Non-Linear Productivity and Investment Dynamics

Giulio Fella, Julio Galvez, Beatriz González,
Juan Carlos Ruiz-Garcia and Tatsuro Senga*

November 23, 2021

Preliminary and incomplete. Please do not cite without permission.

Abstract

The nature of productivity shocks is a central feature of models of heterogeneous-firm dynamics. Using firm-level data from Spain, we show that the observed productivity dynamics differ from those implied by the canonical, linear AR(1) representation with normally-distributed shocks. We document that the productivity process features non-linear persistence and non-normality. Motivated by this, we estimate a flexible stochastic process for productivity which allows for these features and compare its implications with the canonical one. We find that the flexible model fits the productivity data much better. We also estimate non-parametric, semi-reduced form empirical investment functions and find that the two processes imply very different responses to productivity shocks. Those estimated under the flexible process fit much better the relationship between investment and productivity in the data. Finally, we embed both the non-linear and the canonical process in a structural, partial equilibrium investment model and estimate capital adjustment costs under both specifications. The model estimated under the non-linear process fits the data much better and has very different implications for the relative importance of fixed versus quadratic costs of adjustment.

Keywords: Firm dynamics, non-linear productivity process, capital adjustment costs, latent variables, quantile regression.
JEL Codes: E22, E23

*Giulio Fella: Queen Mary University of London, CFM, CEPR and IFS. Julio Galvez: Banco de España. Beatriz González: Banco de España. Juan Carlos Ruiz-Garcia: Universitat de les Illes Balears. Tatsuro Senga: Queen Mary University of London. We thank Manuel Arellano for his useful comments. The views expressed here are those of the authors and are not necessarily the views of CFM, CEPR, IFS, the Banco de España or the Eurosystem.

1 Introduction

It is widely documented that there are large and persistent differences in productivity levels across firms. This empirical finding has led researchers to develop models of firm dynamics with heterogeneity in productivity at the micro-level.[1] In these models, firms make optimal dynamic decisions, such as investment and hiring, based on their expectations of future productivity. Therefore, the conditional distribution of idiosyncratic productivity shocks is key to understand the dynamics of these decisions.

An important limitation of this literature is the assumption of normality and linearity of the productivity dynamics. For example, the workhorse models of heterogeneous firms assume that firm-level productivity follows an AR(1) process with normal innovations. This limitation may lead to understate the salient feature of firm heterogeneity underlying the economy and can matter both qualitatively and quantitatively in drawing any conclusions drawn from these models.

In this paper, we break from this tradition and make the following contributions.

First, we use a comprehensive firm-level dataset for the Spanish economy to extract firm-level productivity through the estimation of a production function at the sectoral level. We show that the resulting productivity series displays substantial non-linearity and non-normality. Motivated by this finding, we use a recent econometric method proposed by Arellano et al. (2017) to estimate a flexible stochastic process for productivity that does not impose either linearity or normality. We show that this flexible process fits the productivity series much better than the alternative one that imposes such restrictions. In order to allow for measurement error and/or different degrees of shock persistence, both processes decompose productivity into two latent components: a persistent (first-order Markov) component and a transitory one. In the case of the restricted process, the two components take the form respectively of an AR(1) and white noise both with normal innovations.

Given the two processes, we estimate by simulation semi-reduced-form empirical

[1] Seminal contributions are Hopenhayn (1992), Cooper and Haltiwanger (2006), Khan and Thomas (2008), and Bloom (2009), among others.

investment functions relating the investment rate to last period capital and the two latent productivity components under either process. As in Arellano et al. (2017) who estimate an empirical consumption rule, our empirical investment rule is also non-linear and non-parametric.[2] We find that the estimated response of investment to shocks is substantially different between the two processes. In particular, the average derivative of investment with respect to productivity shocks when the underlying process is nonlinear (0.00 to 0.80) matches the data (0.02 to 0.70) well, as opposed to when the underlying process is the canonical one (-0.10 to 1.05). We show that this difference is because of the different responses to persistent and transitory productivity. In the model under the nonlinear productivity process, investment responses are mainly driven by the persistent productivity shock; by contrast, in the model under the canonical productivity process, investment responses are mainly driven by the transitory shock.

In the final part of the paper, we estimate a structural model of investment with adjustment costs à la Cooper and Haltiwanger (2006). As in Cooper and Haltiwanger (2006) we allow for both quadratic and fixed adjustment costs, as well as irreversibility. We structurally estimate the adjustment costs and irreversibility parameters under the two processes. The model with the the non-linear productivity process fits the targeted moments way better than the alternative. For instance, both models can match the targeted rate of positive investment spikes and the standard deviation of investment rates, while the two models yield different serial correlation of investment rates. The data moments is 0.150 and the non-linear process yields 0.150 precisely, while the linear process yields 0.351. This stark discrepancy has important implication for aggregate investment dynamics as the sensitivity of investment critically hinges on the price sensitivity, which is tightly linked to how investment is serially correlated at the micro level.[3] Furthermore, the two productivity processes entail important quantitative differences for the estimates of adjustment costs with the canonical case implying a much larger role for fixed as opposed

[2]Our approach is well-suited because investment dynamics are highly non-linear with some notable features like positive and negative spikes, inaction, and a weak serial correlation.

[3]A prime example is the analysis of the monetary policy transmission mechanism. It is commonly understood that we need to dampen the interest rate sensitivity of investment for the canonical New Keynesian model to be able to reproduce an empirically consistent aggregate dynamics following monetary policy shocks.

to convex cost. Intuitively, the model with a linear productivity dynamics has to rely on large fixed costs to match the investment non-linearities in the data. Conversely, in the alternative model it is the productivity dynamics that partially drives the non-linearities in investment.

The remainder of the paper is organized as follows. Section 2 discusses our dataset and the details of the estimation of the production function and the resulting productivity series. Section 3 discusses properties of the productivity series and motivates the need for a flexible, non-linear stochastic process. Section 4 estimates both a flexible, non-linear process and a canonical, linear one and compares their implications. We estimate empirical investment rules in Section 5 while Section 6 estimates a structural investment model under the two alternative processes and compares their implications particularly for the estimates of adjustment costs. We conclude in Section 7.

2 Data and productivity estimates

2.1 Data

We use comprehensive firm-level data for the Spanish economy. The main dataset called *Central de Balances Integrada* (CBI) is hosted by *Banco de España* (BdE). The CBI comprises the quasi-universe of Spanish firms, accurately representing the economy's business structure.[4] The CBI includes very detailed information concerning all the categories of the balance sheet, income statement, and the annual report for each firm since 1995. In the paper, we use data covering all economic sectors from 1999 to 2017, which results, after cleaning, in more than 1.1 million different firms and 7.5 million firm-year observations.[5]

We construct the main variables using the detailed information in the CBI. The difference between the year of the annual accounts and the year of the company's foundation identifies the firm's age. The sector is defined at the 2-digits level according to the

[4]Firms have the legal requirement to submit their annual accounts at the Commercial Registry. BdE compiles and homogenises all the information at the Commercial Registry in a unique dataset, the CBI.

[5]The final sample starts at 1999 instead of 1995 as the lack of capacity and coordination among the commercial registries results in low coverage for the first years of the agreement.

Spanish National Accounts classification (CNAE). The difference between value of production and intermediate goods expenditure delivers the value-added variable, which we deflate using sector-specific value-added deflators from the Spanish National Accounts. We define capital as the sum of intangible and productive assets and deflate it using sector-specific capital deflators.[6] Following the literature, we define labor as the wage-bill in an attempt to control for worker's heterogeneity across firms.[7] We deflate the labor variable using the value-added deflators. Finally, the materials variable measured as intermediate goods is deflated using the value-added deflators. All the deflators are based on 2006.

We focus our analysis on limited liability firms that are either non-lucrative corporations or not controlled by the public sector. There are many firms without or with small economic activity which we remove from our sample.[8] We clean the restricted sample to avoid firms that have reported incorrect figures, variables with erroneous units, and unrealistic values in any of the variables used in the paper. The details of the cleaning procedure and a statistic summary of the variables are reported in Appendix A. We compare the representativeness of the final sample against *el Directorio Central de Empresas* (DIRCE).[9] The final sample has the excellent coverage of 2/3 of all Spanish firms and total employment. Notably, the coverage is stable over time, across main sectors, i.e. manufacturing and services, and more importantly across narrowly defined categories of the firm size distribution, a rare characteristic among the existing firm-level datasets that tend to over-sample or restrict their attention to large corporations.

[6] The literature usually uses the long-term assets to define capital. Our more detailed data allows us to remove the long-term assets that the firm has as unproductive investments, i.e. bonds, real estate, stocks, which do not take part in the production process.

[7] We cannot use the number of workers in equivalent-time units, as we do not observe their characteristics.

[8] Firms with zero employment or minimal economic activity are particularly likely to be used as instrumental corporations in order to avoid taxes or hide heritage to the fiscal authorities. We define firms with no economic activity as those with either their value-added is smaller than 1,000€, use less than 100€ in materials, have less than 500€ as capital stock, employ less than 0.25 people in full-time equivalence, or pay a total wage-bill smaller than 3,000€, in 2006 real €.

[9] DIRCE provides aggregate information on the census of Spanish firms by type of corporation, i.e. limited vs non-limited, economic sector, and firm size measured as the number of employees. Unfortunately, it does not provide information by firm's age.

2.2 Productivity

Total factor productivity (TFP) measures the overall effectiveness with which capital and labor are used in a production process. The usual approach to extract TFP is through the estimation of a structural production function (see, e.g., Olley and Pakes (1996)). Following previous work, we assume that the production function has a Cobb-Douglas form, i.e.,

$$Y_{jht} = \exp(\omega_{jht}) K_{jht}^{\beta_{k,h}} L_{jht}^{\beta_{l,h}}, \tag{1}$$

where j is the firm, h the sector in which it operates, and t the year. By taking logs of equation (1), we obtain the following expression that we estimate from the data:

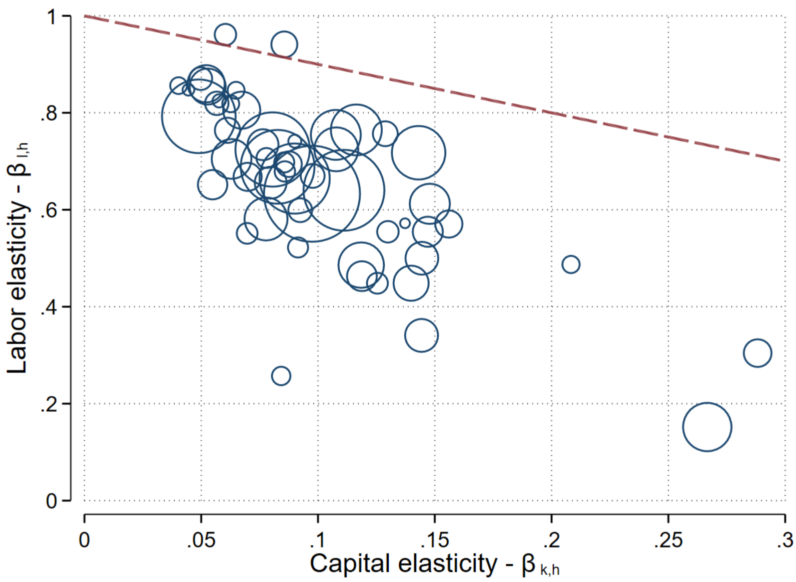
$$y_{jht} = \beta_{h,t} + \beta_{k,h} k_{jht} + \beta_{l,h} l_{jht} + z_{jht} + \varepsilon_{jht}, \tag{2}$$

where we define y_{jht} to be the log value-added —value of production minus intermediate inputs—, k_{jht} the log of physical capital —operating long-term assets—, and l_{jht} the log of labor —wage bill—. The aggregate evolution of TFP in sector h is captured in $\beta_{h,t}$, while firm-idiosyncratic TFP is given by $\omega_{jht} = z_{jht} + \varepsilon_{jht}$, where z_{jht} is the firm's productivity and ε_{jht} is an error term (i.e., a transitory shock or measurement error) that neither the firm nor the econometrician knows.

To estimate the parameters of the production function (2), we modify the Ackerberg et al. (2015) semi-parametric procedure in two respects. First, we utilize a robust estimation routine and systematically employ the continuously updated generalized method of moments (CU-GMM) advocated by Kim et al. (2019). Second, to consider the effect of aggregate shocks in estimating the production function parameters, we apply the econometric framework in Hahn et al. (2020). In this way, we ensure that the extracted TFP measures are free of the effect of aggregate TFP at the sector level in a given year. We provide more details concerning the estimation procedure, together with the assumptions we make, in Appendix B.

Figure 1 shows the production function point estimates of each sector with the dot surface being the size of the sector in the economy in terms of value-added. The capital value-added elasticity —$\beta_{k,h}$— mostly ranges in the 0.05 to 0.15 interval, slightly smaller values than usually found in the literature, while the labor value-added elasticity —$\beta_{l,h}$—, which ranges from 0.4 to 0.8, is in line with previous findings. The dotted red line limits the region —below the line— displaying a decreasing returns to scale technology. The estimates point to a decreasing return to scale technology in most sectors, 52 out of 54. Only two small sectors in the economy —pharmaceutical products and other transportation materials— have a slightly increasing returns to scale technology. The returns to scale —$\beta_{k,h} + \beta_{l,h}$— are mostly estimated around the 0.6 to 0.9 interval, in line with previous studies.

Figure 1: Point estimates of the production function by sector



Note: The figure shows the point estimates of the production function in equation (1). Each dot represents one sector at the 2-digits CNAE classification. The dot's surface represents the weight of the sector in the economy in terms of value-added. The dashed line divides the 2-dimension space according to the returns to scale. Data from the Spanish Central de Balances, from 1999 to 2017.

We report the aggregation of the sector by sector production function estimation in Table 1. The first two rows are the simple average across sectors and the bootstrapped 90% confidence interval, respectively. In the following rows, we weigh each sector according to its value-added contribution to the economy. The parameters of the aggregate production function are precisely estimated with the returns to scale parameter being statistically smaller than one, i.e. decreasing returns to scale. Note that the aggregate estimates are more precise in the weighted specification. The reason is that sectors that contribute more to the economy usually have more firms, being the production function parameters more accurately estimated in those sectors.

Table 1: Summary of the production function estimates

Aggregation	Capital (90% CI)	Labor (90% CI)	RtS (90% CI)
Unweighted	0.101	0.658	0.759
	(0.073, 0.131)	(0.508, 0.802)	(0.627, 0.884)
Weighted	0.101	0.658	0.759
	(0.085, 0.124)	(0.565, 0.744)	(0.676, 0.834)

Note: We report the aggregated parameter estimates of the productivity function. We use value-added weights computed from Spanish National Accounts. 90% confidence intervals are computed via 200 bootstrap samples. Data from the Spanish Central de Balances, from 1999 to 2017.

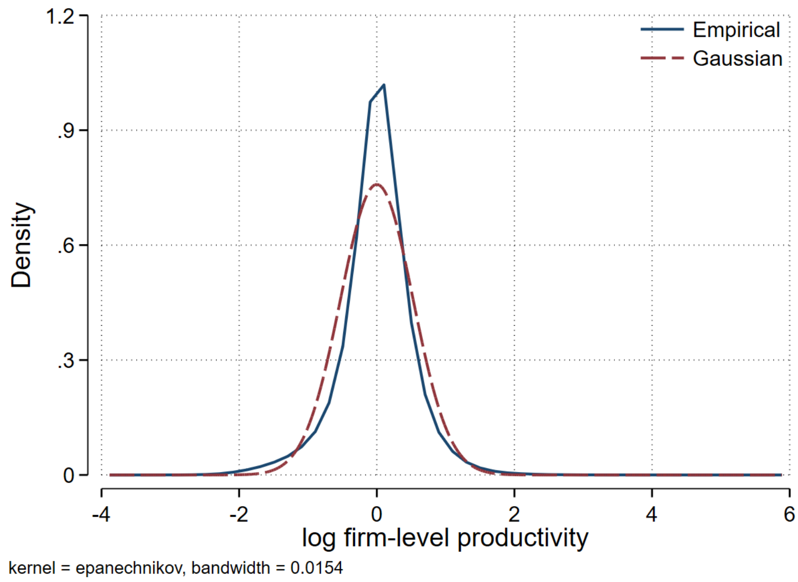
From the estimation of the production function parameters, we then obtain a series of firm-level productivity ω_{iht}:

$$\omega_{jht} = y_{jht} - \widehat{\beta}_{h,t} - \widehat{\beta}_{k,h} k_{jht} - \widehat{\beta}_{l,h} l_{jht}. \tag{3}$$

where $\widehat{\beta}_{h,t}$ is estimated as the sample mean of $y_{jht} - \widehat{\beta}_{k,h} k_{jht} - \widehat{\beta}_{l,h} l_{jht}$ for each sector h and period t. Therefore, we allow for time-varying differences in aggregate productivity across sectors. The idiosyncratic productivity component ω_{jht} has zero mean and represents productivity deviations to the sector average. Figure 2 plots the non-parametric kernel —solid blue— and the Gaussian kernel —dashed red— estimates of the unconditional productivity distribution. Productivity heterogeneity is large, most of the firms a productivity level going from 15% —$exp\{-2\}$— to 7 times —$exp\{2\}$— the average sector productivity. The comparison of the empirical cross-sectional distribution with the log-normal is interesting on its own as the AR(1) productivity dynamics would imply log-normality. The empirical distribution has longer tails than the Gaussian kernel, indicating a larger fraction of very low-productivity and high-productivity firms in the economy.

