor elasticity and underestimate the capital elasticity of the production function, and underestimate the persistence and dispersion of the productivity process. We apply our method to the universe of Chilean firms and confirm these predictions, with factor elasticities varying around 25%, and productivity volatility more than doubling. We find evidence that is in line with the self-financing channel: (i) the reaction of investment to productivity shocks is contingent on the stock of collateral, with larger responses from unconstrained firms; (ii) highly productive firms accumulate wealth after positive and persistent productivity shocks, with a larger effect in wealth-poor firms.

Resumen

La acumulación de riqueza para aumentar el colateral permite a las firmas sobreponerse a fricciones financieras que pueden limitar su inversión. La fuerza de este canal de autofinanciamiento, que puede ser cuantificado por la elasticidad de la acumulación de riqueza a los shocks de productividad, depende del proceso de PTF que enfrentan las firmas y los parámetros asociados con la función de producción. En este trabajo proponemos un nuevo marco empírico para la estimación conjunta de la función de producción, el proceso exógeno que sigue la TFP y el proceso de acumulación de riqueza que decide la firma. Crucialmente, este método es robusto a la existencia de fricciones financieras. Como mostramos, los métodos estándar de estimación de funciones de producción y dinámicas de PTF (por ejemplo, Olley-Pakes) están sesgados si existen restricciones financieras: a diferencia de nuestro marco empírico, sobreestiman la elasticidad del trabajo y subestiman la del capital en la función de producción, y subestiman la persistencia y la dispersión de la productividad. Una aplicación de nuestra metodología a una base de datos construida con registros administrativos del universo de firmas en Chile confirma nuestras predicciones teóricas: al comparar esta metodología con la de Olley-Pakes, las elasticidades de factores estimadas cambian alrededor de 25% y la volatilidad de la productividad casi se dobla. También encontramos evidencia coherente con la

* We are grateful to Daniel Ackerman, Manuel Arellano, Stephane Bonhomme, Francisco Buera, Andrea Caggese, Manuel García-Santana, Emmanuel Farhi, Ivan Fernandez-Val, Benjamin Moll, Josep Pijoan-Mas, Ludwig Straub, Alonso Villacorta, Gianluca Violante, Fabrizio Zilibotti and attendants to seminars at Santiago Macroeconomic Workshop 2020, LACEA-LAMES 2019, PUC Chile, Central Bank of Chile, the 2020 World Congress of the Econometric Society, European Economic Association 2020 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

existencia del canal de autofinanciamiento: (i) la respuesta de la inversión a shocks de productividad depende el nivel de colateral, con una mayor sensibilidad en firmas no restringidas; (ii) firmas altamente productivas acumulan riqueza luego de shocks positivos y persistentes de productividad, con un efecto mayor en firmas con bajo nivel de riqueza.

1 Introduction

A large literature has studied the potential role of financial frictions in explaining cross-country differences in aggregate income, investment, and productivity. Among other consequences, incomplete access to external financing can prevent productive firms with low levels of wealth from operating at their optimal scale. This can lead to an inefficient allocation of factors that lowers aggregate productivity. However, in a dynamic setup, these aggregate effects can be dampened, as firms might endogenously grow out of their financial constraints by accumulating wealth and building collateral—the so-called self-financing channel. The elasticities of investment and wealth accumulation by firms to productivity shocks, and how they depend on the amount of available collateral, are particularly relevant for this mechanism. These elements, and hence the scope of the mitigating effect of the self-financing channel, depend crucially in the persistence and volatility of the firm-level productivity process as well as on the parameters governing preferences and technology.¹

Thus, a precise empirical assessment of the wealth and productivity processes is essential to understand the quantitative effects of financial frictions. However, an analysis of these objects using micro data is currently absent in the literature. Moreover, the standard approaches to estimate the parameters of the firm’s production function and the productivity process are invalid in the presence of financial frictions. This paper explores empirically the strength of the self-financing channel by developing an analytical framework, robust to the presence of financial constraints, to jointly estimate the firm’s production function, the firm-productivity process, and the wealth accumulation process. We implement our novel estimation method on a census of administrative records for formal firms in Chile from 2006 to 2016. The dataset provides information on the firm’s production process and its investment decisions, as well as measures of the firm’s net worth. Our results provide empirical evidence of the presence of financial frictions and the self-financing channel. We identify crucial parameters—such as the response of wealth accumulation to productivity shocks—that allow to discipline quantitative macro models of firm dynamics and financial frictions and to quantify the strength of the self-financing channel.²

An important challenge for our framework is that firm productivity is unobservable. There is a large literature devoted to the estimation of firm production functions and, consequently, of measures of TFP at the firm level. Prevalent methods rely on a proxy variable approach to recover productivity using the firm’s inputs decisions (see [Akerberg et al., 2015](#), for a review). For instance, in their seminal contribution, [Olley and Pakes \(1996\)](#) (OP from now on), recover productivity by inverting an investment demand function which is then used as a nonparametric

¹For instance [Buera and Shin \(2011\)](#) and [Moll \(2014\)](#) focus on the consequences of the degree of persistence of productivity shocks on the strength of the self-financing channel.

²The results of our structural macro model are preliminary, so they are skipped in this version of the paper.

control in the production function regression. One of the main contributions of our paper is to show neatly how these methods fail when financial frictions are present, and to propose an alternative strategy that is robust in a frictional environment. This strategy exploits the theoretical insights behind the self-financing channel, and jointly estimates the production function, the productivity process, and investment and wealth accumulation processes. In line with these predictions we obtain an elasticity of capital in the production function that is 23% larger, and an elasticity of labor that is 30% lower, than OP, when we control for financial frictions. We also find a significantly larger dispersion and persistence of the productivity process relative to OP. Additionally, and consistent with the predictions of models of financial frictions, we document that the marginal effect of productivity on investment is increasing in wealth, and a positive and significant marginal effect of productivity on wealth accumulation, which is stronger for more constrained firms. We take this as evidence that the self-financing channel is active in the data.

Our novel empirical framework consists of a firm production function, a non-linear firm investment policy rule, and a non-linear firm wealth accumulation policy rule. These three equations depend on the latent firm-level productivity process. As in the proxy variable framework, productivity follows a flexible non-linear Markovian process of order one. Our framework has *three main departures* from the proxy variable approach initiated by OP.

First, we include the firm’s stock of wealth as an additional state variable in the investment equation to control for financial constraints, in accordance with the theoretical models of financial frictions in which wealth is pledged as collateral. This is precisely the main reason why OP fails since, in the context of these models, investment variability cannot be completely explained by productivity and initial capital. Intuitively, the OP method assigns differences in investment across firms in the data to differences in unobserved productivity. However, under financing constraints, differences in investment between firms are not only reflecting differences in productivity but also might be driven by differences in borrowing capacity.³

Second, together with the investment equation we estimate the wealth accumulation policy function. This function is our main focus, as it plays a fundamental role to understand the scope of the self-financing channel. Moreover, under financial frictions, the behavior of firms’ wealth may be more informative about unobserved productivity than the evolution of investment, since the collateral of constrained firms is more sensitive to productivity shocks than their investment.

Third, in order to provide reliable empirical estimates of the two policy functions, we allow

³It is worth to note that the failure of OP in the presence of omitted variables in the investment policy function is well known (see for instance [Shenoy, 2020](#)). What we do in this paper is to clearly describe the consequences of this failure and to propose a new empirical framework that solves for it, and is consistent with the widely used macroeconomic models with financial frictions.

for unobservable shocks, in addition to the latent productivity shock.⁴ Unobservable shocks in the investment function may be due to a stochastic component of borrowing costs that is not captured by firm wealth. In the wealth accumulation function, shocks may be capturing idiosyncratic returns to firm assets.

A relevant methodological contribution of this paper is to develop a tractable framework that allows us to take to micro data important aspects of models with rich heterogeneity without the need of specifying functional forms for preferences. In contrast to fully-specified structural approaches, which can be computationally problematic to estimate in environments with large heterogeneity, we model our empirical policy rules non-parametrically, leaving functional forms unrestricted. This modeling approach allows us to get a rich picture of the joint relationship of investment decisions, wealth accumulation, and productivity shocks directly from the data. Although the empirical model is unable to provide direct policy counterfactuals, our estimated parameters may be directly or indirectly used 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 firms’ 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 the financial constraint.⁵

Identification and estimation of our nonlinear model cannot be handled within the proxy variable framework, since our nonlinear policy rules are more flexible and include unobservable shocks in addition to the latent productivity process. Also, a key aspect of our model is to identify and estimate the non-linear policy functions. We show that nonparametric identification of the production function, the productivity process and the policy functions of our model can be established by building on recent developments on nonlinear panel data models with latent variables (Hu and Schennach (2008), Hu and Shum (2012), and Arellano et al. (2017)).

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

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

⁵Our empirical policy functions provide “identified moments” such as the average causal effect of productivity on investment and on wealth accumulation that are useful to estimate parameters of structural models or discriminate between structural models (Nakamura and Steinsson (2018)).

We show that for parsimonious, yet flexible, versions of policy functions, an IV estimation strategy within the proxy variable framework delivers consistent estimates of the model, following the arguments in the nonparametric identification strategy. For more general policy functions, we consider a tractable estimation strategy that is well-suited for non-linear panel data models with latent variables by adapting the approach in [Arellano et al. \(2017\)](#) to a production function setup. The estimation approach is a stochastic EM algorithm that combines simulation methods and GMM estimation. A key aspect of our method is to combine the production function with the information of investment decisions and wealth accumulation dynamics to construct the posterior distribution of the productivity process.⁶

A virtue of our non-linear framework is the possibility to uncover new empirical results for both the production function literature and for the macro literature with financial frictions. Regarding the production function estimates, the results show that the estimated average effect of capital in the production function increases from 0.35 when using OP to 0.43 when we consider financial frictions in the estimation. On the other hand, the estimated marginal effect of labor in the production function decrease from 0.65 in OP to 0.44 when controlling for financial frictions. Using a firm dynamic model with collateral constraints, we show analytically the source of the biases associated to the OP estimator. Intuitively, the presence of financial constraints generate differences in investment, capital and in output between equally productive firms. The OP approach interprets differences in observed investment across firms directly as differences in unobserved productivity. Although the implied variation in output should be captured by variations in capital, the OP approach assigns it to variations in productivity since the proxy equation implies so. As a result, the OP productivity proxy captures an important part of the effect of capital on output, underestimating the marginal effect of capital.

Regarding labor, if frictions are less severe in that market than for investment, OP reads a financially constrained firm as a low-productivity firm that hires too many workers, and produce too much output, relative to its proxy-OP productivity, and hence assigns to labor a high relevance in output. The result is an overestimation of the labor elasticity. Furthermore, the differences in the estimations of factor elasticities translate to significant differences in the measure of returns to scale. In particular, OP results are consistent with constant return to scale whereas our estimates imply decreasing returns to scale with a span of control of 0.87.

Regarding the firm-level productivity process we show that OP significantly underestimates both the dispersion and the persistence in productivities relative to our approach. The 90th to 10th productivity ratio of the firm distribution in any given year is more than twice as large under our methodology. Moreover, the standard deviation of productivity under OP is 0.16, which raises to 0.42 when we control for financial frictions. These results are also consistent with

⁶We show with simulated data and actual data that our estimation method delivers similar results to the GMM proxy variable approach for the OP model.

financial frictions, as relatively productive firms, expected to be more financially constrained according to the canonical model, show larger investment gaps with respect to their optimal levels, leading OP to underestimate their productivity relative to unproductive firms and to shrink the distribution of productivity. Regarding persistence, the first-order autoregressive estimated parameter raises from 0.56 in OP to 0.82 in our model. Given that OP estimated productivity is a combination of true productivity and financial constraints, the underestimation of persistence by OP may suggest that constraints are less persistent thanks to self-financing.

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

The estimated wealth accumulation policy shows that there is a significant and positive effect of productivity shocks on future wealth, which suggest that the self-financing channel is operating in the data. Interestingly, we show that the effect of productivity on wealth accumulation is heterogeneous in the stock of wealth. For highly productive but constrained firms, which are in the lower 10th percentile of the distribution of wealth, the elasticity of productivity on wealth accumulation is 0.95. Thus, for very constrained firms the response in savings to a productivity shock is almost one to one. This response weakens as we move upwards along the wealth distribution. For a highly productive firm in the 90th percentile of the distribution of wealth, the elasticity of wealth accumulation to productivity is only 0.55. This result is consistent with the economic mechanisms driving the self-financing channel in models with financial frictions: Low wealth firms, which are more constrained, have higher incentives to save in order to self-finance future investments when they experience positive and persistent productivity shocks.

The estimated investment policy function shows that there is a positive and significant effect of productivity on investment for almost all of the possible combinations of state variables. Interestingly, keeping initial productivity and capital constant, the effect increases as we move towards higher percentiles of firm wealth. For instance, the marginal effect of productivity on investment is 0.19 for firms in the 10th percentile of wealth, while it is 0.25 for those in the 90th percentile. The fact that the same productivity shock generates a stronger response of investment in firms with higher level of wealth is in line with models of collateral constraints.