The subtle deviations from log-normality in the cross-sectional distribution may prelude richer productivity dynamics than those implied by the first-order autoregressive

Figure 2: Distribution of the idiosyncratic firm-productivity component



Note: The figure shows the non-parametric kernel density estimate for the idiosyncratic firm-level productivity component in the solid blue line and the Gaussian kernel estimate in the dashed red line. Data from the Spanish Central de Balances, from 1999 to 2017.

representation, which we look at in the next section.

3 The dynamics of firms' productivity

This section introduces a flexible stochastic process for idiosyncratic firm productivity. To motivate our approach, Figure 3 reports the second to fourth conditional moments of (the logarithm of) firm-level productivity by percentile of productivity in the previous period. The moments reported are persistence–as measured by the first-order autoregressive coefficients—and quantile-based measures of dispersion

$$\Delta = P90 - P10,$$

skewness

$$S_K = \frac{(P90 - P50) - (P50 - P10)}{P90 - P10}$$

and kurtosis

$$\kappa = \frac{P92 - P8}{P83 - P17}.$$

The measure of skewness S_K captures the asymmetry of the distribution as measured by relative fraction of the overall dispersion $(P90 - P10)$ accounted for by the upper $(P90 - P50)$ versus the lower $(P50 - P10)$ tail. It is negative when the lower tail is longer than the upper one. Kurtosis κ captures the length of the tails. Its value is 1.4 for the normal distribution.

Figure 3 reveals two features of the data. First, the distribution of productivity is not Normal, displaying negative skewness and excess kurtosis (long tails). Second, the conditional moments are not independent of the previous productivity realisation. Productivity is substantially more persistent, and skewness more negative, for firms with high productivity realisations. Also dispersion and kurtosis display substantial variation over productivity ranks.

This contrasts with the typical AR(1) productivity process, assumed in most models of heterogeneous firms' investment dynamics; namely

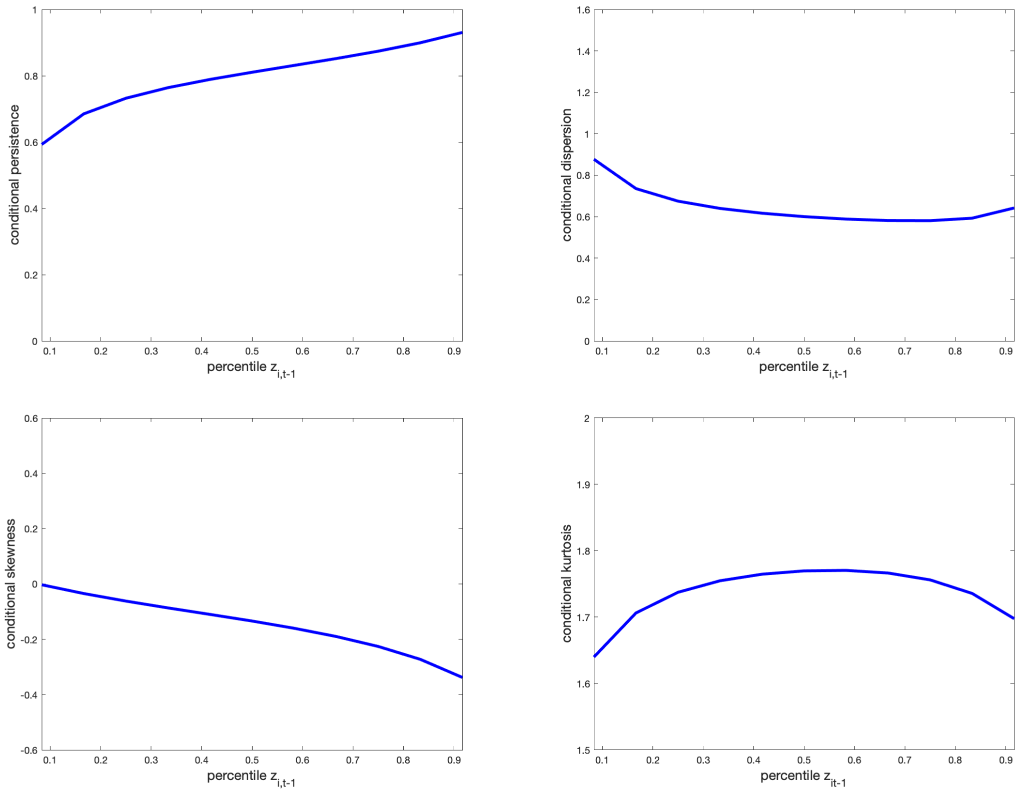
$$z_{jt} = \rho z_{j,t-1} + v_{jt}, \quad v_{jt} \sim N(0, \sigma),$$

where z_{jt} is the logarithm of productivity for firm j at time t. The AR(1) process is linear, which implies that persistence ρ and other second σ and higher-order conditional moments are independent of past histories. Both normality and linearity are at odds with the evidence in Figure 3.

It is this evidence which prompts us to estimate a flexible stochastic process for firms' productivity. In doing so we lean on a recent literature that has shown that the dynamics of both earnings (Arellano et al., 2017; De Nardi et al., 2019; Guvenen et al., 2021) and wages (De Nardi et al., 2021) display substantial deviations from normality and linearity.

In particular, we rely on the quantile-based panel data method proposed by Arellano et al. (2017) to estimate a flexible, non-parametric model that allows for non-normality and non-linearity, but nests the AR(1) process as a special case.

Figure 3: Preliminary evidence: second to fourth moments of productivity distribution conditional on productivity in the previous period.



Note: Top left: persistence (first-order autoregressive coefficient). Top right: dispersion. Bottom left: Kelley skewness. Bottom right: kurtosis.

3.1 A flexible productivity process

Consider a cohort of firms indexed by j and denote by t the firm's age (relative to $t = 1$). Let us assume that z_{it} is the logarithm of firm productivity that we extracted from the production function estimation in subsection 2.2. We can decompose z_{jt} into a persistent component, η_{jt} and a transitory component, ε_{jt}:

$$z_{it} = \eta_{it} + \varepsilon_{it}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{4}$$

wherein the probability distributions of η and ε are absolutely continuous.

The first deviation from the typical AR(1) assumption is that productivity is not restricted to follow a univariate process. This is standard in the literature on earnings dynamics going back to Abowd and Card (1989). The reason behind that is the presence of measurement error which would bias downward the estimated persistence. Allowing for measurement error is likely to be equally, if not more, important in the case of measured productivity. Similar formulations of the AR(1) productivity process have been proposed by Gourio (2008), Roys (2016) and Sterk et al. (2021).

The second deviation is that, although the persistent component η_{jt} is still assumed to follow a first-order Markov process, the process is general rather than restricted to be linear. This can be represented as

$$\eta_{jt} = Q_\eta(u_{jt}|\eta_{j,t-1}, t), \quad u_{jt}|\eta_{j,t-1}, \eta_{j,t-2}, \ldots, \eta_{j1} \sim \text{Uniform}(0,1) \quad t = 1, \ldots, T, \tag{5}$$

where $Q_\eta(\tau|\eta_{j,t-1}, t)$ is the conditional quantile function of $\eta_{j,t}$ given η_{jt-1}, for each $\tau \in (0, 1)$.

Intuitively, the quantile function is the inverse of the cumulative density function.[10] It maps random draws of u_{jt} from the uniform distribution —cumulative probabilities— to corresponding random draws (quantiles) from the conditional distribution of η_{jt}. The AR(1) (linear) case can be seen as a special case of (5) when the quantile function specialises to the linear separable form $\eta_{jt} = \rho\eta_{jt-1} + \Phi_t^{-1}(u_{jt}; \sigma)$, where $\Phi_t^{-1}(u_{jt}; \sigma)$ is the

[10] For a generic random variable v, the quantile function is a mapping from the interval $(0, 1)$ into the support of v. Namely, $v_q = Q(q)$ satisfies $P[v \leq v_q]$ where $P[\cdot]$ denotes the probability distribution of v.

inverse of the cumulative density function of a normal distribution with mean zero and standard deviation σ.

One way to understand the role of non-linearity is in terms of a generalized notion of persistence

$$\rho(\tau|\eta_{j,t-1}, t) = \frac{\partial Q_t(\tau|\eta_{j,t-1}, t)}{\partial \eta}, \tag{6}$$

which measures the persistence of η_{jt-1} when it gets hit by a current shock u_{jt} with rank τ. This quantity depends on the past productivity component η_{jt-1} and the percentile τ of the current shock. Note that while the shocks u_{jt} are i.i.d. by construction, they may differ along the persistence associated with them. One can then think of persistence in this context as *persistence of productivity histories*. Moreover, persistence is allowed to depend on the size and the direction of the shock u_{jt}. As such, the persistence of productivity of firm j in period $t-1$ will evolve depending on the size and sign of current and future shocks $u_{jt}, u_{jt+1}, \ldots, u_{jT}$. In particular, our model allows particular shocks to wipe out the memory of past productivity history. In contrast, in the AR process, $\rho(\eta_{jt-1}, \tau) = \rho$ is constant and, therefore, independent of the past productivity component η_{jt-1} or the shock realization u_{jt}.

A similar unrestricted representation can be used for the, zero-mean, *transitory* component ε_{jt} with the only difference that it is independent over time and of η_{jt}. The same representation can be used for the initial condition η_{j1}.

4 Estimation of productivity processes

4.1 The non-linear productivity process

Following Arellano et al. (2017), we specify the quantile functions for the persistent and transitory components as linear combinations of bivariate (Hermite) polynomial basis functions f_k, for $k = 0, 1, \ldots,$:

$$Q_\eta(\tau|\eta_{jt-1}, age_{jt}) = \sum_{k=1}^{K} a_k^\eta(\tau) f_k(\eta_{jt-1}, age_{jt}) \tag{7}$$

$$Q_\varepsilon(\tau, age_{jt}) = \sum_{k=1}^{K} a_k^\varepsilon(\tau) f_k(age_{jt}) \tag{8}$$

$$Q_{\eta_{j1}}(\tau) = \sum_{k=1}^{K} a_k^{\eta_1}(\tau) f_k(age_{j1}) \tag{9}$$

where age_{jt} is the age of the firm at time t $a_k^\eta(\tau)$, $a_k^{\eta_1}(\tau)$. The coefficients $a_k^\varepsilon(\tau)$ are modelled as piece-wise linear splines on a grid $[\tau_1, \tau_2]$, ..., $[\tau_{L-1}, \tau_L]$, which is contained in the unit interval. We then extend the specification for the intercept coefficients $a_0^\eta(\tau)$, $a_0^{\eta^1}(\tau)$, and $a_0^\varepsilon(\tau)$ to be the quantile of the exponential distribution on $(0, \tau_1]$ (with parameter λ_-^Q) and $[\tau_L, 1)$ (with parameter λ_+^Q).

If the persistent and transitory components of productivity were observed, one could estimate the parameters of the quantile models via ordinary quantile regression. However, as the two components are latent variables, we proceed with a simulation-based algorithm. Starting with an initial guess of the parameter coefficients, we iterate sequentially between draws from the posterior distribution of the latent productivity components and quantile regression estimation until convergence of the sequence of parameter estimates.

We provide a more detailed explanation of the estimation procedure and the identification arguments in Appendix C. We obtain standard errors via both non-parametric and parametric bootstrap, with 500 replications.

4.2 The canonical productivity process

We estimate the canonical process of productivity via a quasi-maximum likelihood procedure, which we explain in Appendix D. As such, we obtain the standard errors via the usual asymptotic variance calculation. The results of the estimation are in Table 2. As can be observed, we find that the persistent component is highly persistent, albeit with root below one.

Table 2: Parameters of the linear AR(1) process

Parameter	Estimate (Std. Err.)
Autoregressive coeff. of persistent component	0.924
	(0.086)
Std. dev. of innovation to persistent component	0.231
	(0.065)
Std. dev. of the transitory component	0.223
	(0.091)
Std. dev. of the initial condition	0.608
	(0.121)

Note: We report the parameter estimates of the linear AR(1) process for productivity. Standard errors are computed via the asymptotic variance calculation. Data from the Spanish Central de Balances, from 1999 to 2017.

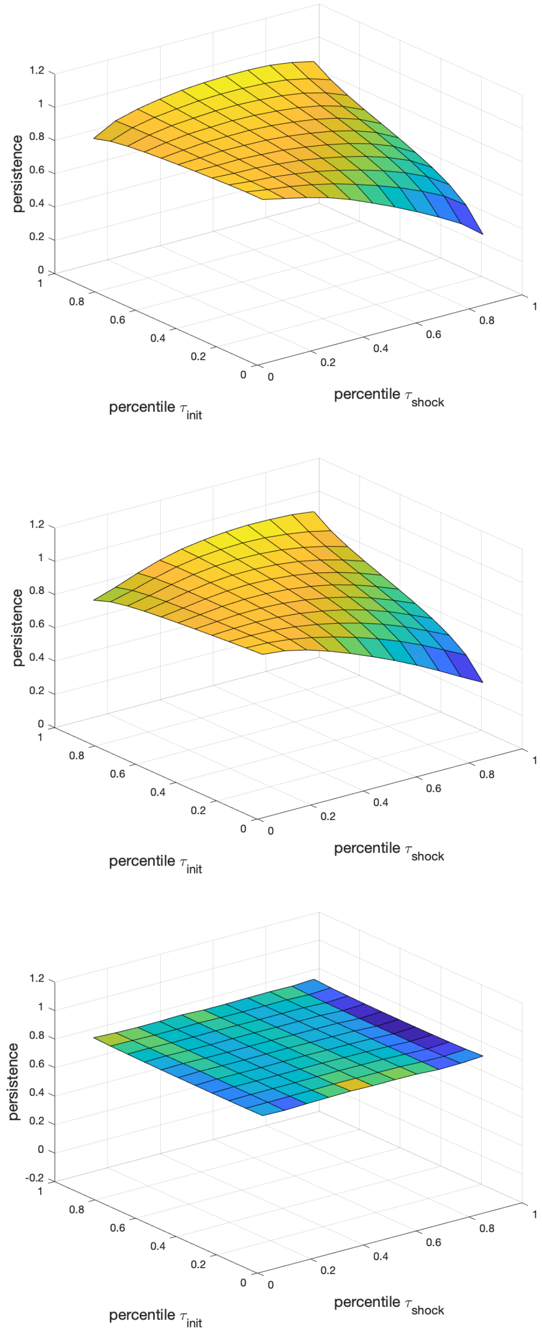
4.3 Comparing the canonical and the non-linear productivity processes

We compare the implications of the canonical and the non-linear productivity process to understand the differences between the two processes.

Figure 4 compares estimates of the persistence of productivity z_{jt} in the Spanish data to those that one obtains by simulating productivity under the two estimated processes. The upper left panels report the generalised persistence coefficient in equation (6) for the log of productivity z_{jt}, as a function of the previous productivity realization $z_{j,t-1}$ and the rank of the innovation τ.[11] The figure confirms the strong evidence of non-linear persistence in the top left panel in Figure 3. More importantly, it shows that the big driver of non-linearity is the fact that very good—high rank—shocks strongly reduce the persistence of previous productivity realisation for firms in the bottom half of the distribution of previous productivity. The upper right panel and bottom panel are the counterpart of the top left panel on but computed on data simulated respectively with the estimated non-linear and linear processes. It is clear that the non-linear process reproduces very well the patterns in the data. In contrast, the linear model fails to match such patterns as it implies constant persistence by construction. We provide the standard errors computed by nonparametric bootstrap in Figure 16 of Appendix E. As the graphs

[11] For clarity, the coefficients plotted in the top left in Figure 3 and the top panel in Figure 4 are related. The former is given by the latter averaged over the innovations.

Figure 4: Non-linear persistence of productivity



Note: The above figures show the estimated persistence implied by the data, the non-linear, and linear model. Upper left: Persistence estimated from the data. Upper right: Persistence estimated from the simulated data implied by the estimation of the non-linear productivity process. Bottom: Persistence estimated from simulated data implied by the estimated linear AR(1) process. Data from the Spanish Central de Balances, from 1999 to 2017.

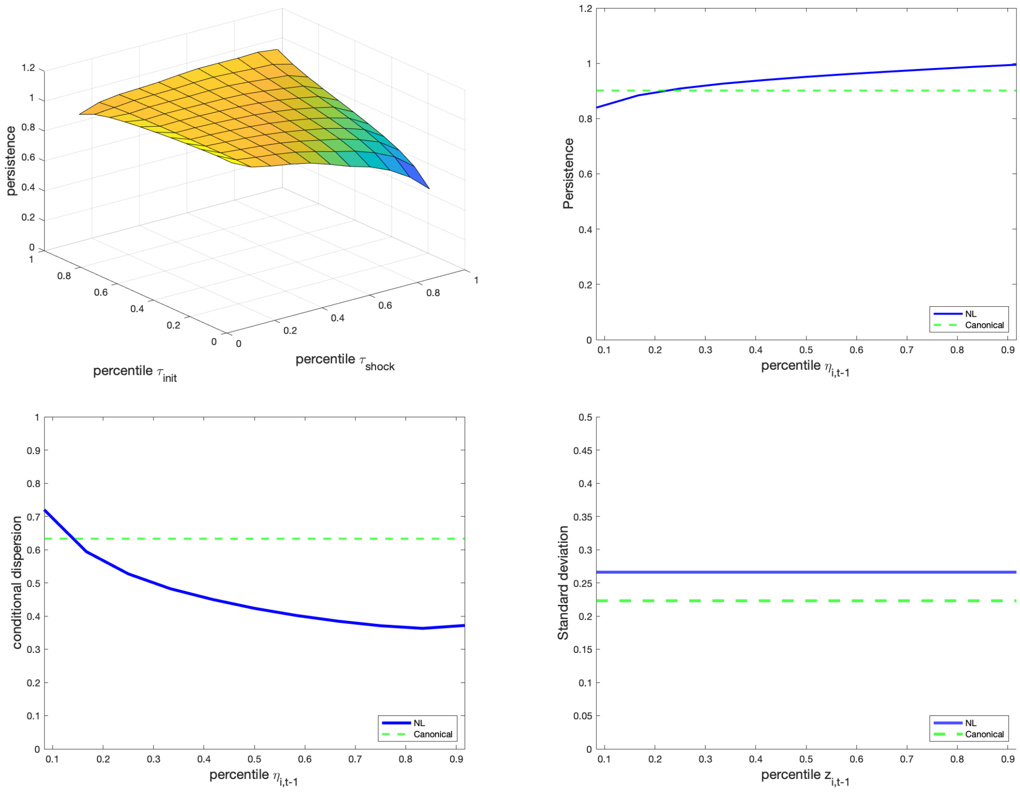
show, we find that our results on nonlinear persistence are precisely estimated.

Figure 5 shows differences in the persistence and conditional variance of the persistent and transitory productivity components between the two processes. The top left panel plots the generalized persistence coefficient in equation (6) for the persistent component η for the non-linear process, while the top right panel reports its average over the innovations and its, constant, counterpart for the canonical process. In the non-linear case, the persistence of η is increasing in the rank of the previous realisation, mainly because above median shocks partially wipe out the memory of previous histories. On the other hand, the linear model with constant persistence underestimates the average persistence of η. The bottom left and bottom right panels plot the dispersion of innovations to η and ϵ respectively. The linear process not only misses the non-linear pattern in the dispersion of shocks to the persistent. It also overestimates the dispersion of persistent shocks and underestimates that of transitory shocks, relative to the flexible process.[12]

Finally, we show the implied marginal distributions of the persistent component η and the transitory component ε for the non-linear model in Figure 6. Althhough there are slight departures from the non-Gaussian distribution in the persistent component η_{jt}, it is the transitory component that shows the clearest departure from a Normal distribution.

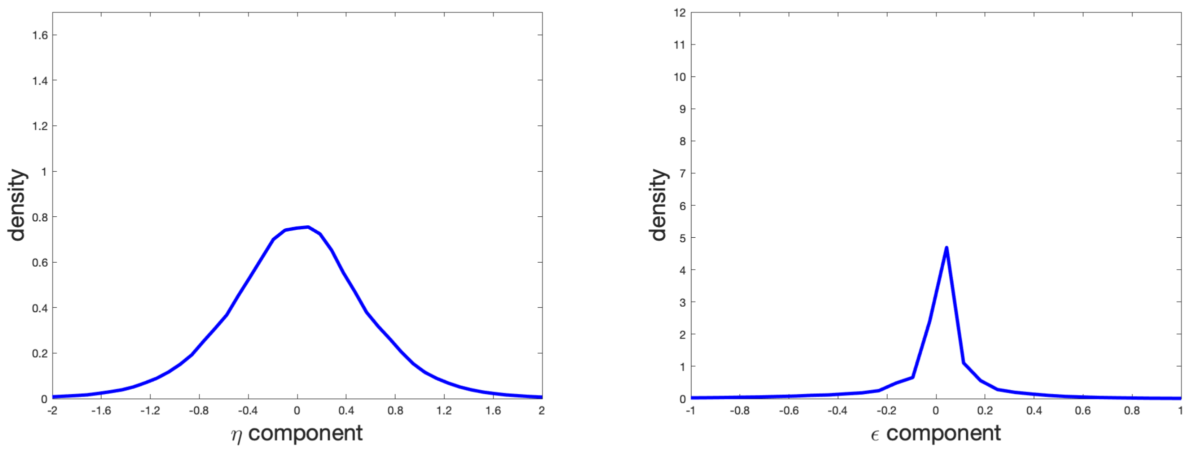
[12]We also show the conditional skewness measures in Figure 15 of Appendix E. As the results indicate, both processes imply differences in conditional skewness, with the nonlinear process being able to replicate the features of the data.

Figure 5: Nonlinear persistence and dispersion of the persistent and transitory components of productivity



Note: The above figures show the persistence and conditional dispersion of η_{jt} and ε_{jt}. Upper left: Persistence of $\eta_{j,t-1}$ as a function of $\eta_{j,t-1}$ and τ_{shock} for the non-linear process. Upper right: Persistence of $\eta_{j,t-1}$ for the non-linear and linear process. Lower left: Conditional dispersion of η_{jt} for the non-linear and linear process. Lower right: Dispersion of ε_{jt} for the non-linear and linear process.

Figure 6: Densities of η_{jt} and ε_{jt}



Note: Densities implied by the estimated non-linear productivity process. The figure on the left shows densities of the persistent component η_{jt}. The figure on the right shows densities of the transitory component ε_{jt}. Data from the Spanish Central de Balances, from 1999 to 2017.

5 Empirical investment policy functions

This section estimates empirical investment policy functions along the lines of Bazdresch et al. (2018); Gala et al. (2020). Policy functions are a mapping from the relevant state to the investment choice that summarises the key predictions of any investment model. As pointed out in Bazdresch et al. (2018); Gala et al. (2020), they are easy to estimate, as flexible polynomial approximations in the state variables, which can accommodate very general assumptions. In the specific case, they can accommodate both the non-linear and canonical productivity processes which allows us to use the same specification to estimate the policy functions under the two processes.

The resulting empirical policy functions are informative about the different implications of the two processes for the dynamics of investment, independently from any additional model structure. They may also used to guide the estimation of a structural model.

5.1 Empirical investment rule

Given the filtration implied by the two productivity processes an appropriate state vector for the investment policy function is given by the capital stock k at the beginning of the period and the two productivity components η and ε. This is true for the structural model in Section 6, but also for a much larger class of models with most types of adjustment costs considered in the literature.

This implies that he following empirical investment policy function is:

$$i_{jt+1} = g_t(k_{jt}, \eta_{jt}, \varepsilon_{jt}, u_{jt+1}), \quad t = 1, \ldots, T, \tag{10}$$

with i_{jt} firm net investment rate as a share of capital, k_{jt} be (log) of capital for firm j at time t. The u_{jt}'s are stochastic determinants of firm investment costs, which we assume to be independent of the state variables and independent over time.

The empirical policy function we specify here is similar to those of Bazdresch et al. (2018) and Gala et al. (2020), who utilize their empirical models to estimate structural models of firm investment. A crucial distinction between the model we present here and previous models in the literature is that we decompose productivity into persistent and transitory components.

5.1.1 Empirical specification

Given the potential non-normality of the investment distribution, we specify the conditional distribution of investment rates given capital and the components of productivity as the following quantile model for a given quantile τ_l:

$$i_{jt} = \sum_{k=0}^{K} a_k^I(\tau_l) f_k(k_{jt}, \eta_{jt}, \varepsilon_{jt}, age_{jt}) + \mathbf{X}_{jt}'\beta(\tau_l) \tag{11}$$

where $f_k(\cdot)$ is a dictionary of functions. In practice, we use tensor products of lower-order Hermite polynomials.[13] The model is a flexible specification of the conditional distribution of investment given the state variables, which does not impose specific distributional assumptions on investment rates. In contrast, both Bazdresch et al. (2018) and Gala et al. (2020) model just the mean of the conditional distribution of investment rates given state variables, thus implicitly assuming normality, which is at odds with the data.[14]

We also specify a vector of controls $\mathbf{X}_{jt}$, that aims to capture aggregate state variables that potentially affect firm investment decisions. In particular, the controls here are time dummies at the sector level that correspond to each year in the panel. As we have a quantile model, we also allow these variables to be quantile-specific.

Similar to the estimation of the productivity process, we model $a_k^I(\tau)$ as piece-wise linear splines on a grid $[\tau_1, \tau_2]$, ..., $[\tau_{L-1}, \tau_L]$, which is contained in the unit interval. We then extend the specification for the intercept coefficient $a_0^I(\tau)$ to be the quantile of the exponential distribution on $(0, \tau_1]$ (with parameter λ_-^Q) and $[\tau_L, 1)$ (with parameter λ_+^Q). We use tensor products of Hermite polynomials for $f_k(\cdot)$, each component of the product taking as argument a standardized variable.

This modelling specification stands in contrast to Gala et al. (2020) and Bazdresch

[13] In particular, in the results we present, the order of polynomials is (2,2,1,1).

[14] We estimated a model that is similar to those specified by Gala et al. (2020) and Bazdresch et al. (2018). Our estimation results indicate that while we are able to get similar derivative effects, we are unable to match the unconditional distribution of investment observed in the data. This is especially important if we would like to use the empirical policy function to generate targets for structural estimation.

et al. (2018), who proceed with a series specification of the empirical policy function. The disadvantage of modelling investment through a series specification is that it does not allow for interactions between different state variables. As it will turn out, these interactions appear to be important for firms' investment decisions.

5.2 Estimation strategy

The main challenge in estimating the empirical policy function that we specify here is that the stochastic components of productivity are unobserved. If we had in hand observations of the persistent and transitory components of productivity, we would be able to proceed with ordinary quantile regressions and estimate the policy functions directly.[15] However, the fact that the two components are latent, i.e. unobserved, requires a more sophisticated estimation algorithm. With this in mind, we proceed with the simulation-based estimation algorithm proposed in Arellano et al. (2017), modified to our specific context. Specifically, we recover estimates of the investment policy functions given the estimates of productivity process obtained earlier.[16]

The estimation procedure consists of two steps. Starting with an initial guess of the parameters of the investment function $\widehat{\mu}^{(0)}$, we iterate the following steps on $s = 1, 2, \ldots$ until convergence of the $\widehat{\mu}^{(s)}$ process:

1. *Stochastic E step:* For each firm, draw observations of the persistent component $\eta_j^{(m)} = (\eta_{j1}^{(m)}, \ldots, \eta_{jT}^{(m)})$ from the posterior distribution $f(\eta_{jt}|k_{jt}, i_{jt+1}, \varepsilon_{jt}; \widehat{\mu}^s, \widehat{\theta})$. We can then compute the associated transitory component $\varepsilon_{jt}^{(m)} = z_{jt}^{(m)} - \eta_{jt}^{(m)}$.

2. *M step:* Compute:

$$\min_{a_0^I(\tau_l),\ldots,a_K^I(\tau_l)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} \rho_{\tau_l}\left(i_{jt+1} - \sum_{k=0}^{K} a_k^I(\tau_l) f_k(k_{jt}, l_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}, age_{it+1}) - \mathbf{X}_{jt}'\beta(\tau_l)\right)$$

for $l = 1, \ldots, L$, where $\rho_{\tau_l}(\cdot)$ is the usual check function of quantile regression.

[15] Gala et al. (2020) circumvent this problem by working with observable proxies for productivity. However, as this paper aims to understand how the underlying productivity process affects firm investment, we do not proceed with this strategy.

[16] We could estimate the production function parameters, the productivity process and the investment policy functions jointly. However, we proceed sequentially as the productivity process is identified from the productivity series alone. Moreover, in a joint estimation approach, the estimates of the productivity process will be partly driven by the investment rule.

Note that the likelihood function can be written in closed form, implying that the E step is straightforward. In practice, we use a random walk Metropolis Hastings sampler that targets an acceptance rate of 30 percent. More importantly, the likelihood function is general enough to admit both the canonical and non-linear productivity processes in the estimation. This will be useful for the corresponding empirical exercise, as we will compare the empirical implications of the productivity process on firm investment. We provide additional details about the estimation algorithm, together with a sketch of non-parametric identification, in Appendix F.

In practice, we first estimate the effect of firm age on mean log productivity, mean log capital, and mean investment rates by regressing them on a quartic polynomial of firm age. The results for the estimation are based on $S = 200$ iterations, with 200 Metropolis-Hastings draws per iteration. Inference is based on non-parametric bootstrap, with 200 bootstrap replications.

5.3 Results

We compare the results of the estimated empirical investment policy functions when the underlying productivity process is non-linear, and when the underlying process is the canonical productivity process.

Our framework allows us to compute the following objects of interest. First, let us compute average investment rates as a function of capital, and the stochastic components of productivity:

$$\mathbb{E}(i_{jt+1}|k_{jt} = k, \eta_{jt} = \eta, \varepsilon_{jt} = \varepsilon) = \mathbb{E}(g_t(k, \eta, \varepsilon, u_{jt+1})) \tag{12}$$

Taking the average derivative of investment with respect to the persistent component of productivity yields the following object:

$$\phi_t(k, \eta, \varepsilon) = \mathbb{E}\left[\frac{\partial g_t(k, \eta, \varepsilon, u_{jt+1})}{\partial \eta}\right]. \tag{13}$$

The object $\phi_t(k, \eta, \varepsilon)$ reflects the degree to which firms increase or decrease their investment with respect to productivity shocks. We can define similar objects with respect to the transitory component of productivity and capital as well, and call these objects *marginal propensities to invest.*

Another set of object of interest that can be computed are the dynamic effects of productivity shocks on firm investment profiles. For example, the contemporaneous effect can be computed as:

$$\mathbb{E}\left[\frac{\partial}{\partial \nu}|_{\nu=\tau} g_t(k, Q_t(\eta, \nu), \varepsilon, u_{jt+1})\right] = \phi_t(k, Q_t(\eta, \tau), \varepsilon)\frac{\partial Q_t(\eta, \tau)}{\partial \nu}. \tag{14}$$

When the productivity process is non-linear, the derivative effect depends on the marginal propensity to invest and the quantity $\frac{\partial Q_t(\eta,\tau)}{\partial \nu}$, as the model allows for general forms of heteroscedasticity and skewness. In the empirical analysis, we will report finite-difference counterparts to the empirical objects we show here.

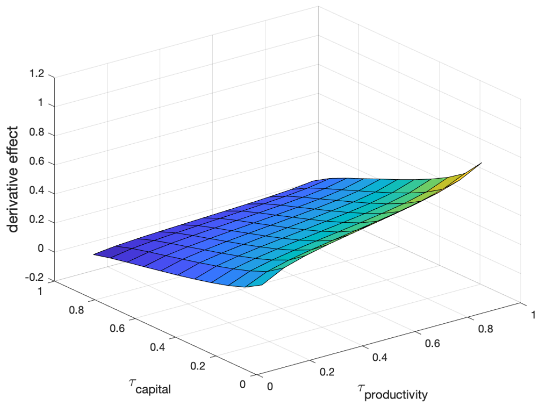
5.3.1 Marginal propensities to invest

We first discuss the marginal propensities to invest, beginning with the effects on productivity, which are in Figure 7. Figure 7a shows the average derivative effect, with respect to z_{jt}, of the conditional mean of i_{jt}, given k_{jt}, z_{jt}, and age_{jt}. This function is evaluated at different quantiles of k_{jt} and z_{jt} ($\tau_{capital}$ and $\tau_{productivity}$, respectively), and is averaged across age_{jt}. The resulting graphs show an effect that ranges from 0.02 to 0.70. Moreover, the graphs indicate that the derivative effect is increasing in productivity and decreasing in capital. This variation in responses suggests the presence of an interaction effect. We then compare the results that come from simulating the estimated semi-structural model of firm productivity and investment. The results, which are in Figure 7b, indicate that the model under the nonlinear productivity process is able to reproduce the patterns that we observe in the data quite well, with magnitudes that are quite similar (from 0.00 to 0.80). This stands in contrast to the model under the canonical productivity process (Figure 7c), where our results indicate that the marginal propensities to invest range from -0.20 to 1.05. The standard errors, which we provide in Figure 22 of Appendix G, show that the results are precisely estimated.