Literature review and outline Our paper makes contributions to two different literatures. First, it connects with the empirical literature that estimates production functions at the firm level using the proxy variable approach ([Olley and Pakes, 1996](#); [Levinsohn and Petrin, 2003](#); [Akerberg et al., 2015](#); [Doraszelski and Jaumandreu, 2013, 2018](#); [Gandhi et al., 2020](#); [Shenoy,](#)

2020). Among these papers [Olley and Pakes \(1996\)](#) and [Akerberg et al. \(2015\)](#) which studied value-added production functions using the investment equation as the proxy variable are the closest. We build on these papers to develop a framework that is robust to financial frictions. [Shenoy \(2020\)](#) is the first paper that studies how the proxy variable approach fails when any type of market frictions distort the firm’s input choices and propose to use the dynamic linear panel data approach ([Arellano and Bond \(1991\)](#), [Blundell and Bond \(1998\)](#)) to estimate the production function. Our paper differs from these papers in several aspects. First, our paper studies the biases that appear when using the proxy variable approach to estimate the production function and the productivity process under the environment of macro models with collateral constraints. Second, our paper uses the insights and economic mechanisms presented in those models to propose a novel strategy that is robust to financial frictions. In this sense our paper is the first paper that uses the self-financing channel to identify the firm productivity process and the firm production function. In terms of the methodology, we allow for more flexible policy rules including, unlike the proxy variable approach, transitory shocks. We propose a novel sequential identification scheme that leads to two novel estimators that jointly exploit the information in the investment and the wealth accumulation policy rules. Finally, an important difference from our framework, is the identification and estimation of the investment and wealth accumulation policy rules, one of the main contributions of this paper.⁷

Second, our paper connects with the macro-finance literature that studies the aggregate effects of financial frictions. We are closer to the set of papers focusing on collateral constraints and the self-financing channel (e.g. [Buera and Shin, 2011](#); [Buera et al., 2011](#); [Song et al., 2011](#); [Buera and Shin, 2013b](#); [Caggese and Cuñat, 2013](#); [Manova, 2013](#); [Moll, 2014](#); [Midrigan and Xu, 2014](#); [Khan and Thomas, 2013](#)), as we guide our empirical specification by the general implications of these models, i.e. self-financing by incumbents undo the effect of financial frictions and allows firms to invest closer to the optimal level.⁸ Our main contribution is to empirically estimate the saving and investment decisions of firms, which in these papers are an endogenous outcome of structural models calibrated with micro-data and built under different assumptions. As suggested by [Hopenhayn \(2014\)](#), this may be the source of a disparity of

⁷The papers that use the proxy variable approach use (only) the investment policy rule as an auxiliary equation to control for unobserved productivity in the production function regression. However, this policy function have not been an object of interest in those papers, and there has been no discussion on how to identify and estimate it when its relationship with productivity is not deterministic. We do this not only for investment but for the wealth accumulation policy function as well.

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

magnitudes reported for the aggregate effects of frictions. Our estimations may help to discipline these models. We provide empirical estimates of key elasticities and, unlike these papers, we exploit not only microeconomic data on real variables but on financial variables as well, for the universe of Chilean firms. Ours is the first paper that provides empirical evidence of the self-financing channel studied in this literature.⁹

A related 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 the firm. In our framework the investment policy function is one 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. Also, since we follow the structural macro models we focus on net-wealth instead of cash flows. Our results show that net-wealth is a significant determinant of investment in our sample of Chilean firms.¹⁰

The rest of the paper is organized as follows. Section 2 presents a simple model of firm dynamics with collateral constraints with the idea to provide intuition about the biases of the OP estimator in a setup with financial frictions. It also motivates the ingredients of the empirical model that we bring to the data. Section 3 introduces the empirical model and their assumptions. Section 4 establish identification of the production function, the productivity process and the policy functions. Section 5 describe the estimation methods. Section 6 presents the results and finally in section 7 we conclude.

2 A Simple Model with Financial Frictions

In this section, we start by describing a model featuring the main ingredients in the macro literature focused on firms investing under financial constraints. The objective is twofold. First, we use the model to illustrate the nature of the biases incurred when estimating the production function using standard methods in the presence of financial constraints. We obtain the sign of the biases we should obtain when applying OP to estimate the production function. Second, we use this setup to derive the form of the firm policy rules of investment and wealth

⁹There is also an extensive margin distorted by financial frictions, consisting of entry decisions to specific markets or other long-run investments. For instance [Manova \(2013\)](#) and [Caggese and Cuñat \(2013\)](#) consider entry into exporting, [Buera et al. \(2011\)](#) and [Buera and Shin \(2013b\)](#) into the manufacturing sector and [Midrigan and Xu \(2014\)](#) into a modern or formal sector. Self-financing is also influential in this extensive margin since firms or entrepreneurs might save out of profits before entering.

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

accumulation. It is important to remark that we do not directly estimate this structural model, but rather use this setup to motivate the ingredients of the empirical model that we will take to the data. Technical details on econometric issues are omitted here since they are discussed at length in the next section.

Following the macro literature featuring firms subject to collateral constraints (see [Buera et al., 2015](#), for a detailed analysis), we present here the simplest model that can generate the 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 them as collateral to obtain resources to invest in the future. The novelty is to show how this type of model relates to the literature on production function estimations.

Although we state the problem recursively we use time indexes to facilitate the mapping to the empirical model. Lower cases variables denote their values in logs. The 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} , and next period wealth A_{it+1} and capital K_{it+1} :

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

where Y_{it} is the value added produced by firm i . Capital K_{it} is decided before observing productivity Z_{it} , while labor L_{it} is decided after that.¹¹ Although not necessary in this analysis, in next sections we explicitly define investment $I_{it} = K_{it+1} + (1 - \delta)K_{it}$ as the decision variable, instead of next period capital K_{it+1} . The function $g(\cdot)$ is assumed to be convex. Since we use linear preferences this assumption rules out corner solutions.¹² The firm discounts future flows at β , capital depreciates at rate δ , and the firm pays interest rate r for its debt, implicitly defined by $K_{it} - A_{it}$.

As it 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 model.

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

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

Financial Constraints We assume firms face collateral constraints. Although our empirical specification doesn't depend on the specific nature of the constraint, we consider in this section the case where collateral defines an upper-bound for debt. A constraint like this rules out equilibrium default and can be developed from a simple limited-enforcement problem (see e.g. Buera et al., 2011). It is widely used in the macro literature due to its simplicity. An alternative that is also consistent with our framework, is to assume that collateral affects borrowing costs.¹³ In both cases OP estimates would fail, as we would have a wedge in the investment optimality condition that depends on collateral, the fact we exploit in our empirical specification.

Following Buera et al. (2015) we consider the following specification

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

Although it is commonly assumed that only net-worth influences the upper-bound on capital κ in this type of models, a more general specification in which intermediaries observe productivity (or value added), and this may increase repayment in the case of default, or it may contain information about default probabilities, would include productivity as well, in line with Aguirre (2017), Brooks and Dovis (2020) or Lian and Ma (2020) (see Buera et al., 2015, for a closer examination).¹⁴

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

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

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

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

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

¹³The 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 increasing on it (see e.g. Bernanke et al., 1999; Quadrini, 2000; Herranz et al., 2015). For applications to consumer's unsecured debt see Chatterjee et al. (2007) and Livshits et al. (2007).

¹⁴If this were not the case investment would not respond to productivity shocks in constrained firms. This invalidates the OP's monotonicity assumption (Shenoy, 2020). It is worth noticing however that one of our proposed empirical methods will be robust to this failure since it relies on the asset accumulation policy function instead of the investment policy function, and monotonicity holds in that case. However when allowing for shocks in the policy functions we need both investment and asset accumulation varying with productivity for both constrained and unconstrained firms.

in both of its arguments. Note that if we had rather assumed that collateral affects borrowing costs, that term would be the spread, and would have been a decreasing function of collateral as well.

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

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

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

If the constraint is not binding, wealth does not play a role, and there is a positive monotonic relationship between investment and productivity, 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. Equation (5) is crucial in our analysis and motivates the empirical specification of the investment policy function that we state in the next section.

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:

$$g'(A_{t+1}) = \beta(1 + r_{it})(1 + \kappa_A E_t[\mu(A_{t+1}, Z_{t+1})]) \quad (6)$$

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 profits 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 marginal product of capital expected for tomorrow, generating a positive correlation between productivity and wealth accumulation. In section 3, we exploit this positive relationship between productivity and wealth accumulation due to the self-financing channel by explicitly using the wealth accumulation policy function to learn about the firm productivity process and the firm production function.

2.1 The bias in the OP estimator under financial frictions

We use the model described above to illustrate the biases that appear when estimating the parameters of the firm production function under standard methods which do not account for financial frictions. In a very influential paper, OP 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, using data on value added y_{it} , capital k_{it} and labor l_{it} :

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \quad (7)$$

where ε_{it} is measurement error in value added.¹⁵ 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 an 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 of the unobserved productivity z_{it} . For example, in the absence of constraints, we can see from the investment policy function (5) that: $k_{it+1} = h(z_{it})$. Under the assumptions that z_{it} is the only unobserved variable for the econometrician in h (known as the scalar unobserved assumption) and the assumption that h is monotonic in z_{it} , we can invert the policy function to recover productivity as $z_{it} = h^{-1}(k_{it+1})$ and construct valid moment conditions. For instance, we can rewrite (7) as:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + h^{-1}(k_{it+1}) + \varepsilon_{it}. \quad (8)$$

Since ε_{it} is assumed to be not correlated with the inputs, OP propose to approximate $h^{-1}(k_{it+1})$ with a high-order polynomial on investment and run an OLS regression of y_{it} on l_{it} , k_{it} , and the polynomial $h^{-1}(k_{it+1})$ to estimate β_l and β_k .¹⁶ Intuitively, the OP method assigns observed differences in investment across firms in the data to differences in unobserved productivity between firms. Hence, by controlling for investment in the production function we can eliminate the endogeneity problem and get consistent estimates of β_l and β_k . However, under borrowing constraints, differences in investment between firms are not only reflecting differences in productivity but also might be driven by difference in borrowing capacity.

In the model with financial frictions described above, the investment function in (5) depends not only on productivity but also on net-worth, through its influence on the strength of financial frictions. Hence: $k_{t+1} = h(z_{it}, a_{it})$, with $h_z > 0$ and $h_a \geq 0$. When we invert the investment policy function in (5) we obtain $z_t = h^{-1}(k_{it+1}, a_{it})$, with $h_k^{-1} > 0$ and $h_a^{-1} \leq 0$. Therefore, for a given level of investment, more severe constraints due to low levels of net-worth, are associated to higher productivity levels. The intuition is direct: For a given level of productivity, an unconstrained firm will invest more than a constrained firm. Therefore for the same level of investment, it must be that the unconstrained is less productive. Replacing z_{it} in the production

¹⁵In the OP framework, ε_{it} could also be a transitory production shock that do not affect the decision of inputs at time t . We allow for that in our empirical specification.

¹⁶ In the OP framework, investment i_{it} is used as the auxiliary equation (instead of k_{it+1}), which is modeled as $i_{it} = h_t(z_{it}, k_{it})$, where h can be time-dependent and a non-linear polynomial to control for adjustment costs. The OLS regression identifies β_L , but it cannot separate β_k from the linear part of $h^{-1}(i_t, k_t)$. Thus, in a second step, OP exploits the markovian process of productivity to estimate β_k by regressing the following model: $\hat{\pi}_t(i_{it}, k_{it}) = \beta_k k_{it} + \rho \hat{\pi}_t(i_{it-1}, k_{it-1}) - \rho \beta_k k_{it-1} + \eta_{it} + \hat{\varepsilon}_{it}$, where $\hat{\pi}_t(i_{it}, k_{it})$ denotes the estimated coefficient of capital in the first step.

function we have:

$$y_t = \beta_l l_{it} + \beta_k k_{it} + h^{-1}(k_{it+1}, a_{it}) + \varepsilon_t \quad (9)$$

Hence, since implementing OP is equivalent to running a regression on y_t , with l_t , k_t and a high-order polynomial k_{t+1} as explanatory variables, the term capturing the severity of the constraint due to net-worth would go to the error term of the OP regression. Thus, if firms operate under borrowing constraints, the OP method could result in a biased estimation of β_l and β_k depending on the correlation of the regressors and the omitted variable a_{it} .¹⁷

In the case of the capital elasticity we know $h_a \geq 0$ in $k_{t+1} = h(z_{it}, a_{it})$. Hence there is a positive correlation between k_{it} and the stock of collateral at the moment the investment decision is taken a_{it-1} . Since wealth accumulation takes time, $h_a^{-1} \leq 0$, implies a negative correlation between the OP residual and k_{it} . This results in a downward bias: $\hat{\beta}_k^{OP} < \beta_k$. Intuitively, financial constraints generate differences in investment and capital even for equally productive firms. The OP framework interprets the observed differences in investment as differences in 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.

To see what happens in the case of labor we replace the expression for z_{it} we obtain after inverting (5) in the FOC for labor (3):

$$l_{it} = c_l + \frac{1}{1 - \beta_l} (\beta_k k_{it} + w + h^{-1}(k_{it+1}, a_{it})) \quad (10)$$

Therefore, after controlling for k_t and k_{t+1} , the correlation between l_{it} and the OP residual is positive.¹⁸ Because OP cannot control for a fraction of productivity, and this goes into the residual term when applying OP to (9), and because 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 with the same level of capital due to differences in collateral. OP will tend to equalize productivity between the two, despite differences in output. The productive firm, that is more financially constrained, will hire more workers, since frictions do not directly affect the labor market.¹⁹ Then OP will assign part of the output explained by productivity to the labor input, resulting in an overestimation of the labor elasticity.

A final observation we make from this analysis is that OP would underestimate (overestimate) the dispersion of productivity across firms if more productive firms are the ones that

¹⁷Note that only net-worth generates this result. If there were an upper-bound on capital that is only a function of productivity then OP would not fail, at least under the specification used in this section.

¹⁸Note that L_{it} depends only on constants c_l and w , and state variables. Then it is linearly dependent with the rest of the regressors in the production function regression (see [Akerberg et al. \(2015\)](#)). To fix this it is assumed some other determinant of labor, which in this case might be a firm-specific iid shock in wages.

¹⁹Other models consider that financial constraints can affect the labor input as well. Still, we should expect an upward bias when the effect of frictions in the labor input are less severe.

are more (less) severely constrained. This depends on the strength of productivity in relaxing constraints both directly, as an argument in κ , and indirectly, through a fastest wealth accumulation. If these effects are not enough to overcome the greater needs of capital of productive firms then, since OP underestimate productivity of constrained firms, we would expect OP to shrink the estimated productivity distribution relative to its actual value.

We conclude from this analysis that OP is likely to underestimate the capital elasticity, overestimate the labor elasticity, and underestimates the dispersion in productivity if constraints are tighter in more productive firms, in the presence of financial constraints. In the next sections, we discuss an empirical framework that is robust to the presence of financial constraints in the estimation of the firm production function. We show that considering empirically all these ingredients are crucial in the estimation.

3 General Empirical Framework

In this section, we discuss our empirical model and the assumptions needed for identifying and estimating the parameters of the production function, the productivity process, and the wealth and investment policy functions.

Empirical Production Model We consider the following Cobb-Douglas production function for valued added expressed in logs,

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it}, \quad (11)$$

where y_{it} is the log of output, k_{it} is the log of capital input, and l_{it} is the log of labor input. All of these variables are observed for the firm and for the econometrician. In contrast, z_{it} and ε_{it} are unobserved variables for the econometrician. The sum $z_{it} + \varepsilon_{it}$ captures a combination (in logs) of Hicks-neutral productivity and measurement error in value-added. Similarly to papers using the proxy variable framework, we assume that z_{it} captures the persistent component of productivity and it is known by the firm before making their input decisions in period t . In contrast, ε_{it} captures a mixture of measurement error in value added and transitory shocks to the production function that are not observable by firms when making their decisions in period t . The ε_{it} is assumed to have zero mean, to be independent over time and independent of z_{it} for all t . On the contrary, the persistent productivity z_{it} evolve following a markovian process:

$$z_{it} = \varphi(z_{it-1}) + \eta_{it}, \quad (12)$$

where η_{it} is a productivity shock that cannot be predicted by the firm with the information known by the firm up to $t - 1$. Therefore, η_{it} is not consider in the firm decisions in period $t - 1$. The assumptions about the stochastic processes of ε_{it} and η_{it} are explained in detail in

the next section. The function $\varphi(z_{it-1})$ is a non-parametric function of z_{it-1} which is known by the firm. Following the literature on production function estimation, the capital input k_{it} is modeled as a dynamic but predetermined input subject to an investment process:

$$k_{it} = \kappa(k_{it-1}, i_{it-1}), \quad (13)$$

where i_{it} denote log-investment. Note that firms produce in time t with the stock of capital accumulated up to $t-1$.²⁰ The empirical model in (11)-(13) is similar to the empirical model in Olley and Pakes (1996) and Akerberg et al. (2015).

Motivated by the simple model stated in section 2, our empirical framework departs from Olley and Pakes (1996) and Akerberg et al. (2015) in the specification of the empirical policy rules:

Investment Policy Rule The empirical policy function of investment under financial frictions is based on:

$$i_{it} = h_t(z_{it}, k_{it}, a_{it}, v_{it}), \quad (14)$$

where h_t is the empirical counterparts of an investment policy function that emerge in a firm-dynamic model with financial frictions as the one discussed in section 2 (see the FOC for investment in equation 4). There are two new ingredients with respect to the investment functions described in Olley and Pakes (1996) and Akerberg et al. (2015). First, equation (14) includes a_{it} as a state variables in order to control for collateral constraints. The nonlinear function h_t allows for heterogeneous effect of productivity on investment depending on the level of collateral and might capture the idea that the investment of constrained firms (firms with low a_{it}) responds less to productivity shocks than the investment of rich firms (firms with high a_{it}). Second, equation (14) includes an additional unobserved shock v_{it} in the policy function. This is in contrast to Olley and Pakes (1996) and Akerberg et al. (2015) which assume that the only unobservable variable in the investment equation is z_{it} (this is known as the scalar-unobservable assumption). The v_{it} is assumed to be independent across periods and independent of state variables a_{it} , k_{it} and z_{it} , and h_t is monotone in v_{it} . An economic interpretation for v_{it} is an investment cost shock like shocks to the loans interest rate. In this sense, v_{it} might capture a stochastic component of the external financing cost that is not captured by the deterministic and persistent firm-specific components in a_{it} and z_{it} .

²⁰ As emphasized by Akerberg et al. (2015) this assumption implies that it takes a full period for new capital to be ordered, delivered, and installed.

Wealth Accumulation Policy Rule A novel feature of our empirical specification is that we explicitly model the wealth accumulation policy function:

$$a_{it+1} = g_{t+1}(z_{it}, a_{it}, k_{it}, w_{it+1}). \quad (15)$$

Equation (15) is the empirical counterpart of a wealth accumulation rule that emerges in a model like the one discussed in section 2 (see equation 6). According to models with financial frictions and the self-financing channel the presence of the financial constraints generates a positive relationship between productivity and wealth accumulation for the whole distribution of firms: including both firms for which the constraint is binding and also for the ones that is not. Also, while these models predict a positive marginal effect of productivity for all firms, this effect is nonlinear and stronger for constrained firms.²¹ The empirical model in (15) is flexible enough to capture non-linear effects of productivity z_{it} on saving decisions a_{it+1} depending on the wealth accumulated up to the current period a_{it} . The empirical policy rule also includes an additional unobserved shock w_{it+1} which is assumed to be independent across periods and independent of state variables a_{it} , k_{it} , and z_{it} . The function g_{t+1} is increasing in w_{it+1} . This w_{it+1} might capture unobserved factors, other than z_{it} , that affect wealth like the interest rate shocks to assets considered in section 2. Similar to the investment policy function, we think that it is important to consider such an econometric error in the empirical specification of the wealth policy function.

We model h_t and g_{t+1} as time-specific functions to capture time-varying aggregate shocks that affects all firms. For instance, this can capture developments in the banking sector across time which would translate in an aggregate relaxation of financial constraints.

Labor Policy Rule Following, [Akerberg et al. \(2015\)](#) we model the labor input as a non-dynamic input in the sense that lagged values do not affect the labor choice:

$$l_{it} = n_t(z_{it}, a_{it}, k_{it}, w_{l,it}), \quad (16)$$

where equation (16) is the empirical labor decision. A difference from the model in section 2 is that in our empirical specification we allow for a potential effect of financial frictions over labor decisions captured by a_{it} in the policy function. The $w_{l,it}$ is assumed to be independent across periods and independent of state variables a_{it} , k_{it} and z_{it} . This $w_{l,it}$ can capture exogenous transitory shocks to wages as in the model in section 2. It can also capture optimization error as the one discussed in [Akerberg et al. \(2015\)](#).

²¹In the absence of shocks w_{it+1} to the wealth policy rule, the fact that the self-financing channel implies that high productive firms accumulate more wealth for a given level of wealth ensures that this policy rule satisfies the monotonicity assumption, necessary in a proxy variable framework. In our model with shocks, the fact that there is a relationship between wealth accumulation and productivity for all the distribution of firms will be important for identification.

4 Identification

In this section, we establish identification of the nonlinear dynamic panel model stated in the previous section. It is important to remark that identification of our model is more challenging than the firm-dynamic models studied in the proxy variable literature (Olley and Pakes (1996), Levinsohn and Petrin (2003) , Akerberg et al. (2015)) due to the presence of additional shocks in the policy functions. Therefore, it is important to show that the model we aim to estimate is identify 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) has established conditions under which dynamic nonlinear model with latent variables are non-parametrically identified under conditional independence restrictions. We built on these papers to provide nonparametric identification of the model stated in section 3.

The goal of this section is to show that $\beta_k, \beta_l, \varphi(z_{it-1}), h_t, g_{t+1}$ are identified from data on $(y_{it}, k_{it}, l_{it}, i_{it}, a_{it}, a_{it+1})$ given that $(z_{it}, w_{it+1}, v_{it}, \varepsilon_{it})$ are not observed by the econometrician and z_{it} is correlated with (l_{it}, a_{it}, k_{it}) . We make the following assumptions where we use the notation $x_i^t = (x_{i1}, \dots, x_{it})$ 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} .
- (ii) **Productivity Shock:** η_{it+s} for all $s \geq 0$ is independent over time and independent of $a_i^{t-1}, z_i^{t-1}, i_i^{t-1}, k_i^{t-1}, l_i^{t-1}$, and y_i^{t-1} .
- (iii) **Policy Functions Shocks:** v_{it} and w_{it+1} are mutually independent and also independent of $z_{i1}, (\varepsilon_{is}, \eta_{is})$ for all s and of v_{is} and w_{is+1} for all $s \neq t$.

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

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

Part (i) and (ii) of Assumption 1 state that current and future productivity and production shocks, which are independent of past productivity and production shocks, to be also independent of current and past wealth stock, capital stock, investment and labor decisions. Initial stock of wealth a_{i1} , initial stock of capital k_{i1} and initial productivity z_{i1} are arbitrarily dependent. Allowing for correlation between a_{i1}, k_{i1} and z_{i1} is important because wealth and capital accumulation upon entry in the sample may be correlated with past persistent productivity shocks. Part (iii) requires that 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 state variables (k_{it}, a_{it}, z_{it}) and mutually independent conditional on

$(l_{it}, k_{it}, a_{it}, z_{it})$. Assumption 1 provides the exclusion restrictions necessary for identification. Assumption 2 is a first order Markov condition on wealth and capital dynamics. It is satisfied in standard models of firm dynamics with financial frictions as the one described in section 2. Assumption 2 (ii) is a standard assumption in the proxy variable framework (see [Akerberg et al. \(2015\)](#)).

4.1 Intuition in a linear model

We first provide intuition for identification of a version of the model with parametric linear policy functions. Then, we generalize this ideas to establish identification in the case with non-parametric policy functions. Consider the following linear version of equations (12), (14) and (15)

$$z_{it} = \rho_z z_{it-1} + \eta_{it}, \quad (17)$$

$$i_{it} = h_z z_{it} + h_a a_{it} + h_k k_{it} + v_{it}, \quad (18)$$

$$a_{it+1} = g_z z_{it} + g_a a_{it} + g_k k_{it} + w_{it+1}, \quad (19)$$

Notice that the standard models in the proxy variable approach assume $h_a=0$ and $v_{it}=0$ and do not model explicitly equation (19).

Using equation (18) z_{it} can be written as a linear separable function of i_{it} , a_{it} , k_{it} and v_{it} .

$$z_{it} = \pi_1 i_{it} + \pi_2 a_{it} + \pi_3 k_{it} + \pi_4 v_{it} \quad (20)$$

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 (20) into the production function:

$$y_{it} = \beta_l l_{it} + (\beta_k + \pi_3) k_{it} + \pi_1 i_{it} + \pi_2 a_{it} + \tilde{\varepsilon}_{it} \quad (21)$$

where $\tilde{\varepsilon}_{it} = \varepsilon_{it} + \pi_4 v_{it}$. In the case that $v_{it} = 0$, $\tilde{\varepsilon}_{it} = \varepsilon_{it}$ and a simple OLS regression between y_{it} on l_{it} , k_{it} , i_{it} and a_{it} identifies β_l , as in the proxy variable approach. The difference with OP is that our regression controls for a_{it} . Note that β_k can not be separately identified from π_3 . As in the proxy variable approach, in a second step (once we have identified β_l), we exploit the markovian assumption of the productivity process in (17) which combined with (21) leads to the following:

$$y_{it} - \beta_l l_{it} = \beta_k k_{it} + \rho_z \pi_3 k_{it-1} + \rho_z \pi_1 i_{it-1} + \rho_z \pi_2 a_{it-1} + \varepsilon_{it} + \pi_4 v_{it-1} + \eta_{it} \quad (22)$$

Again if $v_{it-1} = 0$, an OLS regression of (22) identify β_k . The difference with OP, is that our second stage controls for a_{t-1} .