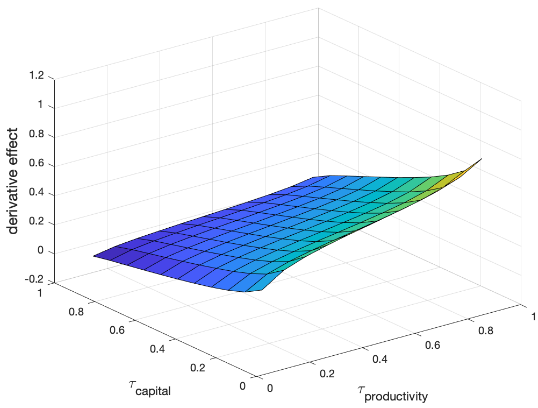
We then investigate the potential sources of differences between the simulated mod-

Figure 7: Marginal propensities to invest with respect to productivity

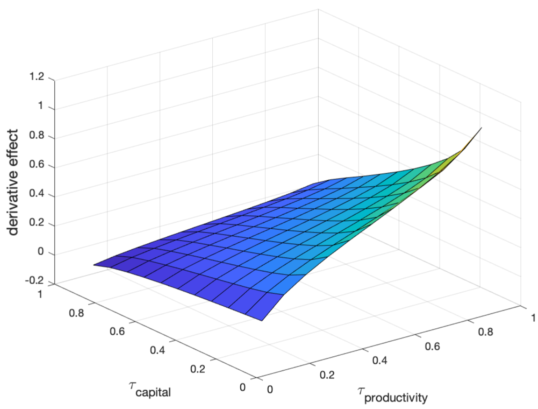
(a) Data



(b) Simulated model with the nonlinear process



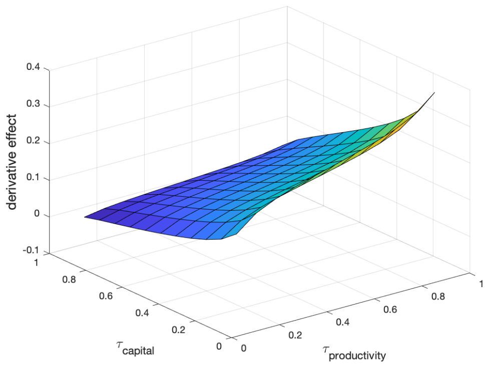
(c) Simulated model with the linear process



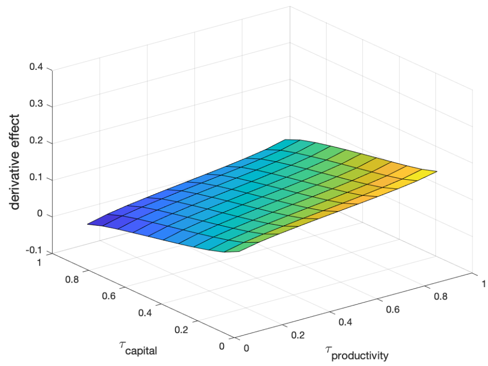
Note: The above figures show the average derivative effect of productivity on firm investment, evaluated at different percentiles of capital and productivity, and averaged across age. Top: investment responses from the data. Middle: investment responses to productivity z_{jt} based on simulated data from the model with the nonlinear productivity process. Bottom: investment responses to productivity z_{jt} based on simulated data from the canonical productivity process.

Figure 8: Marginal propensities to invest with respect to persistent and transitory productivity

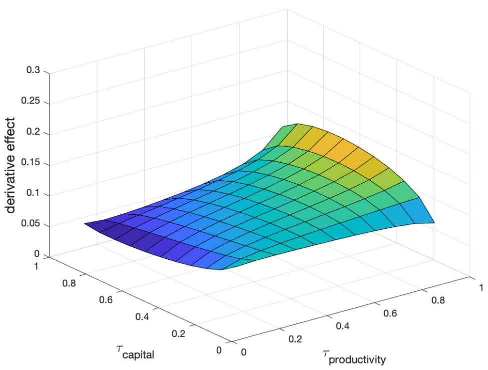
(a) Response to η_{jt}, nonlinear process



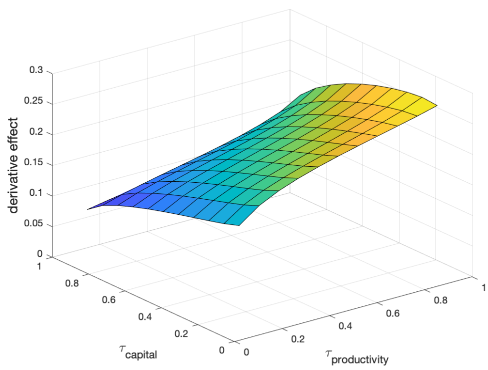
(b) Response to η_{jt}, canonical process



(c) Response to ε_{jt}, nonlinear process



(d) Response to ε_{jt}, canonical process



Note: The above figures show the average derivative effect of productivity on firm investment, evaluated at different percentiles of capital and productivity, and averaged across age. Top left: investment responses to persistent productivity η_{jt} based on the nonlinear process. Top right: investment responses to persistent productivity ε_{jt} based on the nonlinear process. Bottom left: investment responses to persistent productivity η_{jt} based on the canonical process. Top right: investment responses to persistent productivity ε_{jt} based on the canonical process.

els, which we provide in Figure 8. The graphs suggest that our earlier results for the model under the nonlinear productivity process is mainly driven by the persistent component η_{jt}, as can be seen in Figure 8a. Meanwhile, the model under the linear productivity process is mainly driven by the transitory component ε_{jt}, which can be observed on Figure 8d. The standard errors, which we provide in Figure 19 of Appendix G, confirm this result.

Aside from calculating the responses of investment to productivity, we also calculate the average derivative effects with respect to capital, which we show in Figure 20 in Appendix G. As the results indicate, the results are quite similar. Finally, we show the fit of the investment distribution predicted by the non-linear model and that of the data, which we show in Figure 21 of Appendix G. As the graph underscores, we are able to fit the investment distribution quite well.

5.3.2 Persistent responses to productivity shocks

In this part of the results section, we simulate productivity and investment decisions according to the non-linear model and show the evolution of productivity and investment following a persistent productivity shock. With some abuse in terminology, we will call these paths "impulse responses". We report the age-specific medians of log productivity of three types of firms: firms that are hit, at age five, by a large negative shock to the persistent component of productivity, by a large positive shock, and firms that are hit by a median shock to the persistent component. We report age-specific medians across 1,000,000 simulations of the model.

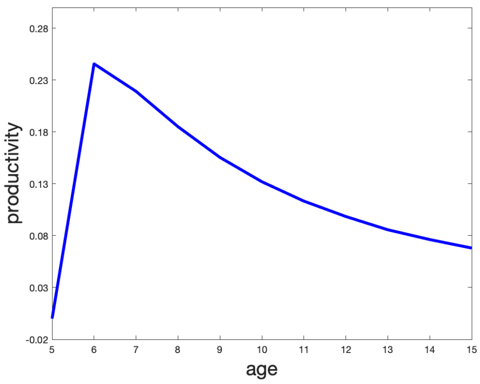
Figure 9 shows the impulse response functions with respect to productivity shocks. The results that we provide here indicate that there are interaction effects between the initial position of the firm in the productivity distribution and the size and sign of the shock to the persistent component of productivity. For example, a large positive shock to the persistent component of productivity is associated with a 25% increase in productivity for low productivity firms, while a similar shock is associated with a 10% increase for high productivity firms. We also find interaction effects with respect to large negative shocks. Heterogeneity in the persistence of the productivity shock is as well observed in the different speeds at which the productivity shock decays. This suggests the presence of asymmetries in the persistence of productivity histories, depending on the previous history of the firm and the size and magnitude of the shock. Figure 10, meanwhile, shows the results based on the canonical model of productivity dynamics. In this model, there are no interaction effects between the firm's position in the productivity distribution and the size and magnitude of the productivity shock. As it is clear, the implications of the non-linear model are different from that of the canonical model of firm dynamics.

Finally, in Figure 11, we report the results of a similar exercise, but we focus on impulse responses to investment.[17] The non-linearities that we observe for productivity also matter for firm investment. As an example, a large positive shock leads to an increase in investment by 20 percent for low productivity firms, while it is associated with an increase in investment by 6 percent for high productivity firms. We find similar results for large negative shocks.

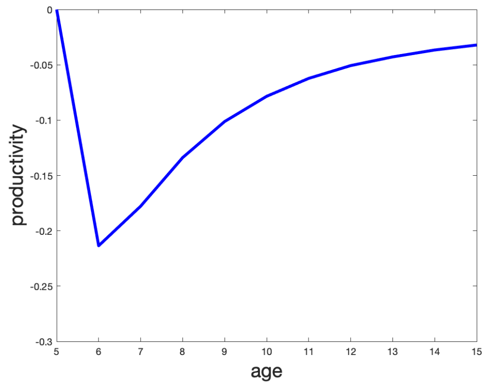
[17] In order to simulate investment paths, we need to make use of the law of motion for capital and specify an initial distribution for capital. For this exercise, we assume that the initial distribution is Normal, with the mean and variance of firms from 1 to 5 years old estimated in the data. The depreciation rate is assumed to be the average observed from the data, which is around 0.114.

Figure 9: Impulse response functions for productivity, non-linear model

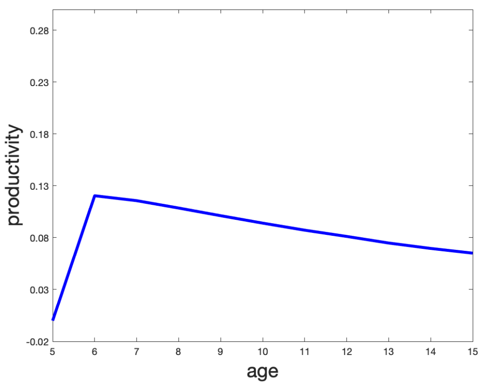
(a) $\tau_{init} = 0.1$, $\tau_{shock} = 0.9$



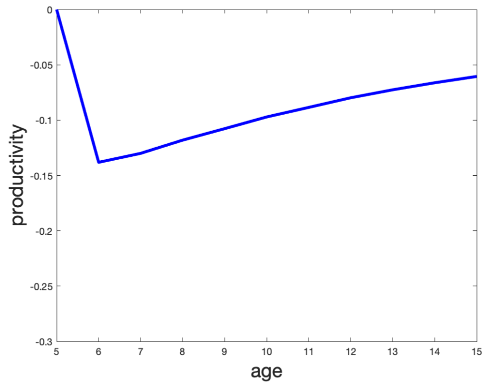
(b) $\tau_{init} = 0.1, \tau_{shock} = 0.1$



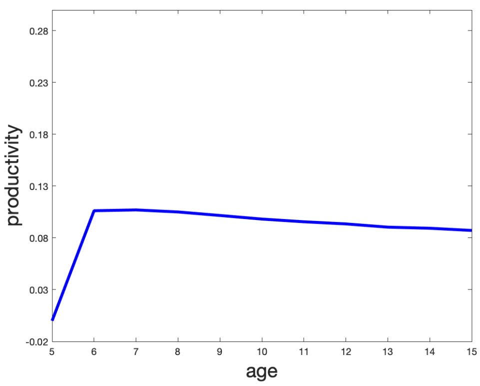
(c) $\tau_{init} = 0.5$, $\tau_{shock} = 0.9$



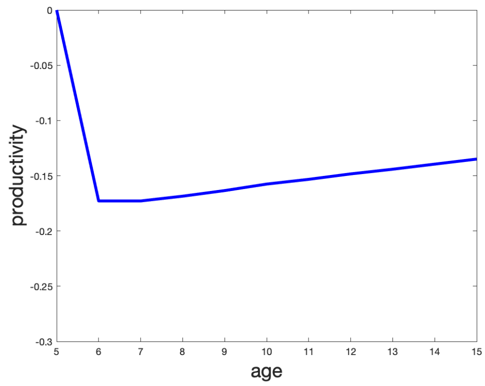
(d) $\tau_{init} = 0.5, \tau_{shock} = 0.1$



(e) $\tau_{init} = 0.9$, $\tau_{shock} = 0.9$



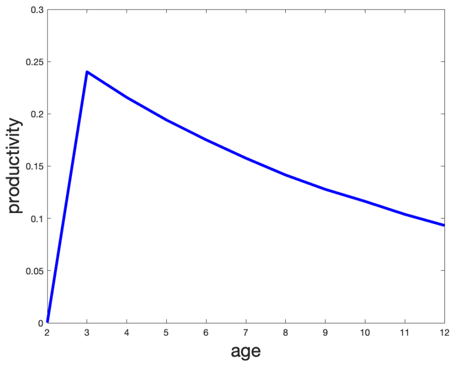
(f) $\tau_{init} = 0.9, \tau_{shock} = 0.1$



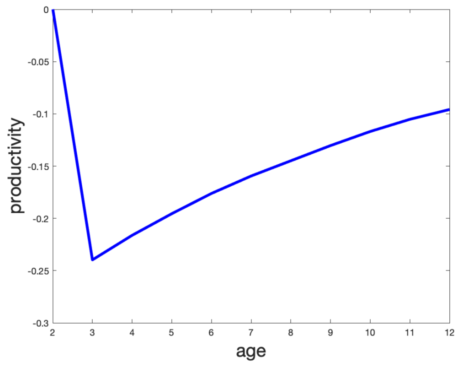
Note: The graphs show the difference between a firm hit by a shock τ_{shock} at age 5, and a firm hit by a 0.5 shock at the same age. Age-specific medians across 1,000,000 simulations. Model under the flexibile productivity process.

Figure 10: Impulse response functions for productivity, canonical model

(a) $\tau_{init} = 0.1$



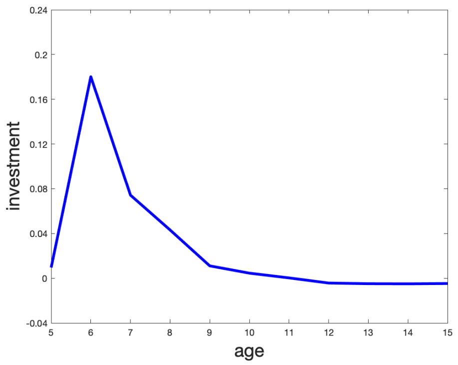
(b) $\tau_{init} = 0.1$



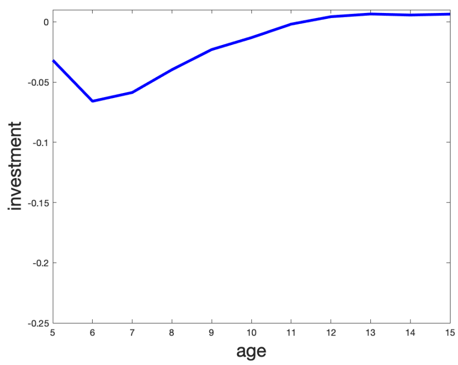
Note: The graphs show the difference between a firm hit by a shock τ_{shock} at age 5, and a firm hit by a 0.5 shock at the same age. Age-specific medians across 1,000,000 simulations. Model under the canonical productivity process.

Figure 11: Impulse response functions for investment

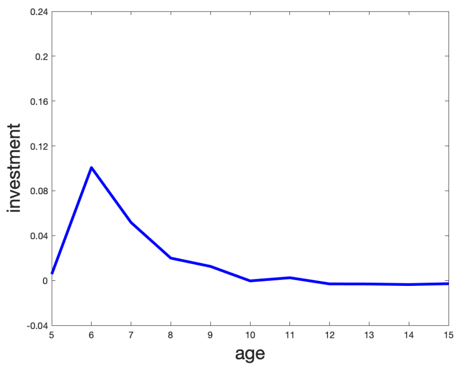
(a) $\tau_{init} = 0.1$, $\tau_{shock} = 0.9$



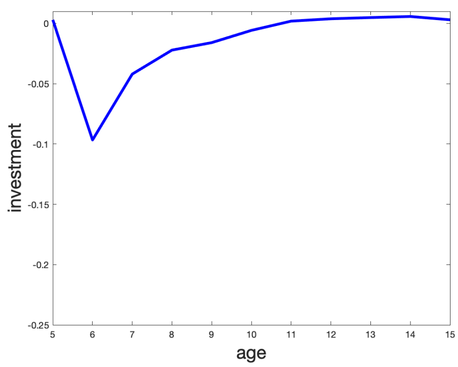
(b) $\tau_{init} = 0.1$,$\tau_{shock} = 0.1$



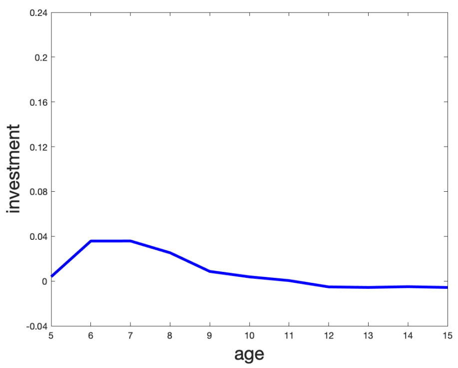
(c) $\tau_{init} = 0.5$,$\tau_{shock} = 0.9$



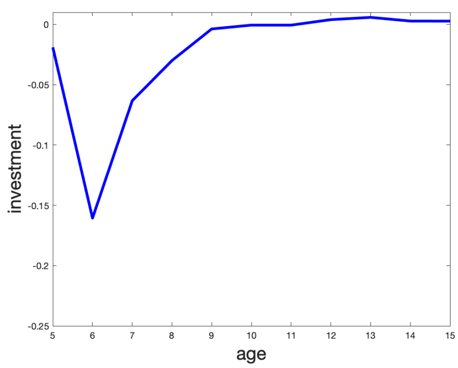
(d) $\tau_{init} = 0.5$,$\tau_{shock} = 0.1$



(e) $\tau_{init} = 0.9$,$\tau_{shock} = 0.9$



(f) $\tau_{init} = 0.9$,$\tau_{shock} = 0.1$



Note: The graphs show the difference between a firm hit by a shock τ_{shock} at age 5, and a firm hit by a 0.5 shock at the same age. Age-specific medians across 1,000,000 simulations. Model under the flexibile productivity process.

6 Structural estimation of capital adjustment costs

Next, we want to gauge the importance of the non-linear productivity process in estimating models of firm dynamics. Most research studying firm dynamics with firm heterogeneity uses linear productivity processes. These papers are trying to gauge the importance of adjustment costs, financial frictions or misallocation, among many others. However, given the evidence presented in the previous sections, their results might be biased since they are not using the right specification of the productivity process. In this section, we go back to one of the canonical models of capital adjustment costs, that of Cooper and Haltiwanger (2006), and we try to gauge the importance of the productivity process for the estimation of the adjustment cost parameters. In order to do so, we follow a similar procedure as they do, targeting a set of investment moments, first with the non-linear and then with the linear productivity process.

6.1 The Model

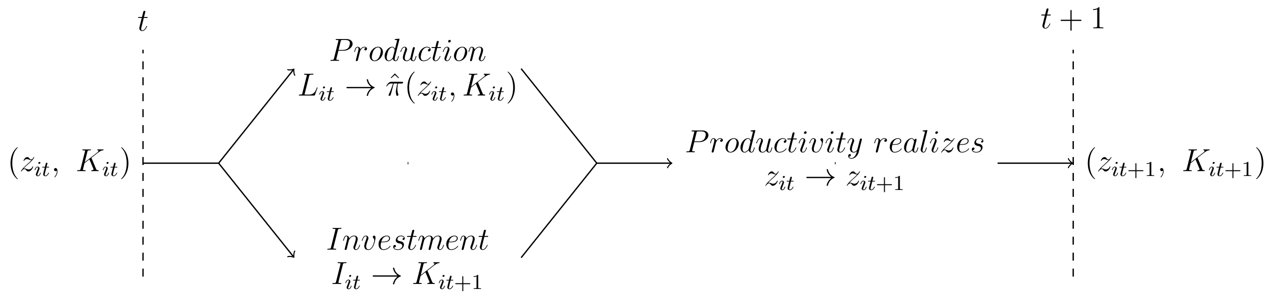
Time is discrete and the horizon is infinite. Each period is a year, and there is no aggregate uncertainty. Firms produce using capital K_{it} and labor L_{it} as inputs, with log productivity z_{it}. We will compare two economies, one where productivity follows the canonical process, and another one where productivity follows a non-linear process, as estimated in Section 4. The production function is Cobb-Douglas and features decreasing returns to scale,

$$f(z_{it}, K_{it}, L_{it}) = \exp(z_{it}) K_{it}^{\alpha_K} L_{it}^{\alpha_L} \quad \alpha_K + \alpha_L < 1. \tag{15}$$

Labor is a static choice, i.e., it is subject to no adjustment frictions, and it can be optimally chosen each period. Capital used for production in the current period is predetermined by investment in the previous period. The timing of the decisions is as shown in Figure 12.

Firms choose labor in every period to maximize current profits, given their produc-

Figure 12: Timing of the problem



tivity and their capital stock

$$\hat{\pi}(z_{it}, K_{it}) = \max_{\{L_{it}\}} \left\{ exp(z_{it}) K_{it}^{\alpha_K} L_{it}^{\alpha_L} - w_t L_{it} \right\}, \tag{16}$$

where $w_t > 0$ is the wage rate, which is assumed to be the same for all the firms. Hence, the optimal labor choice is given by

$$L_{it}^{*}(z_t, K_{it}) = \left[\frac{\alpha_L \ exp(z_{it}) \ K_{it}^{\alpha_K}}{w_t} \right]^{\frac{1}{1-\alpha_L}}; \tag{17}$$

and profits are given by

$$\pi_{it}^{*}(z_t, K_{it}) = (1-\alpha_L) \ \left[\frac{\alpha_L}{w_t} \right]^{\frac{\alpha_L}{1-\alpha_L}} \left[exp(z_{it}) \ K_{it}^{\alpha_K} \right]^{\frac{1}{1-\alpha_L}}. \tag{18}$$

Capital is accumulated and therefore, determined by previous investment following the law of motion

$$K_{it+1} = I_{it} + (1-\delta) K_{it}, \tag{19}$$

where I_{it} is the investment in capital and δ is the depreciation rate of capital.

As in Cooper and Haltiwanger (2006), investment in capital is potentially subject to both convex and non-convex adjustment costs and transaction costs:

- *Convex adjustment costs.* As it is standard in the literature, we adopt a quadratic cost specification

$$\frac{\gamma}{2} \left(\frac{I_{it}}{K_{it}} \right)^2 K_{it} \qquad \text{with} \quad \gamma \ \geq \ 0, \tag{20}$$

where γ is a parameter governing these costs.

- *Non-convex adjustment costs.* We also allow for the possibility of non-convex adjustment cost, which are independent of the amount of investment and only paid if the firm invests either a negative or a positive amount:

  - *Fixed cost.*

$$FC_K \mathbb{1}\left\{\left|\frac{I_{it}}{K_{it}}\right| > 0.01\right\} \qquad \text{with} \quad FC_K \geq 0, \tag{21}$$

    where FC_K is a parameter governing these costs.

  - *Opportunity cost as loss of current operational profits.*

$$\lambda\, \hat{\pi}(z_{it},\ K_{it})\ \mathbb{1}\left\{\left|\frac{I_{it}}{K_{it}}\right| > 0.01\right\} \qquad \text{with} \quad \lambda \in [0,\ 1), \tag{22}$$

    where λ is a parameter governing these costs.

- *Transaction costs.* On top of convex and non-convex adjustment costs, firms incur in transaction costs when selling/buying their capital. Especifically, we consider a lower recovery price for the capital sold. This considers there might be a gap between the buying and selling price of capital, which can be due to capital specifities or a “lemons” problem.

$$I_{it}\ \mathbb{1}\left\{I_{it} > 0\right\} + p_k\ I_{it}\ \mathbb{1}\left\{I_{it} < 0\right\} \qquad \text{with} \quad p_K \in [0,\ 1), \tag{23}$$

  where p_K governs the gap between the buying and selling price of capital.

  The problem of the firm is then given by

$$V(z_{it}, K_{it}) = \max_{I_{it}} \quad \hat{\pi}^*(z_{it}, K_{it}) - C(z_{it}, I_{it}, K_{it}) + \beta\ E\Big[V(z_{it+1}, K_{it+1})\Big| z_{it}\Big] \tag{24}$$

subject to

$$z_{it+1} = G_t(z_{it}, u_{it}), \tag{25}$$

$$K_{it+1} = I_{it} + (1-\delta)K_{it}, \tag{26}$$

$$\pi^*_{it}(z_t, K_{it}) = (1-\alpha_L)\ \left[\frac{\alpha_L}{w_t}\right]^{\frac{\alpha_L}{1-\alpha_L}} [exp(z_{it})\ K_{it}^{\alpha_K}]^{\frac{1}{1-\alpha_L}}\ , \tag{27}$$

$$C(z_{it}, I_{it}, K_{it}) = \frac{\gamma}{2}\ \left(\frac{I_{it}}{K_{it}}\right)^2\ K_{it} +\ FC_k\ \mathbb{1}\left\{\left|\frac{I_{i,t}}{K_{it}}\right|\ >\ 0.01\right\} \tag{28}$$

$$+\,\lambda\ \hat{\pi}(z_{it},\ K_{it})\ \mathbb{1}\left\{\left|\frac{I_{it}}{K_{it}}\right|\ >\ 0.01\right\} +\ I_{i,t}\ \mathbb{1}\left\{I_{i,t}\ >\ 0\right\}\ +\ p_K\ I_{i,t}\ \mathbb{1}\left\{I_{i,t}\ <\ 0\right\}.$$

Note that, given that there are no financial frictions or any other frictions distorting agent's choices, only permanent shocks matter for the investment choice. Firms' productivity z follows a process denoted by $G_t(z_{it}, u_{it})$, which can be an AR(1) process or the non-linear process estimated in Section 4.

6.2 Estimation and preliminary results

We estimate the model following the strategy in Cooper and Haltiwanger (2006). First, we set exogenously certain parameters. We set the annual discount factor β at 0.95 and the annual rate of depreciation δ at 11%, which we compute directly from the CBI data. We set $\alpha_K = 1/3 * 0.85$ and $\alpha_L = 2/3 * 0.85$, implying a span-of-control parameter of 0.85. So far, we assume there is no opportunity cost of capital, i.e. $\lambda = 0$.[18]

The remaining parameter, i.e. convex adjustment costs γ, the fixed cost FC_K, and the transaction cost p_K, are estimated by minimizing the distance between the simulated moments and the data moments. We target three moments: 1)frequency of positive investment spikes; 2)serial correlation of investment; and 3)the standard deviation of investment rates. Table 3 shows the main results of this analysis, with the parameters used for estimation (first panel), the targeted moments (second panel) and the non-targeted moments (third panel). The first column shows the data moments, and the first two columns show non-linear (first column) and the linear (second column) productivity

[18]This is because including the fixed cost and the convex adjustment cost is very similar to the opportunity cost loss.

processes economies with the parameters that minimize the loss function of the non-linear economy. The third and fourth columns show the moments of the economies with non-linear and linear productivity processes respectively, with the parameters that minimize the loss function of the linear economy.[19]. We can first compare the parameters we obtain when using different productivity processes. When using the linear productivity process, the fixed cost of investing is significantly larger than when using the non-linear productivity process. The convex adjustment cost parameter is significantly lower when using the linear process, and the price of selling capital is slightly lower. Hence, if we would be using the linear process to gauge the magnitude of capital adjustment costs in the Spanish economy, we would be understating the importance of convex adjustment costs by more than half, and overstating the magnitude of fixed adjustment costs significantly.

The model with non-linear productivity process matches the targeted moments fairly well (see first column), but with these parameters, the economy with linear productivity process matches poorly the moments, especially the autocorrelation of investment, which is much larger than in the data (second column). Regarding non-targeted moments, the model has a hard time to match the spike rate of negative investment, especially because it is particularly high in our data. Nonetheless, the negative spike rate is larger with the non-linear productivity process than with the AR(1) process. The model also falls short in matching the inaction rate, which is half of the inaction rate in the data. In the case of the AR(1) process, the inaction rate is still much smaller. The model with non-linear productivity overpredicts the fraction of observations with negative investment, although it is closer to the data than in the case of the AR(1) process. The non-linear productivity model also over predicts slightly the average investment rate, although less than the AR(1) model. Finally, there is a higher correlation of the profit shocks and investment in the model than in the data, but still the non-linear productivity process is closer to the data than the AR(1). Finally, the distribution of investment rates is more concentrated on the left tail and less concentrated on the right tail than in the data, but

[19]In each economy, we introduce a hicks-neutral productivity term such that the average size (in employment) of firms in all the economies is constant. This hicks-neutral term is 1, 1.25, 0.998 and 1.245 respectively for each column

Table 3: Targeted and untargeted moments

		Targeting NL		Targeting AR(1)	
	Data	**NL**	**AR(1)**	**NL**	**AR(1)**
Parameters					
γ		0.123	0.123	0.057	0.057
FC_k		0.085	0.085	0.134	0.134
p_k		0.798	0.798	0.733	0.733
Loss function		1.142	899.673	2001.341	434.087
Targeted moments					
Spike rate: positive investment	0.159	0.150	0.202	0.127	0.159
Serial correlation of investment rates	0.150	0.150	0.430	-0.099	0.351
SD investment rate	0.631	0.626	0.533	1.001	0.682
Non-targeted moments					
Spike rate: negative investment	0.204	0.088	0.068	0.087	0.045
Frac of obs with neg investment	0.645	0.742	0.774	0.770	0.828
Inaction rate: investment	0.057	0.026	0.002	0.009	0.001
Average investment rate	0.073	0.099	0.088	0.168	0.114
Correlation profit shocks and investment	0.100	0.364	0.418	0.314	0.418
p10(inv rate)	-0.301	-0.187	-0.119	-0.184	-0.117
p25(inv rate)	-0.179	-0.118	-0.113	-0.118	-0.112
p50(inv rate)	-0.065	-0.103	-0.107	-0.105	-0.108
p75(inv rate)	0.061	0.006	-0.100	-0.022	-0.102
p90(inv rate)	0.490	0.704	0.839	0.538	0.933

Notes: The table shows the outcomes of the model for the same parameters under non-linear productivity process (NL), and the AR(1) process. The first two columns use the parameters that minimize the distance between the targeted data moments and those in the economy with non-linear productivity process (NL).The last two columns use the parameters that minimize the distance between the targeted data moments and those in the economy with linear productivity process (AR(1)).

the match is significantly better than with the same parameters in the AR(1) process.

When we are targeting the moments in the economy with linear productivity process (third and fourth column), the match of the moments improves significantly (see loss function of column four compared to column two), but the model still has difficulties in matching the low serial correlation of investment rates, despite the high fixed cost of investing. The non-linear economy performs poorly with these parameters: the autocorrelation of investment rates is very low and becomes even negative with the lower convex adjustment costs and the higher fixed cost of investing, and the standard deviation of investment increases significantly. Regarding the untargeted moments, the economy with non-linear productivity process even outperforms the linear one in several dimensions: distribution of investment rates, negative spike rates, fraction of observations with negative investment and correlation of profit shocks and investment.

Summing up, estimating capital adjustment costs with the non-linear productivity process deliver significantly different results than with the linear productivity process, with higher convex costs and lower fixed cost of investing when using non-linear productivity process. Furthermore, it matches better the overall investment moments, both targeted and untargeted. These results point at the importance of utilizing the right productivity process when trying to quantify the importance of capital adjustment costs in particular, and this can be extrapolated to several quantification exercises of different nature done with models of firm dynamics in general.

7 Conclusion

This paper proposes a richer stochastic process for firm productivity that features a transitory and a persistent component, allows for nonlinearities in persistence, and non-Gaussian shocks. We find that the estimated nonlinear productivity process accurately captures the observed dynamics in firm productivity data. The estimated productivity dynamics emphasises varying persistence and conditional skewness across productivity history, which are at odds with the canonical AR(1) productivity representation used in the literature.

We then estimate an empirical investment rule under the two productivity dynamics processes. The investment responses to persistent productivity shocks are more prominent when the underlying process is nonlinear compared to the canonical one. We finally estimate a firm investment model that features both convex and non-convex adjustment costs. Our findings indicate that the estimated model under the richer productivity dynamics implies closer-to-data investment dynamics and a different characterisation of the adjustment costs nature with lower fixed costs than the canonical productivity representation.