In contrast, in the more general case with investment shocks in equation (18) (i.e $v_{it} \neq 0$), investment i_{it} can be thought as proxy measure with noise v_{it} for the latent variable z_{it} , conditioned on the observed state variables a_{it} and k_{it} . Therefore, the OLS regressions of (21) and (22) do not identify β_l and β_k given that $E(i_{it}\tilde{\varepsilon}_{it}) \neq 0$ and $E(k_{it}\tilde{\varepsilon}_{it-1}) \neq 0$. Even if the investment shock v_{it} is not correlated with l_{it} , an OLS estimation of (18) will generate a bias in the estimation of β_l through the correlation of l_{it} and the latent variable z_{it} as in the classical linear multivariate model with measurement error in one regressor.

4.1.1 A simple solution: IV-proxy method

Production Function To solve the endogeneity in the proxy variable approach, we notice that the self-financing channel provides a second noisy measure of productivity in a setup with financial frictions. Hence, a_{it+1} can be used as an instrument for investment in equation (21) given the conditional independence assumption in assumption 1 and the relevance condition implied by the self-financing channel $\partial g_{t+1}/\partial z \neq 0$. The observed wealth variable a_{it+1} fulfill the exclusion restriction in equation (21) given the conditional independence assumption (wealth do not have a direct effect in the production function). It is also a relevant instrument for i_{it} through z_{it} . Note that the functions h_t and g_{t+1} are correlated conditional on a_{it} and k_{it} via z_{it} . Therefore, we construct the following IV moment restriction from (21):

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

A regression between $E[y_{it} \mid a_{it+1}, l_{it}, k_{it}, a_{it}]$, which is an object that can be compute from data and $[l_{it}, k_{it}, E[i_{it} \mid a_{it+1}, k_{it}, l_{it}, a_{it}], a_{it}]$ from (23) identify $\{\beta_l, \pi_1, \pi_2\}$ which in turns identify $\{h_z, h_a\}$. Then, β_k is identify from (22) using the following moment condition:

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

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

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

where h_{1t} and h_{2t} are nonlinear functions. The identification of β_l and β_k using the IV-proxy method strategy requires that at least one of the two policy functions is a polynomial of degree

one in z_{it} and separable in z_{it} and the policy shock. If we think that this model is a better approximation for the wealth accumulation policy rule, we should invert this policy in the first step and then use the investment equation as the instrument.

Policy Functions In the linear case, the challenge of the identification of the policy rules relies on the fact that they depend on the unobserved z_{it} . To overcome this, we exploit the markovian process of z_{it} to construct valid instruments. Once β_l and β_k are identified we can have:

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

Replacing (24) in (19):

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

An OLS regression of a_{it+1} on \tilde{y}_{it} , a_{it} and k_{it} from equation (25) do not identify the policy function since $E(\tilde{y}_{it}\varepsilon_{it}) \neq 0$. However, it is possible to use \tilde{y}_{it-1} as an instrument for \tilde{y}_{it} . The markovian assumption of productivity gives the relevance condition since ensures that $E(\tilde{y}_{it}\tilde{y}_{it-1}) \neq 0$ and assumption 1 ensures exogeneity $E(\tilde{y}_{it-1}\varepsilon_{it}) = 0$. A similar strategy identify the investment policy rule in (18).

Productivity Process For a linear productivity process, an IV argument exploiting the markovian assumption identifies the persistence and dispersion parameters. Replacing (24) in (17):

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

From equation (26), we can observe that an OLS regression between \tilde{y}_{it} and \tilde{y}_{it-1} do not identify ρ_z since \tilde{y}_{it-1} is correlated with ε_{it-1} . However, we can exploit the markovian assumption of z_{it} and *assumption 1* to use \tilde{y}_{it-2} as an instrument for \tilde{y}_{it-1} in equation (26). The following moment condition identifies ρ_z :

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

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

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

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

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 model

non-parametrically. A more general model allows for rich interaction between productivity shocks and collateral constraints. The sketch of identification is sequential. First, we establish identification of the production function parameters β_k, β_l . Then we establish identification of the productivity process and finally we show identification of the policy functions h_t and g_t .

As in the linear case discussed above, identification of the production function parameters are based on having two imperfect measures of the unobserved productivity (the investment and the wealth policy function). Once the production function parameters are identified, the productivity process is non-parametrically identified from the dynamic dependence structure of the observable variables in the production function following the ideas in the linear case. Finally, 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.

Production Function To identify the parameters of the production function we rely on results provided by [Hu and Schennach \(2008\)](#) who establish identification of nonlinear models with latent variables and multiple noisy measures of the latent variable. As in the linear case discussed above, we can think of z_{it} as a latent variable with two imperfect observable measures i_{it} and a_{it+1} . Both i_{it} and a_{it+1} are imperfect measures of z_{it} due to the presence of v_{it} and w_{it+1} . 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 $f(y_{it}, i_{it} \mid a_{it+1}, X_{it})$ in terms of some pieces of the model:

$$f(y_{it}, i_{it} \mid a_{it+1}, X_{it}) = \int f(y_{it} \mid z_{it}, k_{it}, l_{it}) f(i_{it} \mid z_{it}, X_{it}) f(z_{it} \mid a_{it+1}, X_{it}) dz_{it} \quad (29)$$

where $X_{it} = (a_{it}, k_{it}, l_{it})$ are the regressors (observable state variables) of the model, $f(y_{it} \mid z_{it}, k_{it}, l_{it})$ is the conditional distribution of the production function, $f(i_{it} \mid z_{it}, X_{it})$ is the conditional distribution of the investment policy rule and $f(z_{it} \mid a_{it+1}, X_{it})$ is the conditional distribution of the latent productivity given the state variables and the wealth accumulation (the distribution of the inverse function of the wealth accumulation rule). We notice that equation 52 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 (see appendix A.1 for the details).

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}) .

Productivity Process In contrast to the proxy variable approach, where the entire path of $z_i^T = \{z_{i1}, \dots, z_{iT}\}$ can be perfectly recovered as a deterministic function of i_{it} and state variables

from the production function setup, in our model we can not perfectly recover z_i^T from h_t and the production function due to the presence of v_{it} . However, using results in [Arellano \(2014\)](#) and [Arellano et al. \(2017\)](#) we can nonparametrically identify the joint distribution of z_i^T , which is one of the main object of interest in this paper and it is also necessary to identify the policy functions h_t and g_{t+1} which are the other important objects of this paper.

Once we identify β_k, β_l we can write the production function:

$$\tilde{y}_{it} = z_{it} + \varepsilon_{it} \quad (30)$$

$$z_{it} = \varphi(z_{it-1}) + \eta_{it} \quad (31)$$

where $\tilde{y}_{it} = y_{it} - \beta_k k_{it} - \beta_l l_{it}$. Given that z_{it} is markovian and ε_{it} is i.i.d over time, equations (30) and (31) are analogous to the income process model with non linear Markovian persistent shocks studied in [Arellano et al. \(2017\)](#). Using assumption 1 (i) and (ii) we show in appendix A.2 that the joint distribution of $(z_{i2}, \dots, z_{iT-1})$ and $(\varepsilon_{i2}, \dots, \varepsilon_{iT-1})$ are identified from the autocorrelation structure of $(\tilde{y}_{i1}, \dots, \tilde{y}_{iT})$ for a panel with $T \geq 3$. Given a stationary assumption, once we have identified $(z_{i2}, \dots, z_{iT-1})$ and $(\varepsilon_{i2}, \dots, \varepsilon_{iT-1})$, we can identify the joint distribution of $\{\varepsilon_{i1}, \varepsilon_{iT}, z_{i1}, z_{iT}\}$.

Policy Functions Let us now turn to the identification of the wealth policy rule $g_{t+1}(z_{it}, a_{it}, k_{it}, w_{it+1})$ for $t = 1, \dots, T$. We proceed in a sequential way starting with the first period. Let $y_i = (y_{i1}, \dots, y_{iT})$, $k_i = (k_{i1}, \dots, k_{iT})$ and $l_i = (l_{i1}, \dots, l_{iT})$ and use f as a generic notation for a density function. Identification of the investment policy rule $h_t(z_{it}, a_{it-1}, k_{it-1}, v_{it})$ follows the same argument.

Period 1 We allow for a flexible correlation of the initial stock of wealth in the data and the initial unobserved productivity of the firm. This is important because $t=1$ represents the first period of data reported for a firm and not the period where the firm was created. As a consequence, the stock of wealth accumulated up to period one in the sample a_{i1} might depend on past persistence productivity shocks that are summarized in z_{i1} .

$$f(a_1 | y, k, l) = \int f(a_1 | z_1, y, k, l) f(z_1 | y, k, l) dz_1, \quad (32)$$

by *assumption 1*, $f(a_1 | z_1, y, k, l) = f(a_1 | z_1)$ equation (32) can be expressed as:

$$f(a_1 | y, k, l) = \int f(a_1 | z_1) f(z_1 | y, k, l) dz_1. \quad (33)$$

Equation (33) can be rewritten as the following moment restriction:

$$f(a_1 | y, k, l) = E[f(a_1 | z_1) | y_i = y, k_i = k, l_i = l] \quad (34)$$

where the expectation is taken with respect to the density of z_{i1} given y_i, k_i, l_i and for a fixed value of a_1 . Provided that the distribution of $(z_{i1} | y_i, k_i, l_i)$, which is identified from the production function structure (see discussion above) is complete in $(y_{i1}, k_{i1}, l_{i1} \dots, y_{iT}, k_{iT}, l_{iT})$, the unknown density $f(a_1 | z_1)$ is identified from (34). The density $f(a_1, z_1 | y, k, l) = f(a_1 | z_1) f(z_1 | y, k, l)$ is also identified.

As discussed in Arellano et al. (2017), in IV terms, equation (34), in $t = 1$ for a fixed a_1 , is analogous to a nonlinear IV problem where z_{i1} is the endogenous regressor and $y_i = (y_{i1}, \dots, y_{iT})$, $k_i = (k_{i1}, \dots, k_{iT})$ and $l_i = (l_{i1}, \dots, l_{iT})$ are the vector of instruments. The difference with a standard nonlinear IV is that the "endogenous regressor" in the moment condition in (34) is a latent variable. However, this is not a problem since we have identified $(z_{i1} | y_i, k_i, l_i)$ using the production function.

Period 2 Like the analysis in period 1, we can use *assumption 1* to express $f(a_2 | a_1, y, k, l)$ as:

$$f(a_2 | a_1, y, k, l) = \int f(a_2 | z_1, a_1, k_1) f(z_1 | a_1, y, k, l) dz_1 \quad (35)$$

where $f(a_2 | a_1, y, k, l) = f(a_2 | z_1, a_1, k_1)$. Equation (35) can be rewritten in terms of the following moment restriction:

$$f(a_2 | a_1, y, k, l) = E[f(a_2 | z_1, a_1, k_1) | a_{i1} = a_1, y_i = y, k_i = k, l_i = l] \quad (36)$$

The identification of $f(a_2 | z_1, a_1, k_1)$ allow us to recover the wealth policy function $a_{i2} = g_2(z_{i1}, a_{i1}, k_{i1}, w_{i2})$ in $t = 2$ since g_2 is the conditional quantile function of a_2 given z_1, a_1 and k_1 . Equation (36) can be interpreted as a nonlinear IV restriction where a_{i1}, k_{i1} are the controls (they are arguments in the wealth function in $t=2$), and $y_i = (y_{i1}, \dots, y_{iT})$, $k_i = (k_{i2}, \dots, k_{iT})$ and $l_i = (l_{i1}, \dots, l_{iT})$ are the excluded instruments for the the endogenous unobserved variable z_{i1} in model $f(a_2 | z_1, a_1, k_1)$. Equation (36) provides identification for $f(a_2 | z_1, a_1, k_1)$ as long as $f(z_1 | a_1, y, k, l)$, which is identified in period 1, is complete in $(y_{i1}, l_{i1}, y_{i2}, l_{i2}, k_{i2} \dots, y_{iT}, k_{iT}, l_{iT})$.

Using the first order Markovian assumption (Assumption 2), the identification of the policy function from the third period and onwards follows the same steps in period 1 and period 2.

5 Empirical Strategy

In this section we discuss three approaches to estimate different versions of the empirical model presented in section 3 and discussed in section 4. First, we consider a model without shocks in the policy functions as in the proxy variable approach. For this model we propose two new

proxy variables for estimating the model by GMM . Second, we consider a model that includes shocks in the policy functions but 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 consist of an IV regression within the proxy variable framework of [Olley and Pakes \(1996\)](#) and [Akerberg et al. \(2015\)](#), following the identification strategy presented in section 4.1. Finally, we consider a more flexible model that allows for shocks in the policy functions and nonlinear effects of productivity. For this model we introduce a flexible estimation method well suited for nonlinear panel data models with latent variables.

5.1 Policy functions without shocks: proxy variable approach

Augmented OP In a model where the investment equation is a deterministic function of the state variables of the model (z_{it}, k_{it}, a_{it}) , it is possible to identify and estimate the model with financial frictions using the proxy variable approach by a simple modification of the moment conditions used by OP to control for collateral constraints. Under the assumption that the function h_t in (15) is monotonic in z_{it} , it is possible to invert the investment policy function to recover z_{it} as a function of observable variables:

$$z_{it} = \pi_t(i_{it}, k_{it}, a_{it}) \quad (37)$$

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

$$y_{it} = \beta_l l_{it} + \phi(i_{it}, k_{it}, a_{it}) + \varepsilon_{it}, \quad (38)$$

where $\phi(i_{it}, k_{it}, a_{it}) = \beta_k k_{it} + \pi_t(i_{it}, k_{it}, a_{it})$. Using *assumption 1*, we can define the following moment condition from (38):

$$E(\varepsilon_{it} \mid l_{it}, k_{it}, a_{it}, i_{it}) = 0, \quad (39)$$

$$E(\eta_{it} + \varepsilon_{it} \mid k_{it}, k_{it-1}, a_{it-1}, i_{it-1}) = 0 \quad (40)$$

The moment condition in (39) - (40) allows us to identify β_l and β_k . These two moments corresponds to the first and second stages of OP but controlling for a_{it} .

Wealth accumulation policy rule as the proxy variable Note that in the absence of shocks in the wealth accumulation policy rule we can also invert (15) and use the wealth accumulation as the proxy variable:

$$z_{it} = \pi_t(a_{it+1}, k_{it}, a_{it}), \quad (41)$$

where $\pi_t = g_t^{-1}$. Then, we can use similar moment conditions but using a_{it+1} instead of i_{it} in (39) and using a_{it} instead of i_{it-1} in (40). This approach is novel, since we are the first paper to use the self-financing channel as the proxy variable for the production function estimation. We refer to this novel estimator that use the wealth accumulation policy function to construct the proxy variable as *Proxy-Wealth*. Since z_{it} is perfectly recover, estimation of the productivity process and the policy functions are straightforward.

5.2 Policy functions with shocks

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

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

$$a_{it+1} = g(z_{it}, k_{it}, a_{it}, w_{it}) = g_1(k_{it}, a_{it}) + g_2(k_{it}, a_{it}) z_{it} + w_{it+1}, \quad (42)$$

As in the proxy variable approach we can invert equation (42):

$$z_{it} = \pi_1(k_{it}, a_{it}) + \pi_2(k_{it}, a_{it}) a_{it+1} + \omega_{it+1} \quad (43)$$

where $\pi_1(k_{it}, a_{it}) = -g_1(k_{it}, a_{it})/g_2(k_{it}, a_{it})$, $\pi_2(k_{it}, a_{it}) = 1/g_2(k_{it}, a_{it})$ and $\omega_{it+1} = -w_{it+1}/g_2(k_{it}, a_{it})$. Replacing (43) in the production function:

$$y_{it} = \beta_l l_{it} + \phi(k_{it}, a_{it}) + \pi_2(k_{it}, a_{it}) a_{it+1} + \omega_{it+1} + \varepsilon_{it}, \quad (44)$$

where $\phi(k_{it}, a_{it}) = \beta_k k_{it} + \pi_1(k_{it}, a_{it})$. As we emphasize in section 4, an OLS regression of (44) does not deliver a consistent estimator of β_l since $E(\omega_{it+1} | a_{it+1}) \neq 0$. However, given *assumption 1*, i_{it} can be use as an instrument for a_{it+1} in equation (44). Therefore, we propose the following two-stage procedure:

First Stage: Estimate (44) with an IV estimator using $\pi_2(k_{it}, a_{it}) i_{it}$ as the instrument for $\pi_2(k_{it}, a_{it}) a_{it+1}$. The IV regression delivers a consistent estimator of β_l , $\phi(k_{it}, a_{it})$ and $\pi_2(k_{it}, a_{it}) a_{it+1}$. For instance, in the linear case where $g_2(k_{it}, a_{it}) = 1$, i_{it} will be the instrument for a_{t+1} .

Second Stage: Combining equation (43) with the markovian model of the productivity process $z_{it} = \rho_z z_{it-1} + \eta_{it}$:

$$z_{it} = \rho_z \pi_1(k_{it-1}, a_{it-1}) + \rho_z \pi_2(k_{it-1}, a_{it-1}) a_{it} + \rho_z \omega_{it} + \eta_{it}, \quad (45)$$

Replacing equation (45) into the production function:

$$y_{it} - \beta_l l_{it} = \beta_k k_{it} + \rho_z \pi_1(k_{it-1}, a_{it-1}) + \rho_z \pi_2(k_{it-1}, a_{it-1}) a_{it} + \rho_z \omega_{it} + \eta_{it} + \varepsilon_{it}, \quad (46)$$

using *assumption 1* we can define the following moment condition from equation (46)

$$E(\omega_{it} + \eta_{it} + \varepsilon_{it} \mid k_{it}, k_{it-1}, a_{it-1}, i_{t-1}) = 0, \quad (47)$$

The moment condition in (47) allows us to identify β_k . If we replace β_l , $\pi_1(k_{it-1}, a_{it-1})$ and $\pi_2(k_{it-1}, a_{it-1})$ by their IV estimates from the first stage, an OLS regression of (46) 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.

Full empirical model 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 $t = 1, \dots, T$

$$\begin{cases} y_{it} = \beta_l l_{it} + \beta_k k_{it} + z_{it} + \varepsilon_{it} \\ z_{it} = \sum_{r=1}^R \alpha_r^{\varphi} \phi_r^{\varphi}(z_{it-1}) + \eta_{it} \\ i_{it} = \sum_{r=1}^R \alpha_r^h \phi_r^h(z_{it}, k_{it}, a_{it}, \delta_t^h) + v_{it} \\ a_{it+1} = \sum_{r=1}^R \alpha_r^g \phi_r^g(z_{it}, k_{it}, a_{it}, \delta_t^g) + w_{it+1} \\ a_{i1} = \sum_{r=1}^R \alpha_r^{g1} \phi_r^g(z_{i1}, \delta_1^{g1}) + w_{i1} \\ l_{it} = \sum_{r=1}^R \alpha_r^n \phi_r^n(z_{it}, k_{it}, a_{it}) + w_{l,it+1} \end{cases} \quad (48)$$

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

Stochastic EM Estimation Algorithm (SEM) To estimate our nonlinear model with latent variables, we adapt a stochastic EM algorithm to our production function framework. Let $X_i^T = (y_i^T, k_i^T, l_i^T, a_i^T,)$ and z_i^T the history of observables and productivity for firm i , respectively. Given *assumption 1*, the full model in (48) imply the following integrated moment restrictions:

$$E \left(\int \begin{bmatrix} \sum_{t=2}^T \left(a_{it+1} - \sum_{k=1}^K \alpha_k^g \phi_k^g(z_{it}, k_{it}, a_{it}, \delta_t^g) \right)^2 \\ \sum_{t=1}^T \left(i_{it} - \sum_{k=1}^K \alpha_k^h \phi_k^h(z_{it}, k_{it}, a_{it}, \delta_t^h) \right)^2 \\ \sum_{t=1}^T \left(l_{it} - \sum_{k=1}^K \alpha_k^n \phi_k^n(z_{it}, k_{it}, a_{it}) \right)^2 \\ \sum_{t=1}^T (y_{it} - \beta_l l_{it} - \beta_k k_{it} - z_{it})^2 \\ \sum_{t=1}^T \left(z_{it} - \sum_{k=1}^K \alpha_k^\varphi \phi_k^\varphi(z_{it-1}) \right)^2 \\ \left(a_{i1} - \sum_{k=1}^K \alpha_k^{g1} \phi_k^g(z_{i1}) \right)^2 \end{bmatrix} f(z_i^T | X_i^T, \theta) dz \right) \quad (49)$$

where $f(z_i^T | X_i^T, \theta)$ is the posterior density of the vector z_i^T given the data. The vector $\theta = [\theta^y, \theta^h, \theta^g, \theta^{g1}, \theta^n, \theta^\varphi]$ contains all the parameters of the model in (48), $\theta^y = [\beta_k, \beta_l, \sigma_\epsilon]$, $\theta^h = [\alpha_1^h \dots \alpha_K^h, \sigma_v]$, $\theta^g = [\alpha_1^g \dots \alpha_K^g, \sigma_w]$, $\theta^\varphi = [\alpha_1^\varphi \dots \alpha_K^\varphi, \sigma_\eta]$. Note that (49) are the integrated version of the unfeasible OLS regressions of the equations in (48). 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 (48), 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 $f(z_i^T | X_i^T, \theta)$ and simple OLS estimation of the parameters in θ .²² Arellano et al. (2017) use a similar approach in a nonlinear panel model with latent variables to estimate an income process and nonlinear consumption and assets policy rules from household data.

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

1. *Stochastic E-step:* For each firm i , draw $\{z_{i1}^{(m)} \dots z_{iT}^{(m)}\}$ M realizations of z_i^T from $f(z_i^T | X_i^T, \theta)$. Using *assumptions 1 and 2* we can express the posterior distribution of z_{it} as a

²²For instance, if we specify our nonlinear functions as second-order polynomials, the model in (48) will contains around of 100 parameters to be estimated.

function of the likelihoods of the equations in (48).

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

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

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

$$\begin{cases} \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \left(y_{it} - \beta_l l_{it} - \beta_k k_{it} - z_{it}^{(m)} \right)^2 \\ \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \left(i_{it} - \sum_{k=1}^K \alpha_k^h \phi_k^h \left(z_{it}^{(m)}, k_{it}, a_{it}, \delta_t^h \right) \right)^2 \\ \sum_{i=1}^N \sum_{t=2}^T \sum_{m=1}^M \left(a_{it+1} - \sum_{k=1}^K \alpha_k^g \phi_k^g \left(z_{it}^{(m)}, k_{it}, a_{it}, \delta_t^g \right) \right)^2 \\ \sum_{i=1}^N \sum_{t=2}^T \sum_{m=1}^M \left(z_{it}^{(m)} - \sum_{k=1}^K \alpha_k^\varphi \phi_k^\varphi \left(z_{it-1}^{(m)} \right) \right)^2 \\ \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^M \left(l_{it} - \sum_{k=1}^K \alpha_k^n \phi_k^n \left(z_{it}^{(m)}, k_{it}, a_{it} \right) \right)^2 \end{cases} \quad (50)$$

In practice, we stop the iterative procedure after $S=500$ iterations and check the convergence of the estimates. In each iteration of the chain we simulate 100 draws from step 1 (i.e $M=100$). We start the algorithm from different initial values (OP, OPA or Proxy-IV) and we get similar results. The statistical properties of a similar stochastic algorithm has been studied in (Nielsen 2000) in a likelihood context and in [Arellano and Bonhomme \(2016\)](#) in a GMM context where the M-step consist of quantile-based regressions. [Arellano and Bonhomme \(2016\)](#) show that the estimates of the stochastic EM algorithm for parametric models are asymptotically normally distributed as M and N tend to infinity 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:

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

where Σ is the asymptotic variance of the GMM estimator of (49).

6 The data

Our database comes from administrative records generated by Chile's tax collection agency (*Servicio de Impuestos Internos* - SII). The dataset covers all formal firms in the Chilean economy between 2005 and 2016 (firms' identifiers are anonymized to guarantee confidentiality). 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 for now 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 defined on annual basis.

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 information on individual workers that were employed on the firm. This allows us to have a measure of the number of workers that were employed in the firm in any given year, adjusted by the number of months. The combined dataset can be used to construct measures of value added, net investment (as the annual change in the firm's proxied capital stock), as well as estimates of firm-level total factor productivity.

Crucially, form F22 also provides information on the firms' balance sheets. In particular, we build a measure of net worth, defined as the difference between the firm's 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.

7 Empirical Results

In this section, we present the estimation results. We compare the different methods defined in section 5, which we re-enumerate here:

- (i) OP: standard approach following Olley-Pakes which uses investment as an auxiliary equation to recover productivity. This method is not robust to financial frictions.
- (ii) OPA-Inv: augments Olley-Pakes approach by including firm wealth in the auxiliary investment equation.
- (iii) Proxy-Wealth: a proxy variable approach that instead uses the wealth accumulation equation as auxiliary equation.

Methods (i), (ii), and (iii) rely on the ability of the auxiliary equations to perfectly recuperate the unobserved productivity (under a scalar unobservable assumption and a monotonicity assumption). Instead, the following methods allow for shocks to the auxiliary equations:

- (iv) Proxy-IV: uses both the investment equation and wealth accumulation equation through an IV regression.
- (v) SEM: non-linear approach that uses the full information of both investment and wealth accumulation equations.

We find significant differences in the estimates across the different approaches. These differences are in line with the predictions from the simple model described in Section 2. Overall we provide evidence of the role of firm net wealth for firm decisions, and of financial frictions in the data. Overall this means that frictions are relevant in the data and that the severity of these does depend on collateral in the form of firm's net wealth. Although an endogenous response of net wealth to frictions exists, something we confirm later in our policy function estimations, the fact that significant differences persist means that self-financing is not strong enough to ignore frictions in production function estimations.