References

Abowd, J. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–45.

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

Arellano, M. (2003). *Panel data econometrics*. Oxford university press.

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

Bazdresch, S., R. J. Kahn, and T. M. Whited (2018). Estimating and testing dynamic corporate finance models. *The Review of Financial Studies 31*(1), 322–361.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

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

De Nardi, M., G. Fella, M. Knoef, G. Paz-Pardo, and R. Van Ooijen (2021). Family and government insurance: Wage, earnings, and income risks in the netherlands and the u.s. *Journal of Public Economics 193*, 104327.

De Nardi, M., G. Fella, and G. Paz-Pardo (2019, 03). Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare. *Journal of the European Economic Association 18*(2), 890–926.

Gala, V. D., J. F. Gomes, and T. Liu (2020). Investment without q. *Journal of Monetary Economics 116*, 266–282.

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.

Gourio, F. (2008). Estimating Firm-Level Risk. Mimeo.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). What do data on millions of u.s. workers reveal about lifecycle earnings dynamics? *Econometrica 89*(5), 2303–2339.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020). Estimation with aggregate shocks. *The Review of Economic Studies 87*(3), 1365–1398.

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150.

Khan, A. and J. K. Thomas (2008). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*.

Kim, K. I., Y. Luo, and Y. Su (2019). A robust approach to estimating production functions: Replication of the acf procedure. *Journal of Applied Econometrics 34*(4), 612–619.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies 70*(2), 317–341.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263.

Peñaranda, F. and E. Sentana (2015). A unifying approach to the empirical evaluation of asset pricing models. *Review of Economics and Statistics 97*(2), 412–435.

Roys, N. (2016). Persistence of shocks and the reallocation of labor. *Review of Economic Dynamics 22*, 109–130.

Sterk, V., P. Sedláček, and B. Pugsley (2021). The nature of firm growth. *American Economic Review 111*(2), 547–79.

A Data appendix

This appendix provides further details about our primary dataset —*Central de Balances Integrada* (CBI) of *Banco de España* (BdE)—, the cleaning procedure that led to the final sample, the final sample representativeness in the economy, and the investment statistics used in the model section.

A.1 Construction of the Final Sample

The CBI contains firm-level data of Spanish firms since 1995. The data has an administrative nature as limited liability firms have the legal obligation to deposit their annual accounts (balance sheet, income statement and annual report) at the Commercial Registry every year. BdE has had an agreement with the Commercial Registries to access that information since 1995. It compiles and homogenizes the information received in a unique dataset called CBI. The lack of capacity and coordination among the Commercial Registries result in lower coverage for the first years of the agreement. Therefore, we focus on the period from 1999 —coverage stabilizes at a high level— to 2017 —last year with a final sample at the writing of the paper—.

We use the following raw variables

- id: Firm identifier,
- cif: Fiscal identifier,
- any: Year of the annual accounts,
- anyconst: Year the firm was set up,
- cnae09: Sector of the main activity at the 4 digits of the *Clasificación Nacional de Actividades Económicas* (2009-CNAE),
- grup: Indicator of the firm ownership,
- production: Value of production,
- materials: Value of intermediate inputs,

- wagebill: Total expenditure in employment,
- totalassets: Value of the total assets of the firm,
- longtermassets: Value of the long-term assets,
- intangibles: Value of the long-term of the intangible assets,
- productive: Value of long-term assets used in production,
- shorttermassets: Value of the short-term assets,
- workers: Number of employees in full-time equivalence units,

to construct our main variables. The firm's age is defined as the difference between any and anyconst variables. The sector of the firm corresponds to the first 2-digits of the cnae09 variable. Value-added is given by the difference between the production and materials variables. We define capital as the sum of the intangibles and productive variables. Finally, labor is given by the wagebill variable. We construct price indexes with 2006 base year to deflate the nominal variables at the 2-digits 2009-CNAE disaggregation. We use the value-added price index computed with the information from the National Accounts to deflate production, value-added, and materials. We use the investment price index to deflate the capital variable. We also deflate the labor variable using value-added price index.

We focus our analysis on limited liability firms that are not held by the public sector. We identify limited liability firms using the cif —Spanish fiscal identifier— variable. Firms with a cif starting by A are large limited liability corporations, while those with a cif starting by B are small limited liability firms. We identify the firms not controlled by the public sector with the information in the grup variable. We then create flags and drop those observations for which the sector is not identified, age is not identified or is 0, have 0 or negative production, value-added, materials, capital, labor, or wage bill.

Some firms are purely instrumental; their purpose is to reduce the tax bill of their owners and hide heritage to the fiscal authorities. Those firms usually have negligible economic activity. Therefore, we restrict our attention to firms with a reasonable level of activity defined as production larger or equal to 1,000 €, value-added larger or equal to 1,000 €, materials larger or equal to 100 €, capital larger or equal to 500 €, employment larger or equal to 0.25, and wage-bill larger or equal to 3,000 €; all in real 2006 €.

We next implement the following cleaning strategy. First, we do a winsorization of bottom and top wages at 1%, across years, within years and sector-year. This step detects wrong units by linking a monetary —the wage bill— and non-monetary —the number of workers—. Second, we search for rank discrepancies in the distributions of value-added, materials, capital, and labor, i.e. a firm in the first decile of the value-added distribution, but in the top percentile of the materials distribution, or a firm in the top decile of the value-added distribution, but in first percentile of the materials distribution. Third, we do a winsorization of bottom and top ratios at 1%, both within and across sectors, of value-added (va) with other variables —va/m, va/k, va/l—, materials (m) with other variables —m/k, m/l—, capital (k) with other variables —k/l—, and labor (l). Table 5 has a summary of all the steps to go from the raw data to the final sample. The final sample has contains 7,032,977 firm-year observations from 1,148,756 different firms.

Table 5 reports the summary statistics of the main variables used in the table. And, table 6 summarizes the panel dimension of the data.

A.2 Representativeness

[TBC]

A.3 Investment Data

We now lay down the steps to recover investment data from the CBI. Consider the capital accumulation equation of a firm i from period $t-1$ to period t

$$K_{i,\ t} \ = \ K_{i,\ t-1} \left(1 \ - \ \delta_{i,\ t}\right) \ + \ Inv_{i,\ t},$$

where $\delta_{i,\ t}$ is the depreciation rate at which capital has depreciated in firm i from period $t-1$ to period t, and $Inv_{i,\ t}$ is the investment undertaken for firm i from period $t-1$ to period t. After rearranging, the investment rate of firm i from period $t-1$ to period t is

Table 4: Summary of the cleaning procedure

Cleaning Steps	Version 7
Raw Dataset	17,697,126
Flags on	9,393,862
Sector	1,361,472
Age	1,277,586
Production	4,270,574
Value Added	4,686,449
Materials	1,651,509
Capital	3,798,581
Labor	6,442,065
Wage Bill	5,598,756
Intermediate Sample	8,303,264
Sample selection I on	65,639
No Limited Liability	54,054
Public Sector	12,477
Intermediate Sample	8,237,625
Sample selection II on	279,425
Production	8,314
Value Added	28,307
Materials	1,335
Capital	148,581
Labor	69,448
Wage Bill	78,795
Intermediate Sample	7,958,200
Cleaning on	925,223
Wages	203,963
Rank Discrepancies	14,118
Ratios	771,950
Final Sample	7,032,977

Note: We report the summary of the steps to go from the raw data to the final sample. Data from the Spanish Central de Balances, from 1999 to 2017.

Table 5: Summary of panel dimension of the final sample

Variable	Min	P_{25}	P_{50}	P_{75}	Max	Mean	Std. Dev.
Year	1999	2005	2009	2013	2017	2008.96	5.11
Age	1	5	10	17	149	11.89	9.00
Value Added	1.00	51.35	117.99	291.11	8,072,300	616.74	13,736.49
Materials	0.56	71.97	198.53	613.10	1.71e+7	1,736.04	42,743.29
Capital	0.50	18.41	67.34	252.99	1.61e+7	816.91	36,081.80
Labor	3.00	41.81	93.91	224.15	2,634,868	413.48	6,818.45

Note: We report the summary statistics of the final sample. Data from the Spanish Central de Balances, from 1999 to 2017.

Table 6: Summary of the cleaning procedure

	Number of obs.		Number of firms with longest spell at least			
Sector	Firm-Year	Firms	2 periods	3 periods	4 periods	5 periods
Smallest	124	36	26	21	5	3
2_{nd} Smallest	1,333	276	212	152	118	88
3_{rd} Smallest	3,667	424	375	324	291	186
Largest	1,080,605	198,227	153,401	119,405	94,900	75,608
All	7,032,977	1,148,756	922,161	741,756	609,542	503,036

Note: We report the summary of the panel dimension length of the final dataset. Data from the Spanish Central de Balances, from 1999 to 2017.

given by

$$inv_{i,\ t} \ \equiv \ \frac{Inv_{i,\ t}}{K_{i,\ t-1}} \ = \ \frac{K_{i,\ t} \ - \ K_{i,\ t-1}}{K_{i,\ t-1}} \ + \ \delta_{i,\ t}.$$

The investment rate (gross) has two components:

- Net investment rate: $\frac{K_{i,\ t} \ - \ K_{i,\ t-1}}{K_{i,\ t-1}}$, which is observable from the balance sheet information. We winsorize the two tails at the 1% level.

- Depreciation rate: $\delta_{i,\ t}$, which is not observable from the balance sheet information.

We recover the depreciation rate ($\delta_{i,\ t}$) using information from the income statement. Particularly, we use amortizations, depreciation and provisions. This has some caveats. First, it contains the amortization of all the assets and not only our measure of capital. Second, it contains negative changes in prices of all assets, i.e. depreciation. Finally, it contains savings due to likely events that the firm may face in the future, i.e. provisions. Unfortunately, we do not have information to disentangle them. Taking into account that our measure of capital is wide and depreciations and provisions are barely used, we correct the depreciation rate as follows. First, Define the provisional depreciation rate as follows

$$\delta_{i,\ t}^{prov} \ = \ \frac{\text{amortizations, depreciation and provisions from period } t-1 \text{ to } t}{K_{i,\ t-1}}.$$

We next set to missing the values that are likely to contain other things than just amortization. If the $\delta_{i,\ t}^{prov}$ is smaller to 0.02 or bigger than 0.2, they imply 50 and 5 years respectively of capital life expectancy under a linear system. Then, we run a regression of $\delta_{i,\ t}^{prov}$ on a 3-order degree polynomial of age, capital (in logs) and labor (wage bill in logs) in period $t-1$, with all the interactions for each sector-year level. We use this auxiliary model to predict the depreciation rate for all the firms. Finally, we reset the bounds of the predicted depreciation rate at the firm level to 0.02 and 0.20. This only happens very rarely 0.33 firms of each 1,000 for each bound. An statistic summary of the resulted depreciation rate $\delta_{i,\ t}$ is in table 7.

Table 8 reports several statistics of the net investment rate. Mean, standard deviation and autocorrelation, correlation with the productivity in period $t-1$, in logs,

Table 7: Summary of panel dimension of the final sample

Variable	P_1	P_{25}	P_{50}	P_{75}	P_{99}	Mean	Std. Dev.	Auto-corr
$\delta_{i,\ t}$	0.042	0.096	0.119	0.134	0.158	0.113	0.027	0.929

Note: We report the summary statistics of the panel length in the final sample. Data from the Spanish Central de Balances, from 1999 to 2017.

$\rho(inv_t,\ a_{t-1})$, correlation with the productivity shock from period $t-2$ to period $t-1$, $\rho(inv_t,\ \Delta A_{t-2})$, where $\Delta A_{t-2} = \frac{A_{t-1}\ -\ A_{t-2}}{A_{t-2}}$, inaction rate defined as $\mid inv_{i,\ t} \mid \leq\ 0.01$, negative investment rate, $inv_{i,\ t}\ <\ -0.01$ and negative spyke, $inv_{i,\ t}\ \leq\ -0.2$, positive investment rate, $inv_{i,\ t}\ >\ 0.01$ and positive spyke, $inv_{i,\ t}\ \geq\ 0.2$, and finally, some percentiles of the $inv_{i,\ t}$ distribution conditional on not being in inaction rate, $\mid inv_{i,\ t} \mid \leq\ 0.01$. Figure 13 plots the histogram of net investment rates. Finally, table 9 and figure 14 show the summary statistics and the histogram for gross investment rates.

Table 8: Summary statistics of net investment rates

Statistic	Value	Inaction Rate			5.7%
Mean(inv_t)	0.073	Negative	64.5%	P_{10}	-0.301
SD(inv_t)	0.631	Spyke (-)	20.4%	P_{25}	-0.179
$\rho(inv_t,\ inv_{t-1})$	0.150	Positive	29.9%	P_{50}	-0.065
$\rho(inv_t,\ a_{t-1})$	0.180	Spyke (+)	15.9%	P_{75}	0.061
$\rho(inv_t,\ \Delta A_{t-2})$	0.100			P_{90}	0.490

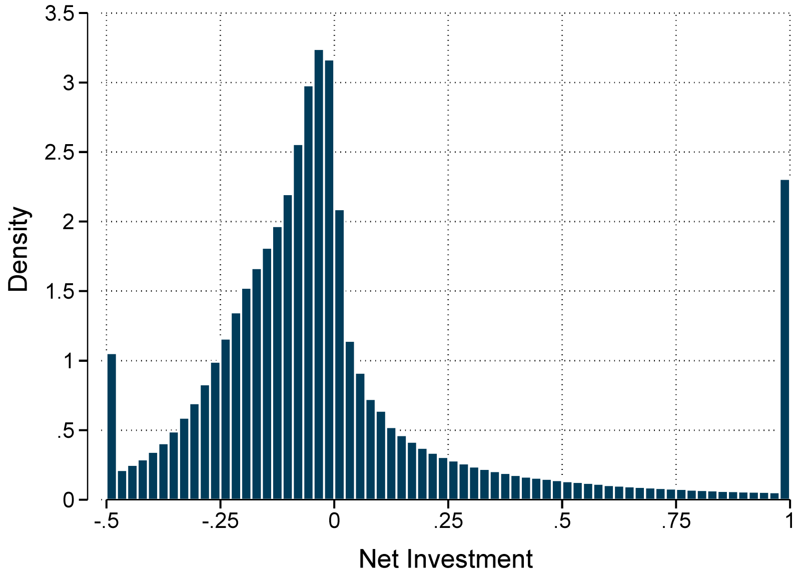
Note: We report the summary statistics of the net investment rates. Data from the Spanish Central de Balances, from 1999 to 2017.

Table 9: Summary statistics of gross investment rates

Statistic	Value	Inaction Rate			6.0%
Mean(inv_t)	0.185	Negative	34.5%	P_{10}	-0.175
SD(inv_t)	0.634	Spyke (-)	7.7%	P_{25}	-0.060
$\rho(inv_t,\ inv_{t-1})$	0.154	Positive	59.6%	P_{50}	0.048
$\rho(inv_t,\ a_{t-1})$	0.180	Spyke (+)	21.4%	P_{75}	0.171
$\rho(inv_t,\ \Delta A_{t-2})$	0.108			P_{90}	0.614

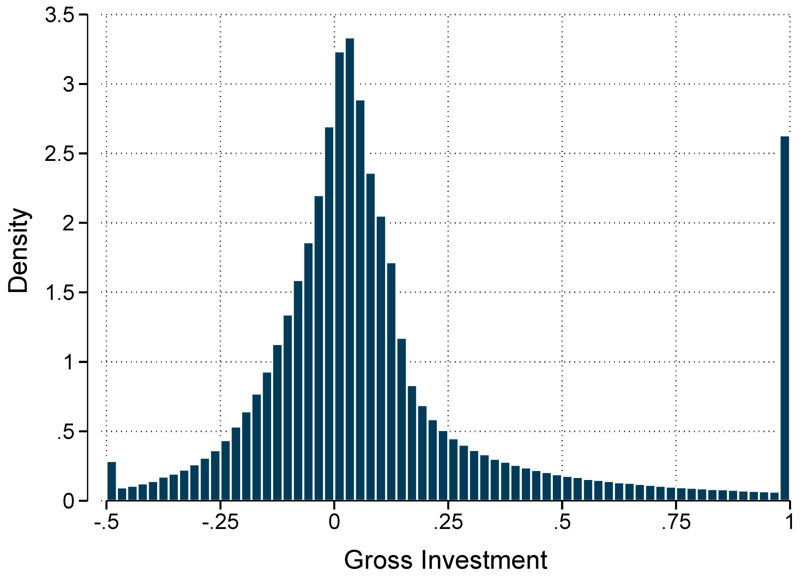
Note: We report the summary statistics of the gross investment rates. Data from the Spanish Central de Balances, from 1999 to 2017.