Moreover, the estimates from OPA-IV and SEM support the evidence of the influence of iid shocks to the policy functions, and confirming that the joint estimation of the policy functions improves the results.

After presenting the evidence on the estimation of the production function. We turn to the estimation results of the policy functions. These policies are estimated in a non-linear and very flexible way, so we show how the marginal effects of productivity and current wealth on investment and wealth accumulation vary across different values of our state variables. These estimated patterns correspond with predictions of theoretical models. In particular, the estimation results of these policy rules provide evidence of the importance of the self-financing channel, and are crucial to guide the calibration of quantitative macro models with financial frictions.

7.1 Production Function

Recall from the analysis of the theoretical model in Section 2 that we expect OP to underestimate the capital elasticity, and to overestimate the labor elasticity. OP assigns differences in value added due to financial constraints to differences in productivity, and not capital, as it should. But since productivity is biased, output and labor co-move more than it would be expected, and hence a larger elasticity of labor is needed to accommodate this.

We start by describing the results of the first stage regressions of the production function estimation for methods (i) - (iv). Recall that, in this stage, the coefficient on labor delivers the estimate of β_l , while the coefficient associated to β_k is recovered in the second stage.

Table 1 displays the results of the first stage, which regress y_{it} on l_{it} , k_{it} , and the proxy for productivity. The goal of this table is to give intuition of the information added by the different proxies each method uses for capturing productivity, so it uses a linear proxy for all the specifications to make the comparison simpler. Table 2 instead shows the results with non-linear policy rules.

The OP estimate in column 2, which controls for (i_{it}, k_{it}) as proxy for productivity, delivers an estimate of β_l of 0.65. The OPA-Inv estimate in column 3, which adds a_{it} as a control, delivers a lower β_l of 0.53. We can see that the stock of wealth is significant, indicating that it plays an important role in the investment equation and should be included when constructing the proxy variable for productivity. The estimate of Proxy-Wealth in Column 4, which instead controls for $(a_{it+1}, k_{it}, a_{it})$, delivers a β_l of 0.50. We can see that a_{it+1} is significant suggesting that the wealth accumulation policy contains important information about productivity. While the estimates of β_l are close in OPA-Inv and Proxy-Wealth, they are statistically different, which suggests some misspecification in the auxiliary equations (this difference persists even when we model nonlinear policies as in table 2). Finally, column 5 displays the estimates of Proxy-IV, which exploits both policy rules and allows for shocks to these auxiliary equations. In this case, the estimate of β_l decreases to 0.43. Note that this estimate is closer to the value of the aggregate labor share in Chile of 0.44 according to the Penn World Tables.

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

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

Table 2 displays the estimation of the production function parameters (β_l, β_k) using methods (i)-(v). The estimates of β_l are different than the ones in table 1 because Table 2 uses non-linear policy functions.

There are significant differences between the estimators, with a pattern that is in line with the presence of financial constraints. Regarding the labor elasticity, the estimate of β_l is 0.65

for OP, and decreases for all the estimates that are robust to financial constraints: to 0.51 in OPA-Inv, to 0.48 in Proxy-Wealth, to 0.43 to Proxy-IV, and to 0.44 to SME. The opposite pattern is obtained for the elasticity of capital, the estimate of β_k is 0.35 for OP, and increases to 0.41 for OPA-Inv, to 0.43 for Proxy-Wealth, to 0.45 to Proxy-IV, and to 0.43 to SME.

These results are in line with the predictions of the model discussed in section 2. The presence of financial constraints leads OP to overestimate β_l and underestimate β_k . Controlling for wealth in the policy functions helps us to discriminate between productivity and collateral constraints. For instance, while two firms with the same level of investment but different levels of wealth are considered as firms with the same productivity by OP, our estimators will assign a higher productivity to the firm with lower wealth but same level of investment. In addition, by relying on the co-movements between the wealth accumulation and investment, controlling for the current stock of wealth, we can further disentangle productivity shocks from transitory shocks that affects investment and saving decisions. The differences between the estimates of OPA-Inv and Proxy-Wealth from Proxy-IV and SME confirm the presence of these transitory shocks in the policy functions.

Finally, the differences in the estimations of input elasticities in the production function translate to significant differences in the measure of returns to scale. In particular, OP results are consistent with constant returns to scale, while OPA-Inv, Proxy-Wealth, Proxy-IV and SME predict decreasing returns to scale. The estimate of Proxy-IV or SEM leads to a span of control around 0.87. Although lower than the OP estimate the value we obtained is in the upper-end of the range used in the related literature ([Buera and Shin \(2013a\)](#) use 0.79, [Restuccia and Rogerson \(2008\)](#) 0.85, and [Cageti and Di Nardi](#) 0.88).

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

TABLE 2: Production Function

7.2 Productivity

Now, we turn to the comparison of the firm productivity process estimated by the different methods.

Figure 1 depicts the distribution of productivity for OP, OPA-Inv, and SEM.²³ We can see a significant change in the dispersion of productivity. In OP, the standard deviation is 0.16, while it equals 0.31 and 0.42 under OPA-Inv and SME, respectively (see Table 3). These results are also consistent with financial frictions. More productive firms depict relatively low investment when they are constrained, and as a result productivity is underestimated under OP. The opposite is true with unconstrained less productive firms. These firms would appear to invest more, and consequently OP assigns them a higher productivity. Hence, by ignoring firm wealth, OP underestimates productivity for high productive firms and overestimates it for low-productive ones, leading to a compression of the distribution of productivity relative to the methods that are robust to frictions.

Table 3 also presents the results for the persistence of productivity. The first row displays the autocorrelation of the estimated productivity ρ_z .²⁴ We can see that the persistence is considerably lower for OP. The estimated value of ρ_z raises from 0.53 under OP to 0.7 in OPA and to 0.82 under SME, respectively. The persistence of the productivity process is a crucial parameter in quantitative models that assess the strength of the self-financing channel and the importance of financial frictions on aggregate productivity and misallocation. For instance, Moll (2014) highlights that if productivity shocks are relatively transitory, self-financing is ineffective since by the time an entrepreneur is able to save enough collateral her productivity is likely to be different. Using an elegant and tractable model of firm dynamics with financial frictions, he shows that when the persistence is low the effects of financial frictions over aggregate TFP in the steady state are large, whereas for high persistent the effects of financial frictions over aggregate TFP are low.

7.3 Policy Functions

This section presents the results of the policy functions of investment and wealth accumulation. Typically, the literature on production function estimations use the policy rules as auxiliary equations to control for unobserved productivity. But, these policy functions has not been the object of interest in this literature. We instead pay special attention to the estimated policy functions because they are key in understanding the role of financial frictions and the

²³The distribution of productivity for Proxy-Wealth is similar to the one estimated by OPA-Inv.

²⁴Let us clarify that the estimation does not assume an AR(1) process for productivity.

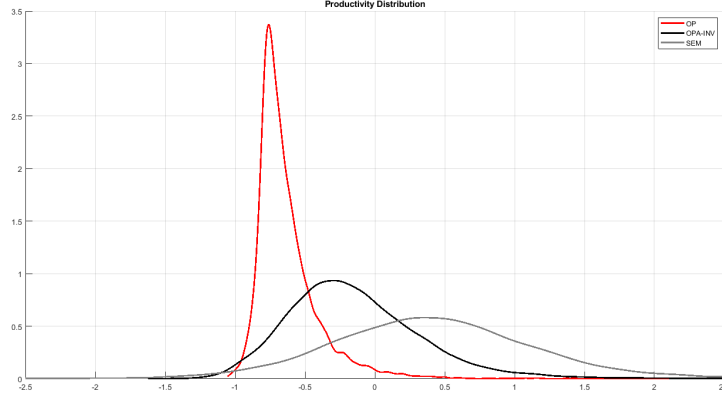


FIGURE 1: Estimated distribution of productivities

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

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

TABLE 3: Productivity Process

self-financing channel.

The following subsections show the results of the policy function using the more complete methods (iv) Proxy-IV and (v) SEM.

7.3.1 Investment Policy Function

First, we describe the estimates of the OPA-IV approach in Figure 2. This method restricts to be linear the effect of productivity on investment, still such elasticity is allowed to vary with the level of wealth and capital. Panel (a) of Figure 2 displays the estimated marginal effect of productivity on investment as a function of the level of wealth a_t , and for three different

values (the percentile 10th, 50th, and 90th) of the stock of capital k_t . Overall we find that this marginal effect is positive and significant for almost all of the possible combinations of state variables. Interestingly, the effect increases as we move through higher values of wealth, irrespectively of the stock of capital. This pattern in which a productivity shock generates a stronger response on investment for firms with higher level of wealth is in line with collateral constraints since these firms would be less constrained and can adjust their investment more easily when their productivity changes. Also, we can see that, for given wealth, the higher the initial level of capital, the smaller the effect of productivity on investment. This indicates that firms with lower wealth-to-capital ratios (high leverage) seem to be more constrained, so their investment react less to productivity.

Panel (b) displays the marginal effect of wealth on investment also for also as a function of wealth and for the three different values of the stock of capital. We can see that the elasticity is decreasing in wealth and increasing in stock of capital. This is also in line with financial constraints, firms with low wealth or (high leverage) are expected to be more constrained, so their investment reacts strongly to changes in their wealth (collateral). Instead, the level of wealth does not play a role for wealthy unconstrained firms.

Figure 3 displays the estimated marginal effect of productivity on investment using method (v) SEM. This method allows the investment policy function to be non-linear on productivity z_t . Hence, the three-dimensional graph presents how this elasticity changes for different values of a_t and z_t (we keep k_t at its median level in this graph). We can see that the previous patterns from figure 2 remain: the effect of productivity increases with a_t as firms get less constrained. In addition, this non-linear method allow us to uncover the pattern as we change z_t . Interestingly, the effect of productivity on investment increases with z_t , indicating that, for a given value of wealth, a high productive firm is able to change its investment more easily. This is in line with models of financial constraints in which productivity can affect firm lending contracts and the amount of borrowing, as it is the case in the models of Clementi and Hopenhayn (2006), DeMarzo and Fishman (2007), or Lian and Ma (2020) in which firms can use their future cash-flows as collateral.

7.4 Assets' Accumulation Policy Function

As in the last subsection, we start by describing the estimates of the OPA-IV approach in Figure 4. This method also restricts to be linear the effect of productivity on wealth accumulation, while allowing the effect to vary with the level of wealth and capital. Panel (a) of Figure 4 displays the estimated marginal effect of productivity on wealth accumulation as a function of the level of wealth a_t , and for three the percentiles 10th, 50th, and 90th of the stock of capital k_t . Overall we find that this marginal effect is positive and significant for almost all of

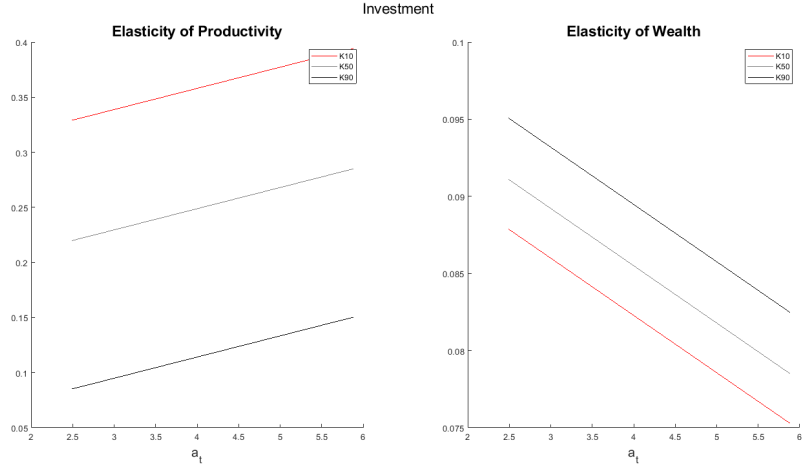


FIGURE 2: Marginal effect of productivity and wealth on investment

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

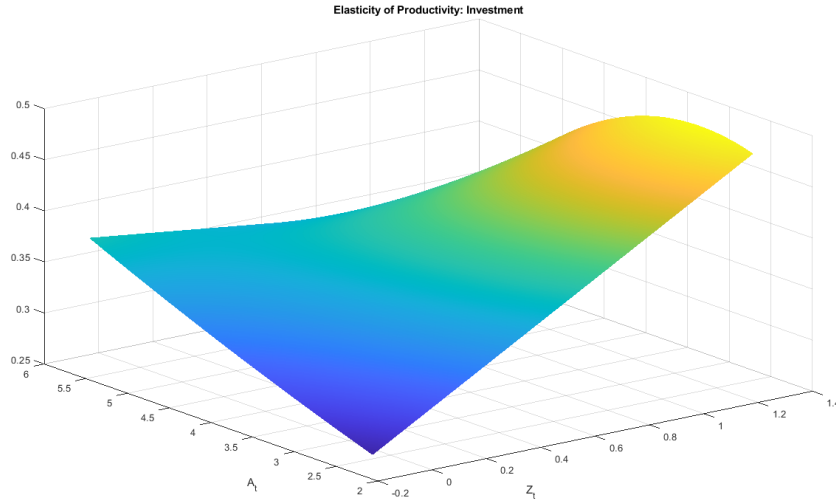


FIGURE 3: Marginal effect of productivity on investment

Notes: The figure exhibits the estimated derivative effect of productivity over investment using the SEM method. The estimated model is highly non-linear, so the figure displays the marginal effect for different values of productivity and stock of wealth, keeping the stock of capital at its median level.

the possible combinations of state variables. In contrast to the investment policy function, the effect of productivity on the wealth accumulation policy decreases as we move through higher values of wealth, irrespectively of the stock of capital. The fact that the same productivity shock generates a stronger response of wealth accumulation for firms with lower wealth is also consistent with models of financial constraints: more constrained firms have more incentives to increase savings when they face persistent productivity shocks in order to build collateral to finance future investments. Moreover, it can be observed that, for given wealth, the higher the initial level of capital (leverage), the stronger the effect of productivity on wealth accumulation. Consistent with the results on the investment equation, this indicates that firms with lower wealth-to-capital ratios (high leverage) seem to be more constrained, so their savings react more to productivity shocks.