Figure 13: Histogram of net investment rates



Note: The figure shows the histogram of the net investment rates. Data from the Spanish Central de Balances, from 1999 to 2017.

Figure 14: Histogram of gross investment rates



Note: The figure shows the histogram of the gross investment rates. Data from the Spanish Central de Balances, from 1999 to 2017.

B Estimation of the production function

To estimate TFP, we appeal to the estimation procedure of Ackerberg et al. (2015) (hereafter referred to as ACF), albeit with slight modifications. The choice of this estimation procedure, as opposed to other popular procedures such as Olley and Pakes (1996) and Levinsohn and Petrin (2003), is due to the flexibility of ACF to deal with the possibility of adjustment costs in capital and labor.[20] We outline the assumptions for our estimation here:

1. *Information set:* The information set Ω_{jht} includes past and current productivity shocks $\{z_{jh\tau}\}_{\tau=0}^{t}$, but not future ones. Transitory shocks satisfy $E[\varepsilon_{jht}|\Omega_{jht}] = 0$.

2. *First-order Markov:* Individual productivity z_{jht} evolves according to a first-order Markov process known by the firms, i.e.,

$$p(z_{jht}|\Omega_{jht}) = p(z_{jht}|z_{jht-1}). \tag{29}$$

3. *Timing of input choice:* Capital at time t is determined by the law of motion

$$k_{jht} = x_{jht-1} + (1-\delta)k_{jht-1}, \tag{30}$$

   where x_{jht-1} is investment at $t-1$ and δ is the depreciation rate. Labor input has potential dynamic implications and is chosen at period $t-1$.[21]

4. *Intermediate inputs:* Intermediate inputs are given by

$$m_{jht} = f_t(k_{jht}, l_{jht}, z_{jht}), \tag{31}$$

   where the function $f_t(k_{jht}, l_{jht}, z_{jht})$ is invertible and strictly increasing in z_{jht}.

From the assumptions outlined above, ACF propose to estimate the parameters of the production function via the following two-step procedure:

[20] An equally flexible estimation procedure is the one proposed by Gandhi et al. (2020), who use the optimization conditions implied by the firm problem to achieve identification of the structural parameters. However, this procedure requires the availability of price data, which we do not have.

[21] This is important to account for the duality of the Spanish labor market.

1. *First stage.* Invert the conditional input demand function and perform an ordinary least squares (OLS) regression of y_{jht} on l_{jht}, k_{jht}, and $f_t^{-1} = \varphi_t(k_{jht}, l_{jht}, m_{jht})$:

$$y_{jht} = \beta_{0,h} + \beta_{l,h} l_{jht} + \beta_{k,h} k_{jht} + \varphi_t(k_{jht}, l_{jht}, m_{jht}) + \varepsilon_{jht} \equiv \Phi_t(k_{jht}, l_{jht}, m_{jht}) + \varepsilon_{jht} \tag{32}$$

2. *Second stage.* The first-order Markov assumption on productivity allows us to write the following expression:

$$\omega_{jht} = g(\omega_{jht-1}) + u_{jht}.$$

Plugging this into the production function

$$y_{jht} = \beta_{0,h} + \beta_{l,h} l_{jht} + \beta_{k,h} k_{jht} + g(\omega_{jht-1}) + \varepsilon_{jht} + u_{jht},$$

we will be able to obtain the following moment condition:

$$\begin{aligned} E(\varepsilon_{jht} + u_{jht} | \Omega_{jht-1}) = E(&y_{jht} - \beta_{0,h} - \beta_{l,h} l_{jht} - \alpha_{k,h} k_{jht} - \\ &g(\Phi_{t-1}(k_{jht-1}, l_{jht-1}, m_{jht-1}) - \alpha_0 - \alpha_l l_{jht-1} - \alpha_k k_{jht-1}) | \Omega_{t-1}) = 0. \end{aligned} \tag{33}$$

Because the estimation problem in (33) is non-linear in nature and to reduce the dimensionality, ACF propose a concentrated GMM procedure. Specifically, for a given guess of $\beta_{l,h}$ and $\beta_{k,h}$, we compute an estimate of $\beta_{0,h} + \omega_{jht}$ as:

$$\beta_{0,h} + \widehat{\omega_{jht}(\beta_{l,h}}, \beta_{k,h}) = \Phi_t(k_{jht}, l_{jht}, m_{jht}) - \beta_{l,h} l_{jht} - \beta_{k,h} k_{jht}.$$

We can then regress $\beta_{0,h} + \widehat{\omega_{jht}(\beta_{l,h}}, \beta_{k,h})$ on a non-parametric function of $\beta_{0,h} + \widehat{\omega_{jht-1}(\beta_{l,h}}, \beta_{k,h})$ and calculate the residual $\chi_t(\beta_{l,h}, \beta_{k,h})$. Finally, we can estimate the parameters of the production function using the following unconditional moment conditions:

$$\mathbb{E}\left(\chi_t(\beta_{l,h}, \beta_{k,h}) \otimes \begin{pmatrix} k_{jht} \\ l_{jht-1} \end{pmatrix}\right) = 0. \tag{34}$$

As discussed earlier in the main text, we modify the estimation procedure suggested by ACF to take into account (i.) potential finite sample problems and (ii.) the effect of aggregate shocks. Kim et al. (2019) show that the proposed estimation procedure of ACF suffers from a "spurious minimum" problem in that for some initial values, i.e., the resulting parameter estimates are a local minimum and not the global minimum.[22] Meanwhile, Hahn et al. (2020) show that in the presence of aggregate shocks, the parameters $\beta_{k,h}$ and $\beta_{l,h}$ are biased in that the function $g(\cdot)$ is not independent of time. Moreover, they show that the usual procedure of introducing time dummies does not remove the bias coming from the presence of the aggregate shock.

To deal with these concerns, we modify the proposed estimation procedure of ACF, which we outline below:

1. *Modified first stage.* Regress y_{jht} on $\tilde{\Phi}_t(k_{jht}, l_{jht}, m_{jht})$, which is the non-parametric function $\Phi_t(k_{jht}, l_{jht}, m_{jht})$ interacted with time dummies, following Hahn et al. (2020), via OLS:
$$y_{jht} = \tilde{\Phi}_t(k_{jht}, l_{jht}, m_{jht}) + \varepsilon_{jht}. \tag{35}$$

2. *Modified second stage.* As in Ackerberg et al. (2015), we proceed with a concentrated GMM procedure, wherein we concentrate out the parameter $\beta_{0,h}$ and the parameters of the polynomial function $g(\cdot)$ to minimize the dimensions of the non-linear search. In this regard, we proceed with the following estimation:

   (a) *Obtaining residuals* $\chi_t(\beta_{l,h}, \beta_{k,h})$. We compute an estimate of $\beta_{0,h} + \omega_{jht}$ from a given guess of $\beta_{l,h}$ and $\beta_{k,h}$ via the following equation
   $$\beta_{0,h} + \widehat{\omega_{jht}(\beta_{l,h}}, \beta_{k,h}) = \tilde{\Phi}_t(k_{jht}, l_{jht}, m_{jht}) - \beta_{l,h} l_{jht} - \beta_{k,h} k_{jht}.$$
   and regress $\beta_{0,h} + \widehat{\omega_{jht}(\beta_{l,h}}, \beta_{k,h})$ on a third-order polynomial of $\beta_{0,h} + \widehat{\omega_{jht-1}(\beta_{l,h}}, \beta_{k,h})$. From this estimation, we obtain residual $\chi_t(\beta_{l,h}, \beta_{k,h})$.

   (b) *Estimating $\beta_{l,h}$ and $\beta_{k,h}$.* Once we obtain the residual $\chi_t(\beta_{l,h}, \beta_{k,h})$, we estimate the parameters $\beta_{l,h}$ and $\beta_{k,h}$ with the following overidentified moment

[22] Moreover, in the case where the technology is constant returns to scale, it can be shown that there is a global identification problem in $\tilde{\beta}_k = \beta_l + \beta_k$ and $\tilde{\beta}_l = 0$.

conditions:

$$\mathbb{E}\left(\chi_t(\beta_{0,h},\beta_{l,h},\beta_{k,h}) \otimes \begin{pmatrix} 1 \\ k_{jht} \\ l_{jht-1} \\ l_{jht-2} \\ k_{jht-1} \end{pmatrix}\right) = 0. \tag{36}$$

Interacting the non-parametric function with the time dummies is sufficient to remove the aggregate shocks from the estimation of the production function parameters and recover the firm-specific productivity series, see Hahn et al. (2020), if the interest of the researcher is in these objects.[23] The moment conditions in (36) mirror the ones in Kim et al. (2019), who suggest the use of lagged instruments to help with the identification of the parameters of the production function. These moment conditions are also consistent with a model where firms face adjustment costs in capital and/or labor.

Given the moment conditions in equation (36), we utilize the CU-GMM procedure proposed in Peñaranda and Sentana (2015) to estimate the structural parameters $\beta_{l,h}$ and $\beta_{k,h}$. In particular, this procedure regresses 1s on the moment conditions with an OLS routine that is robust to singularities in the covariance matrix of the influence functions implied in (36).[24] We then obtain the standard errors of the parameter estimates via non-parametric and parametric bootstrap, with 500 replications.

C Estimating the non-linear productivity process

The following section provides a more formal description of the non-linear productivity process and a heuristic identification argument of productivity's persistent and transitory components.

[23]Of course, if we wanted to recover the aggregate shock as well, we would need to model the distribution of the aggregate shock more formally. However, as it is not of interest here, we pursue this procedure.

[24]In simulation exercises not reported here, we compare the performance of our proposed estimation procedure with those advocated by ACF using a modification of the data generating process in the published version of the paper. We find that our proposed modification works well in finite samples.

C.1 Identification argument - sketch

In this part of the appendix, we sketch the identification arguments for the non-linear productivity process. The starting point for the identification of the productivity process is the observation that the assumptions underlying Ackerberg et al. (2015) permit the identification of the production function parameters β_k and β_l. Given identification of these two parameters, we can then extract the productivity series z_{it} and assume that it is known, or rather, it is observed[25].

Given the assumptions we made for the persistent component η and the transitory component ε, we can apply the arguments in Arellano et al. (2017) for non-parametric identification. That is, assuming that the distributions of $(z_{jt}|z_{jt-1})$ and $(\eta_{jt}|z_{jt-1})$ satisfy completeness conditions, and given that in a model with three periods of log-productivity, (z_{j1}, z_{j2}, z_{j3}) are conditionally independent given η_{j2}, it follows that the marginal distributions of ε_{jt} are non-parametrically identified.[26] Serial independence of the ε then gives non-parametric identification of the joint distribution $(\varepsilon_{j2}, \varepsilon_{j3}, \dots, \varepsilon_{jT-1})$. As long as the characteristic functions of the transitory components do not vanish in the real line. The joint distribution $(\eta_{j2}, \eta_{j3}, \dots, \eta_{jT-1})$ are identified via a deconvolution argument. Subsequently, the distributions $f(\eta_{jt}|\eta_{jt-1})$ are identified for $t = 3, \dots, T-1$ and the marginal distribution of η_{j2} is identified. Hence, we would need $T = 4$ to identify at least one Markov transition.

C.2 Estimation algorithm

As it was described in the main text, the estimation algorithm that we pursue follows Arellano et al. (2017), who use a stochastic EM algorithm. In particular, for an initial guess $\widehat{\theta}^0$, we iterate on the following two steps on $s = 1, 2, \dots$ until convergence of the $\widehat{\theta}^s$ process:

[25]The estimation of the production function parameters is analogous to estimating the deterministic components of log-earnings in the case of the estimation of the earnings process.

[26]Completeness in this context is operator injectivity. For example, the first condition requires that the only function h that satisfies the equation $\mathbb{E}[h(z_{jt})|z_{jt-1}] = 0$ is $h = 0$. In effect, we just need that there is some statistical dependence between η_{jt} and η_{jt-1}, without specifying the form of dependence. Otherwise, we will be unable to distinguish the two components from each other.

1. *Stochastic E-step:* Draw observations of the persistent and transitory components $\eta_j^{(m)} = (\eta_j^{(m)}, \ldots, \eta_{jT}^{(m)})$ from the posterior distribution $f(\cdot; \widehat{\theta}^s)$

2. *M-step:* Compute

$$\widehat{\theta}^{s+1} = \sum_{j=1}^{N} \sum_{m=1}^{M} R(z_j, \eta_j^{(m)}; \theta) \tag{37}$$

wherein $R(\cdot)$ is a particular known objective function.

The E-step is straightforward, as the likelihood is available in closed form, see Arellano et al. (2017) for the complete form of the likelihood. In practice, we use a random-walk Metropolis-Hastings sampler with a target acceptance rate of 25 to 30 percent. To estimate the productivity process, we would need to compute a sequence of quantile regressions for the persistent component, its initial distribution, and the transitory component of productivity. In particular, to update the following set of parameters: $a_k^{\eta}(\tau_l)$ (the parameters of the persistent component), $a_k^{\eta 1}(\tau_l)$ (the parameters of the initial persistent component), and $a_1^{\varepsilon}(\tau_l)$ (the parameters of the transitory component), we estimate the following quantile regressions:

$$\min_{a_1^{\eta}(\tau_l),\ldots,a_K^{\eta}(\tau_l)} \sum_{j=1}^{N} \sum_{t=2}^{T} \sum_{m=1}^{M} \rho_{\tau l} \left(\eta_{jt}^{(m)} - \sum_{k=1}^{K} a_k^{\eta}(\tau_l) f_k(\eta_{jt-1}^{(m)}, age_{jt}) \right) \text{ for } l = 1, \ldots, L, \tag{38}$$

$$\min_{a_1^{\eta 1}(\tau_l),\ldots,a_K^{\eta 1}(\tau_l)} \sum_{j=1}^{N} \sum_{m=1}^{M} \rho_{\tau l} \left(\eta_{j1}^{(m)} - \sum_{k=1}^{K} a_k^{\eta 1}(\tau_l) f_k(age_{j1}) \right) \text{ for } l = 1, \ldots, L, \text{ and} \tag{39}$$

$$\min_{a_1^{\varepsilon}(\tau_l),\ldots,a_K^{\varepsilon}(\tau_l)} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \rho_{\tau l} \left(z_{jt} - \eta_{jt}^{(m)} - \sum_{k=1}^{K} a_k^{\varepsilon}(\tau_l) f_k(age_{jt}) \right) \text{ for } l = 1, \ldots, L, \tag{40}$$

in which $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \leq 0\})$ is the usual "check" function of quantile regression. In practice, we first estimate the age effect of productivity by a linear regression on a quartic polynomial in age. We then impose that ε_{it} is uncorrelated with age, although we allow for age effects in the variance and quantiles of ε_{it}. We take $M = 1$, stop the chain after a large number of iterations, and report an average across the last $\tilde{S}$ values, where $\tilde{S} = \frac{1}{\tilde{S}} \sum_{s=S-\tilde{S}+1}^{S} \widehat{\theta}^s$. The results for the productivity parameters are based on $S = 500$ iterations, with 200 Metropolis-Hastings draws per iteration, and we take $\tilde{S} = \frac{S}{2}$. We start the algorithm from different initial parameter values, and select the estimates that yield the highest log-likelihood. The non-selected values are similar to the ones that we report in the paper.

For inference, we perform both parametric and non-parametric bootstrap with 500 bootstrap replications.

D Estimating the canonical linear process

D.1 Identification

A more formal statement of the assumptions behind the canonical productivity process are the following:

1. $|\rho| < 1$.

2. $\eta_{jt} \perp \xi_{it} \perp \varepsilon_{jt}$.

3. $\eta_{jt} \sim iid\ N(0, \sigma_\eta^2)$, $\xi_{jt} \sim iid\ N(0, \sigma_\xi^2)$, $\varepsilon_{it} \sim iid\ N(0, \sigma_\varepsilon^2)$

Given these assumptions, we can formally identify the parameters of interest in this model from the autocovariance function alone, following standard arguments. Identification of the parameters requires four periods of data. The arguments for identification are reproduced below.