Panel (b) displays the marginal effect of wealth a_t on the stock of wealth at the next period a_{t+1} (conditional persistence in the wealth accumulation equation). This effect is non-linear as it depends on the level of a_t . We can see that the elasticity is increasing in wealth and decreasing in stock of capital.

Figure 5 displays the estimated marginal effect of productivity on wealth accumulation using method (v) SEM. This method allows the wealth accumulation policy function to be non-linear on productivity z_t . Hence, the three-dimensional graph presents how this elasticity changes for different values of a_t and z_t (we keep k_t at its median level in this graph). We can see that the previous patterns from figure 4 remain: the effect of productivity decreases with a_t as firms get less constrained for almost all levels of productivity. Again, this non-linear method allow us to uncover the pattern as we change z_t . Interestingly, this marginal effects is larger for high productive firms with low level of wealth. This is also in line with the model described in section 2, which state that very productive firm with low level of net wealth are the ones that are more constrained. Using the SEM estimates we can see that for a high productive firm (located in the highest quintile of the productivity distribution) but with low net wealth (equal to the first quintile of the distribution on wealth), the marginal effect is almost one (0.95). This implies that very high productive firms transfer the entire increase in income (due to the persistent income shock) to savings. In contrast, a low productive firm (located in the lower quintile of the productivity distribution) but with the same level of low wealth, transfers 76 percent of the increase in income (the marginal effect is 0.76). The marginal effect of productivity to wealth accumulation reduce to 0.67 when me move to the higher quintile of the distribution of wealth for both, the high productive firm and the low productive firm.

Although we specify a non-linear policy function as in the case of investment, here we present results from a simpler linear specification. We do this mainly because, unlike the investment policy function, non-linearities are less relevant in this case. Moreover we think these simpler specifications are useful to calibrate quantitative models of collateral constraints, as we do in

the next section.

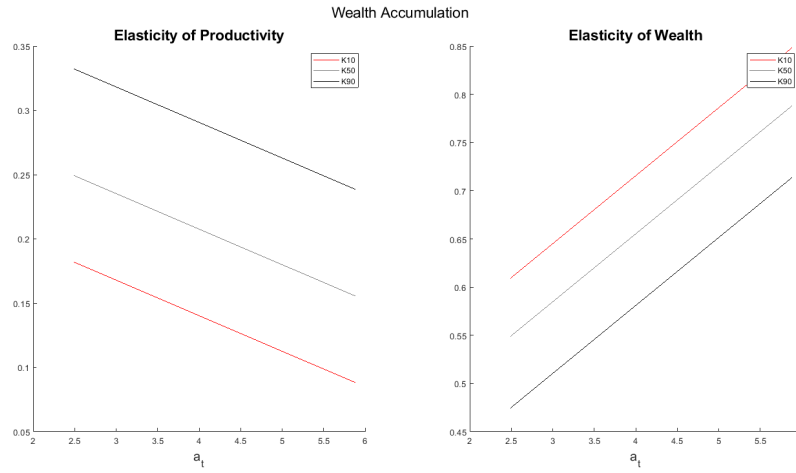


FIGURE 4: Estimated distribution of productivities

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

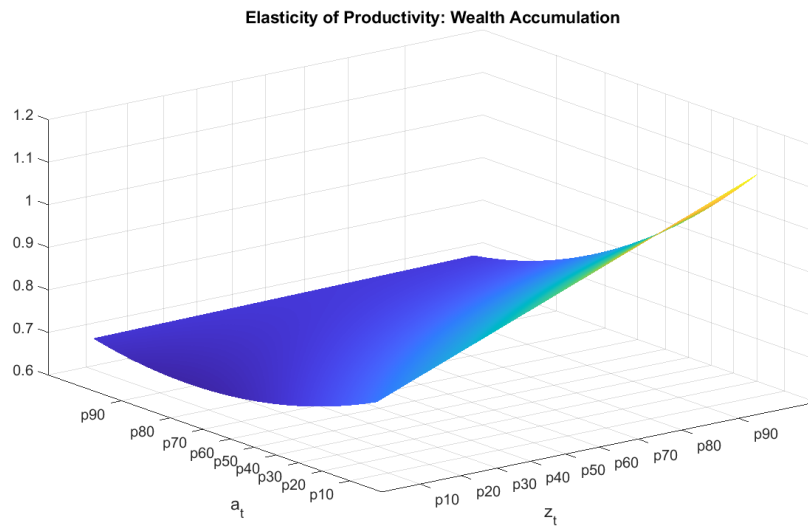


FIGURE 5: Estimated distribution of productivities

Notes: The figure exhibits the estimated derivative effect of productivity in the wealth accumulation policy function using the SEM method. The estimated model is highly non-linear, so the figure displays the marginal effect for different values of productivity and stock of wealth, keeping the stock of capital at its median level.

8 Conclusion

In this paper we study empirically the firm wealth accumulation dynamics and its relationship with the productivity process and financial frictions for the Chilean economy. To do so we need to estimate the firm productive process which is unobservable. We first notice that standard approaches to recover productivity process from production function estimations fail under the presence of financial frictions.

Under financial constraints the investment decision exploited as an auxiliary equation by OP might not hold. For instance, in a model with credit constraints like [Moll \(2014\)](#) and [Buera and Shin \(2013b\)](#), the investment decision also depends on the firm wealth or net worth for self-financing. This paper argues that not considering empirically the firm wealth when inverting the investment demand function, as in the OP approach, will render a considerable bias in the estimation of the parameters of a production function and, therefore, in the estimation of the distribution of the productivity process.

In this paper we extend the OP approach to a financial friction framework where we consider wealth and unobservable firm-specific shocks in the investment demand function. We propose a flexible framework to jointly model and estimate the firm wealth accumulation dynamics and the unobservable productivity process. We study the bias of an OP approach in a model with financial frictions and we show that this bias is quantitatively important. The marginal productivity of capital is underestimated by 30 percentage points. We show analytically that this downward bias appears because of the negative correlation between capital and the cost of financing (which is negatively correlated with the stock of wealth). Intuitively, what the OP approach does is use the observed investment as a proxy for the unobserved productivity of a firm in the production function regression. Therefore, if in the data we observe two firms with different levels of capital, investment and output, OP will assign a higher productivity to the firm with higher capital and higher investment and this will attenuate the marginal effect of capital over output because it will assign part of the higher output to the fact that this firm is more productive. In a context with financial constraints, the firm with lower capital and investment might be more productive but because it is financially constrained it can not invest at its optimal level. Therefore, higher investment in the data do not necessarily means higher productivity. So replacing the unobserved firm productivity by the observed investment level in an environment where firms face financial friction will attenuate the marginal effect of capital.

Our results show that the estimated capital elasticity in the production function increases from 0.35 when using OP to 0.43 when we estimating a model that allows for financial frictions. In contrast, the labor elasticity in the production function decreases from 0.65 in OP to 0.44 when we use our estimator that is robust to financial frictions. We also show that OP underestimates the dispersion in productivities significantly relative to our method. The 90th to 10th

productivity ratio of the firm distribution in any given year is more than twice as large under our estimator. Total Factor Productivity estimates using our framework also have a larger degree of inter temporal persistence, while the cross-section correlation between productivity and measures of firm capital and assets is more positive. We also document that the marginal effect of productivity on investment is increasing in wealth, which suggest the presence of financial frictions. Finally, we find a positive and significant marginal effect of productivity on wealth accumulation, which is stronger for more constrained firms. We take the latter as evidence that the self-financing channel is active in the data.

Appendix A.1

Using the conditional independence assumption in *assumption 1* we can write the following conditional distribution of the observed variables $f(y_{it}, i_{it} | a_{it+1}, X_{it})$ in terms of some pieces of the model, where $X_{it} = (a_{it}, k_{it}, l_{it})$ are the regressors (observable state variables) of the model.

$$f(y_{it}, i_{it} | a_{it+1}, X_{it}) = \int f(y_{it} | z_{it}, i_{it}, a_{it+1}, X_{it}) f(i_{it} | z_{it}, a_{it+1}, X_{it}) f(z_{it} | a_{it+1}, X_{it}) dz_{it}, \quad (51)$$

where $f(y_{it} | z_{it}, k_{it}, l_{it})$ 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 $f(y_{it} | z_{it}, i_{it}, a_{it+1}, X_{it}) = f(y_{it} | z_{it}, k_{it}, l_{it})$ and $f(i_{it} | z_{it}, a_{it+1}, X_{it}) = f(i_{it} | z_{it}, X_{it})$, and we can re-write (51) as

$$f(y_{it}, i_{it} | a_{it+1}, X_{it}) = \int f(y_{it} | z_{it}, k_{it}, l_{it}) f(i_{it} | z_{it}, X_{it}) f(z_{it} | a_{it+1}, X_{it}) dz_{it} \quad (52)$$

Now, the identification challenge is to recover the latent conditional density of the production function $f(y_{it} | z_{it}, k_{it}, l_{it})$ given the observed conditional density $f(y_{it}, i_{it} | a_{it+1}, X_{it})$. Once we identify $f(y_{it} | z_{it}, k_{it}, l_{it})$ we can construct $E[y_{it} | z_{it} = 0, k_{it}, l_{it}] = \beta_l l_{it} + \beta_k k_{it}$ and identify β_k, β_l with a regression between $E[y_{it} | z_{it} = 0, k_{it}, l_{it}]$ and (l_{it}, k_{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 $f(y_{it} | z_{it}, k_{it}, l_{it})$ is identified from the data. 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 above.

Definition 1: Integral Operators Let $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(Z)$ be spaces of functions defined on the domains of X and Z respectively. The integral operator based on the conditional density $f(x | z)$ is a function that maps a function $g(z)$ in $\mathcal{F}(Z)$ into a function in $\mathcal{F}(\mathcal{X})$

$$[L_{x|z}](x) = \int_{\mathbb{Z}} f(x | z) g(z) dz$$

Equation (52) can be expressed in terms of integral operators:

$$L_{y;I|a,X} = L_{y|z,k,l} D_{I;z|X} L_{z|a,X} \quad (53)$$

where $L_{y;I|a,X} = \int f(y_{it}, i_{it} | a_{it}, X_{it}) p(a_{it} | X_{it}) da$ and $D_{I;z|X}$ is a "diagonal" matrix operator mapping the function $g(z)$ to the function $f(i_{it} | z_{it}, X_{it}) g(z)$ for a given i . Integrating both sides of (52) with respect to I :

$$L_{y|a,X} = L_{y|z,k,l} L_{z|a,X} \quad (54)$$

From (54), 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 (54) in (53) we get:

$$L_{y;I|a,X} L_{y|a,X}^{-1} = L_{y|z,k,l} D_{I;z|X} L_{y|z,k,l}^{-1} \quad (55)$$

Note that the observed quantity $L_{y;I|a,X} L_{y|a,X}^{-1}$ in (55) 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 (55). If $L_{y|z,k,l}$ is identify, then $f(y_{it} | z_{it}, i_{it}, a_{it}, X_{it})$ is identify.

Rank Condition (Injectivity) To identify $L_{y|z,k,l}$ from (55), the inverse of $L_{y|a,X}$ has to exist. Looking at (54) we can show that $L_{y|a,X}$ has an inverse if $L_{y|z,k,l}$ and $L_{z|a,X}$ are invertible. This is the case under the assumption that the conditional characteristic functions of $f(y | z, k, l)$ and $f(a_{it+1} | z_{it}, a_{it}, k_{it})$ do not vanish on the real line. The operators $L_{y|z,k,l}$ and $L_{a|z,X}$ are injective (and invertible) if there is sufficient variation in the densities $f(y | z, k, l)$ and $f(a_{it+1} | z_{it}, a_{it}, k_{it})$ for different values of z_{it} . The condition for $f(y | z, k, l)$ is directly fulfill by the Cobb Douglas production function with Hicks neutral productivity. A monotonic relation between a_{it+1} and z_{it} in equation (15) fulfill the condition for $f(a_{it+1} | z_{it}, a_{it}, k_{it})$. In the IV terminology, the later is a relevance condition, that ensures that a_{it+1} is valid instrument for z_{it} .

Note that the expression $L_{y;I|a,X} L_{y|a,X}^{-1}$ in (55) 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} .

Appendix A.2

Following [Arellano et al. \(2017\)](#), we first show nonparametric identification of the distribution of ε_{it} for all t and 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:

$$\tilde{y}_{it} = \psi(\tilde{y}_{it-1}) + \zeta_{it} \quad (56)$$

$$\tilde{y}_{it-1}\tilde{y}_{it} = \phi(\tilde{y}_{it-1}) + v_{it} \quad (57)$$

where $E[\zeta_{it} | \tilde{y}_{it-2}] = 0$ and $E[v_{it} | \tilde{y}_{it-2}] = 0$, and $\psi(\cdot)$ and $\phi(\cdot)$ are the solutions of an IV regression where \tilde{y}_{it-2} is the instrument of \tilde{y}_{it-1} in (56) and (57): $E[\tilde{y}_{it} - \psi(\tilde{y}_{it-1}) | \tilde{y}_{it-2}] = 0$ and $E[\tilde{y}_{it-1}\tilde{y}_{it} - \phi(\tilde{y}_{it-1}) | \tilde{y}_{it-2}] = 0$. The solutions $\psi(\cdot)$ and $\phi(\cdot)$ exist and are unique if both the conditional distributions of $\tilde{y}_{it} | \tilde{y}_{it-1}$ and $\tilde{y}_{it-1} | \tilde{y}_{it}$ are complete. This is a nonlinear relevance assumption that is ensured by the markovian condition of z_{it} . Identification of $\psi(\cdot)$ and $\phi(\cdot)$ relies on the autocorrelation structure in the data $(\tilde{y}_{i1}, \dots, \tilde{y}_{iT})$. Note that both $\psi(\cdot)$ and $\phi(\cdot)$ 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} . Hence:

$$E(\tilde{y}_{it} | z_{it-1}) = E(\psi(\tilde{y}_{it-1}) | z_{it-1}), \quad (58)$$

$$z_{it-1}E(\tilde{y}_{it} | z_{it-1}) = E(\phi(\tilde{y}_{it-1}) | z_{it-1}). \quad (59)$$

Equation (58) uses the condition that $E(\psi(\tilde{y}_{it-1}) | z_{it-1}, \tilde{y}_{it-2}) = E(\psi(\tilde{y}_{it-1}) | z_{it-1})$ and $E(\tilde{y}_{it} | z_{it-1}, \tilde{y}_{it-2}) = E(\tilde{y}_{it} | z_{it-1})$, while equation (59) uses also the condition that $E(\varepsilon_{it-1} | z_{it-1}) = 0$ and $E(\phi(\tilde{y}_{it-1}) | z_{it-1}, \tilde{y}_{it-2}) = E(\phi(\tilde{y}_{it-1}) | z_{it-1})$.

Since $\psi(\cdot)$ and $\phi(\cdot)$ are identified from (56) and (57) and data on $\{\tilde{y}_{it-2}, \tilde{y}_{it-1}, \tilde{y}_{it}\}$, we can use equation (58) and (59) to identify the distribution of ε_{it-1} for a fixed value of z :

$$E_{\varepsilon_{it-1}}[z\psi(z + \varepsilon_{it-1})] = E_{\varepsilon_{it-1}}[\phi(z + \varepsilon_{it-1})] \quad (60)$$

By a deconvolution we can recover the density of ε_{it-1} from (60). 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 $(\varepsilon_{i2}, \dots, \varepsilon_{iT-1})$, we can identify the distribution of $(z_{i2}, \dots, z_{iT-1})$ given the observed data on $(\tilde{y}_{i2}, \dots, \tilde{y}_{iT-1})$.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification properties of recent production function estimators," *Econometrica*, 83(6), 2411–2451.

- AGUIRRE, A. (2017): “Contracting institutions and economic growth,” *Review of Economic Dynamics*, 24, 192–217.
- ARELLANO, M. (2014): “Uncertainty, persistence, and heterogeneity: A panel data perspective,” *Journal of the European Economic Association*, 12(5), 1127–1153.
- ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2017): “Earnings and consumption dynamics: a nonlinear panel data framework,” *Econometrica*, 85(3), 693–734.
- ARELLANO, M., AND S. BOND (1991): “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.
- ARELLANO, M., AND S. BONHOMME (2016): “Nonlinear panel data estimation via quantile regressions,” .
- BERNANKE, B. S., M. GERTLER, AND S. GILCHRIST (1999): “The financial accelerator in a quantitative business cycle framework,” *Handbook of macroeconomics*, 1, 1341–1393.
- BLUNDELL, R., AND S. BOND (1998): “Initial conditions and moment restrictions in dynamic panel data models,” *Journal of econometrics*, 87(1), 115–143.
- BROOKS, W., AND A. DOVIS (2020): “Credit market frictions and trade liberalizations,” *Journal of Monetary Economics*, 111, 32–47.
- BUERA, F. J., J. P. KABOSKI, AND Y. SHIN (2011): “Finance and development: A tale of two sectors,” *The American Economic Review*, 101(5), 1964–2002.
- (2015): “Entrepreneurship and Financial Frictions: A Macro-Development Perspective,” *Annual Review of Economics*.
- BUERA, F. J., AND Y. SHIN (2011): “Self-insurance vs. self-financing: A welfare analysis of the persistence of shocks,” *Journal of Economic Theory*, 146(3), 845–862.
- (2013a): “Financial frictions and the persistence of history: A quantitative exploration,” *Journal of Political Economy*, 121(2), 221–272.
- (2013b): “Financial Frictions And The Persistence Of History: A Quantitative Exploration,” *Journal of Political Economy*, 121(2), 221–272.
- CAGGESE, A., AND V. CUÑAT (2013): “Financing constraints, firm dynamics, export decisions, and aggregate productivity,” *Review of Economic Dynamics*, 16(1), 177–193.

- CHATTERJEE, S., D. CORBAE, M. NAKAJIMA, AND J.-V. RÍOS-RULL (2007): “A quantitative theory of unsecured consumer credit with risk of default,” *Econometrica*, 75(6), 1525–1589.
- CLEMENTI, G. L., AND H. A. HOPENHAYN (2006): “A theory of financing constraints and firm dynamics,” *The Quarterly Journal of Economics*, 121(1), 229–265.
- DEMARZO, P. M., AND M. J. FISHMAN (2007): “Optimal long-term financial contracting,” *The Review of Financial Studies*, 20(6), 2079–2128.
- DORASZELSKI, U., AND J. JAUMANDREU (2013): “R&D and productivity: Estimating endogenous productivity,” *The Review of Economic Studies*, 80(4), 1338–1383.
- (2018): “Measuring the bias of technological change,” *Journal of Political Economy*, 126(3), 1027–1084.
- FAZZARI, S., R. G. HUBBARD, AND B. C. PETERSEN (1987): “Financing constraints and corporate investment,” Discussion paper, National Bureau of Economic Research.
- GANDHI, A., S. NAVARRO, AND D. A. RIVERS (2020): “On the identification of gross output production functions,” *Journal of Political Economy*, 128(8), 2973–3016.
- HERRANZ, N., S. KRASA, AND A. P. VILLAMIL (2015): “Entrepreneurs, risk aversion, and dynamic firms,” *Journal of Political Economy*, 123(5), 1133–1176.
- HOPENHAYN, H. A. (2014): “Firms, misallocation, and aggregate productivity: A review,” *Annu. Rev. Econ.*, 6(1), 735–770.
- HU, Y., G. HUANG, AND Y. SASAKI (2020): “Estimating production functions with robustness against errors in the proxy variables,” *Journal of Econometrics*, 215(2), 375–398.
- HU, Y., AND S. M. SCHENNACH (2008): “Instrumental variable treatment of nonclassical measurement error models,” *Econometrica*, 76(1), 195–216.
- HU, Y., AND M. SHUM (2012): “Nonparametric identification of dynamic models with unobserved state variables,” *Journal of Econometrics*, 171(1), 32–44.
- KHAN, A., AND J. K. THOMAS (2013): “Credit shocks and aggregate fluctuations in an economy with production heterogeneity,” *Journal of Political Economy*, 121(6), 1055–1107.
- LEVINSOHN, J., AND A. PETRIN (2003): “Estimating production functions using inputs to control for unobservables,” *The Review of Economic Studies*, 70(2), 317–341.

- LIAN, C., AND Y. MA (2020): “Anatomy of corporate borrowing constraints,” *The Quarterly Journal of Economics*, 136(1), 229–291.
- LIVSHITS, I., J. MACGEE, AND M. TERTILT (2007): “Consumer bankruptcy: A fresh start,” *American Economic Review*, 97(1), 402–418.
- MANOVA, K. (2013): “Credit constraints, heterogeneous firms, and international trade,” *Review of Economic Studies*, 80(2), 711–744.
- MIDRIGAN, V., AND D. Y. XU (2014): “Finance and misallocation: Evidence from plant-level data,” *The American Economic Review*, 104(2), 422–458.
- MOLL, B. (2014): “Productivity losses from financial frictions: Can self-financing undo capital misallocation?,” *American Economic Review*, 104(10), 3186–3221.
- NAKAMURA, E., AND J. STEINSSON (2018): “Identification in macroeconomics,” *Journal of Economic Perspectives*, 32(3), 59–86.
- OLLEY, S., AND A. PAKES (1996): “The dynamics of productivity in the telecommunications equipment industry,” *Econometrica*, 64, 1263–97.
- QUADRINI, V. (2000): “Entrepreneurship, saving, and social mobility,” *Review of economic dynamics*, 3(1), 1–40.
- RESTUCCIA, D., AND R. ROGERSON (2008): “Policy distortions and aggregate productivity with heterogeneous establishments,” *Review of Economic dynamics*, 11(4), 707–720.
- SHENOY, A. (2020): “Estimating the Production Function Under Input Market Frictions,” *Review of Economics and Statistics*, pp. 1–45.
- SONG, Z., K. STORESLETTEN, AND F. ZILIBOTTI (2011): “Growing like china,” *American economic review*, 101(1), 196–233.

Documentos de Trabajo Banco Central de Chile	Working Papers Central Bank of Chile
NÚMEROS ANTERIORES	PAST ISSUES
<p>La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica:</p> <p>www.bcentral.cl/esp/estpub/estudios/dtbc.</p> <p>Existe la posibilidad de solicitar una copia impresa con un costo de Ch\$500 si es dentro de Chile y US\$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.</p>	<p>Working Papers in PDF format can be downloaded free of charge from:</p> <p>www.bcentral.cl/eng/stdpub/studies/workingpaper.</p> <p>Printed versions can be ordered individually for US\$12 per copy (for order inside Chile the charge is Ch\$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.</p>

DTBC – 903

Earnings Cyclicalities of New and Continuing Jobs: The Role of Tenure and Transition Length

Elías Albagli, Gabriela Contreras, Matías Tapia, Juan M. Wlasiuk

DTBC – 902

The Internal Labor Markets of Business Groups

Cristobal Huneeus, Federico Huneeus, Borja Larrain, Mauricio Larrain, Mounu Prem

DTBC – 901

A strategic analysis of “Expectations and the neutrality of money”

Gent Bajraj, Neil Wallace

DTBC – 900

Forecasting Brazilian Inflation with the Hybrid New Keynesian Phillips Curve: Assessing the Predictive Role of Trading Partners

Carlos Medel

DTBC – 899

Searching for the Best Inflation Forecasters within a Consumer Perceptions Survey: Microdata Evidence from Chile

Carlos Medel

DTBC – 898

Capital Flows and Emerging Markets Fluctuations

Jorge Lorca

DTBC – 897

Financial Constraints: a Propagation Mechanism of Foreign Shocks

Rosario Aldunate

DTBC – 896

Sovereign Default Risk, Macroeconomic Fluctuations and Monetary-Fiscal Stabilization

Markus Kirchner, Malte Rieth

DTBC – 895

Productivity Gaps and Job Flows: Evidence from Censal Microdata

Elías Albagli, Mario Canales, Chad Syverson, Matías Tapia, Juan Wlasiuk

DTBC – 894

The potential impact of financial portability measures on mortgage refinancing: Evidence from Chile

Carlos Madeira

DTBC – 893

Macroprudential Policy and the Inward Transmission of Monetary Policy: the case of Chile, Mexico, and Russia

Georgia Bush, Tomás Gómez, Alejandro Jara, David Moreno, Konstantin Styrin, Yulia Ushakova

DTBC – 892

Labor Earnings Dispersion in Chile: Decomposition, Dynamics and the Role of Firms

Rosario Aldunate, Gabriela Contreras, Matías Tapia

DTBC – 891

Determinantes del rating soberano: el caso de Chile

Natalia Gallardo, Mauricio Hitschfeld

DTBC – 890

The impact of the Covid public policies on the Chilean households

Carlos Madeira



BANCO CENTRAL
DE CHILE