First, we can identify ρ from the slope:

$$\begin{aligned}
\frac{\text{Cov}(\xi_{j0}, \xi_{j3}) - \text{Cov}(\xi_{j0}, \xi_{j2})}{\text{Cov}(\xi_{j0}, \xi_{j2}) - \text{Cov}(\xi_{j0}, \xi_{j1})} &= \frac{\rho^3 \sigma_\xi^2 - \rho^2 \sigma_\xi^2}{\rho^2 \sigma_\xi^2 - \rho \sigma_\xi^2} \\
&= \frac{(\rho^3 - \rho^2)(\sigma_\xi^2)}{(\rho^2 - \rho)(\sigma_\xi^2)} = \rho.
\end{aligned}$$

The difference between the covariances allows us to obtain σ_ξ:

$$\begin{aligned}
\text{Cov}(\xi_{j0}, \xi_{j2}) - \text{Cov}(\xi_{j0}, \xi_{j1}) &= \rho^2 \sigma_\xi^2 - \rho \sigma_\xi^2 \\
&= (\rho^2 - \rho)(\sigma_\xi^2).
\end{aligned}$$

The difference between the variances allows us to obtain σ_η:

$$\begin{aligned}\mathrm{Var}(\xi_{j1}) - \mathrm{Var}(\xi_{j0}) &= (\rho\sigma_\xi^2 + \sigma_\eta^2 + \sigma_\varepsilon^2) - (\sigma_\xi^2 + \sigma_\varepsilon^2) \\ &= (\rho - 1)\sigma_\xi^2 + \sigma_\eta^2.\end{aligned}$$

Finally, the variance allows us to identify σ_ε:

$$\mathrm{Var}(\xi_{j0}) = \sigma_\xi^2 + \sigma_\varepsilon^2.$$

D.2 Estimation

The standard estimation strategy is to use minimum distance estimation, where the goal is to choose the parameters that minimize the distance between the empirical and theoretical moments. An alternative, which we implement here, is to estimate the parameters via pseudo-maximum likelihood estimation, following Arellano (2003). That is, if $u_j \sim \mathcal{N}(0, \Omega(\theta))$, then the pseudo maximum likelihood estimator of θ solves:

$$\hat{\theta}_{PML} = \arg\min_c \left\{ \log\det(\Omega(c)) + \frac{1}{N}\sum_{j=1}^{N} \hat{u}_j \Omega(c)^{-1} \hat{u}_j \right\}.$$

This is equivalent to:

$$\hat{\theta}_{PML} = \arg\min_c \left\{ \log\det(\Omega(c)) + \mathrm{tr}(\Omega(c)^{-1}\hat{\Omega}) \right\},$$

where **tr** is the trace of the resulting matrix, and $\hat{\Omega} = \sum \hat{u}_j' \hat{u}_j$. We can then use the asymptotic covariance matrix to compute the standard errors.

The assumptions on the stochastic productivity process imply

$$z_{jt} = \rho^{t-1}\eta_{j1} + \sum_{k=2}^{t} \rho^{t-k}\xi_{jk} + \varepsilon_{jt} \tag{41}$$

from which the following moments

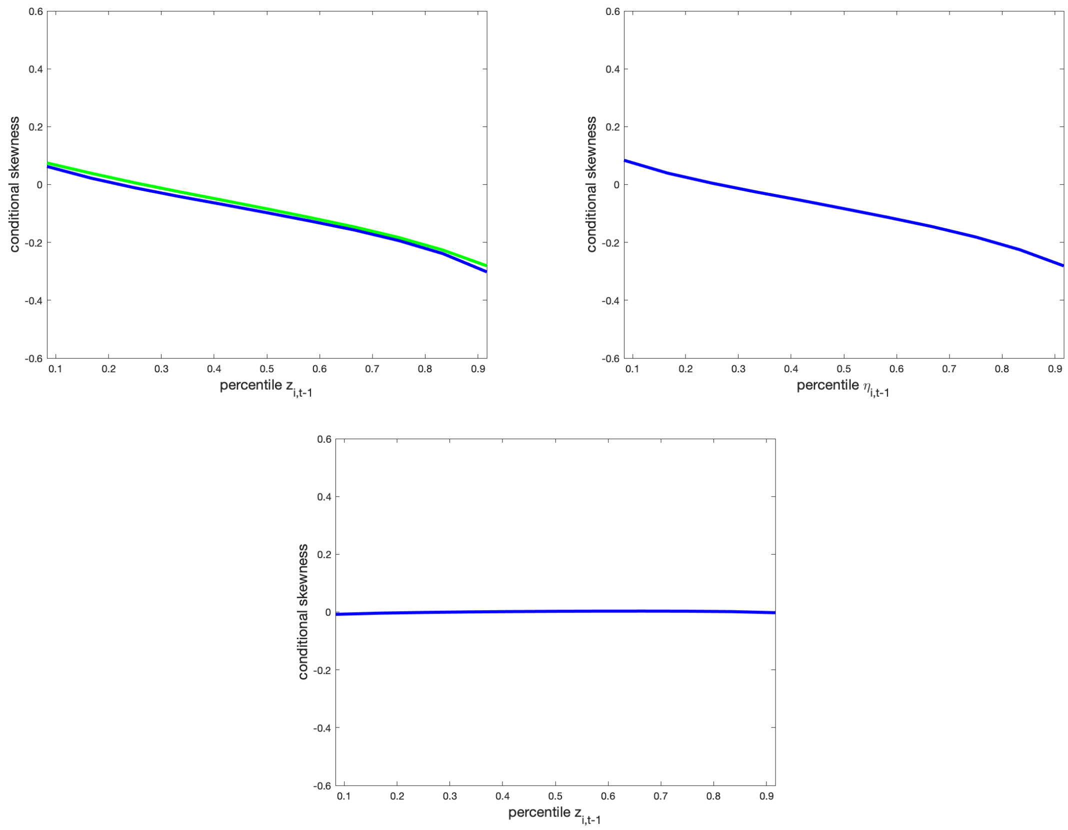
$$var(z_{jt}) = \rho^{2(t-1)}\sigma^2_{\eta 1} + \sum_{k=2}^{t} \rho^{2(t-k)}\sigma^2_{\xi} + \sigma^2_{\varepsilon}, \text{and} \tag{42}$$

$$cov(z_{jt}, z_{jt-1}) = \rho^{2t-1}\sigma^2_{\eta 1} + \sum_{k=2}^{t} \rho^{1+2(t-k)}\sigma^2_{\xi} \tag{43}$$

allow us to identify the parameters.

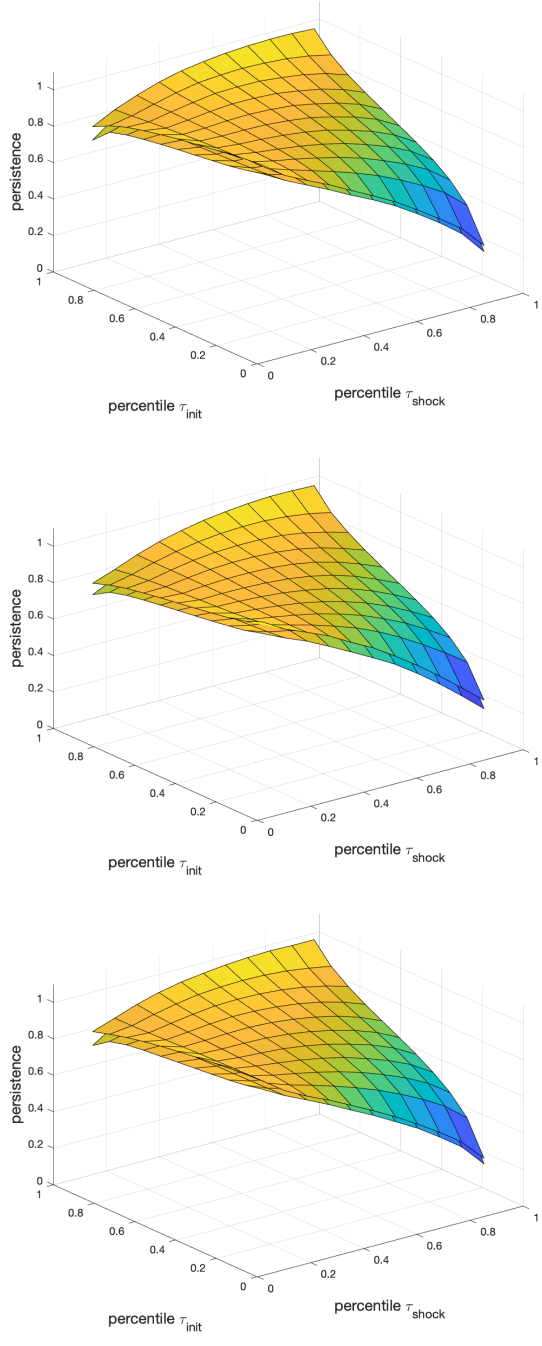
E Additional results for the productivity process

Figure 15: Conditional skewness, productivity process.



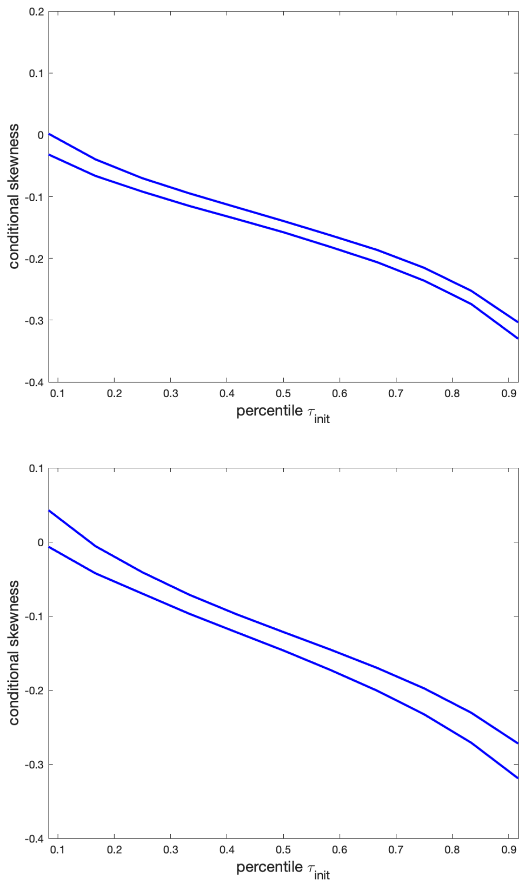
Note: These graphs provide the conditional skewness measures implied by the nonlinear productivity process and the canonical productivity process. Top left: Data (blue) and simulated data (green) from the model under the nonlinear productivity process. Top right: Conditional skewness of η, nonlinear productivity process. Bottom: Conditional skewness of η, canonical productivity process.

Figure 16: Graphs of nonlinear persistence, nonparametric bootstraps.



Note: These graphs provide the 95% confidence bands obtained from a nonparametric bootstrap for the nonlinear persistence measure as a function of the initial position in the productivity distribution and the shock. Top: Data. Middle: Simulated data from the model under the nonlinear productivity process. Bottom: Nonlinear persistence of η.

Figure 17: Graphs of conditional skewness, nonparametric bootstraps.



Note: These graphs provide the 95% confidence bands obtained from a nonparametric bootstrap for the conditional skewness measure as a function of the initial position in the productivity distribution. Top: Data. Bottom: Simulated data from the nonlinear productivity process.

F Empirical investment policy functions

F.1 Identification argument - sketch

To proceed with the identification of the investment policy function, we make the following assumptions.

For all $t \geq 1$:

1. u_{jt+s} and ε_{jt+s} are independent of $(k_{j1}, \ldots, k_{jt})$, $(\eta_{j1}, \ldots, \eta_{jt-1})$, and $(z_{j1}, \ldots, z_{jt-1})$. ε_{j1} is independent of k_{j1} and z_{j1}.

2. Conditional on $(k_{jt}, i_{jt}, \eta_{jt}, \varepsilon_{jt})$,capital at $t+1$ k_{jt+1} is independent of $(k_{j1}, \ldots, k_{jt-1})$, $(i_{j1}, \ldots, i_{jt-1})$, $(\eta_{j1}, \ldots, \eta_{jt-1})$, and $(z_{j1}, \ldots, z_{jt-1})$.

3. The cost shifter ν_{jt} in the investment policy function is independent of η_{j1}, $(u_{js}, \varepsilon_{js})$ for all $s \neq t$, and $(k_{j1}, \ldots, k_{jt})$.

The first part of this assumption requires that current and future productivity shocks to be independent of current and past firm capital. At the same time, we allow some dependence between the initial capital distribution and the persistent component of productivity. This is important because capital accumulation upon entry may be correlated to past productivity shocks through the firm's previous investment decisions. The second part of the assumption is a first-order Markov assumption on capital accumulation. This is satisfied by the law of motion for capital in the standard firm investment model. Finally, the third part of the assumption requires that cost shifters are serially independent, independent of the productivity components, and independent of past and future capital. In particular, this assumption rules out unobserved heterogeneity in firm investment.

First period. Identification of the investment policy function follows an induction argument. We begin with the first period investment policy function, which we can write as the following:

$$f(i_1|z, k_1) = \int f(i_1|k_1, \eta_1, z) f(\eta_1|k_1, z) d\eta_1. \tag{44}$$

Note that, if k_1 and η_1 are independent given z, an injectivity argument will provide identification of the first-period investment policy rule.[27] However, this conditional independence restriction is unlikely to hold when firms enter the sample. To show that this policy function is identified, then we would need to show identification of $f(k_1|z)$. Given the model assumptions, we can show that this can be written as:

$$f(k_1|z) = \int f(k_1|\eta_1)f(\eta_1|z)d\eta_1. \tag{45}$$

Because we can identify $f(\eta_1|z)$ from the productivity data alone, by completeness of this distribution, $f(k_1|\eta_1)$ is identified. Identification of this density implies that $f(\eta_1|k_1, z)$ is identified. Hence, under completeness in $(z_{j1}, \ldots, z_{jt})$ of $(\eta_{j1}|k_{j1}, z)$, $f(i_1|k_1, \eta_1, z)$ and $f(i_1, \eta_1|k_1, z)$ are thus identified.

Second period. We next proceed with the identification of second-period capital. Notice that we can write the density of $f(k_2|k_1, i_1, z, \eta_1)$ as:

$$f(k_2|k_1, i_1, z) = \int f(k_2|k_1, i_1, z, \eta_1)f(\eta_1|k_1, i_1, z)d\eta_1 \tag{46}$$

Provided that the distribution $(\eta_{j1}|k_{j1}, i_{j1}, z)$ is complete in z (notice that the density $f(\eta_1|k_1, i_1, z)$ is identified given the arguments above), the density $f(k_2|k_1, i_1, z, \eta_1)$ is non-parametrically identified. Moreover, we can show that:

$$f(\eta_2|k_1, z, k_2, i_1) = \int f(\eta_2|\eta_1, k_1, z, k_2, i_1)f(\eta_1|k_1, z, k_2, i_1)d\eta_1 \tag{47}$$

and because we can identify $f(\eta_1|k_1, z, k_2, i_1)$ from the previous argument, under completeness of this density on $(z_{j1}, \ldots, z_{jt})$, we can show that $f(\eta_2|\eta_1, k_1, z, k_2, i_1)$ is identified. Finally, we can then show that:

$$f(i_2|k_2, k_1, i_1, z) = \int f(i_2|k_2, \eta_2, z)f(\eta_2|k_1, z, k_2, i_1)d\eta_2 \tag{48}$$

[27] To see this, it suffices to show that because of independence of η_1 and k_1, the density above collapses to $f(\eta_1|z)$, which we can identify from productivity alone. Hence, by operator injectivity, the first-period investment rule is non-parametrically identified.

is non-parametrically identified.

Finally, through an induction argument, and given the first-order Markov assumptions, we can show identification from periods three onwards using the same arguments. As has been argued in Arellano et al. (2017), one can think of the identification arguments through their link to non-parametric IV problems. For example, in the investment policy function, the endogenous regressor is η_{j1} and conditional on i_{j1} and k_{j1}, the z's can be thought of as instruments for the endogenous regressor. In subsequent periods, together with the leads and lags of log productivity, we use lags of capital and investment as instruments in a non-parametric sense.

F.2 Model restrictions

The empirical policy function of investment implies the following model restrictions:

$$a_0^I(\tau_l), \ldots, a_K^I(\tau_l) = \underset{a_0^I(\tau_l),\ldots,a_K^I(\tau_l)}{\arg\min} \sum_{t=1}^{T} \mathbb{E}\Bigg[\int \rho_{\tau_l}\Bigg(i_{jt+1} - \sum_{k=0}^{K} a_k^I(\tau_l) f_k(k_{jt}, \eta_{jt}, \varepsilon_{jt}, age_{jt+1})\Bigg) \\ \times g_j(\eta_j^T; \bar{\theta}, \bar{\mu}) d\eta_i^T\Bigg] \tag{49}$$

where $g_j(\eta_j^T; \bar{\theta}, \bar{\mu}) = f(\eta_j^T | i_j^T, k_j^T, z_j^T; \theta, \mu)$ denotes the posterior density of the persistent component of productivity given investment, capital, and productivity data.

The tail parameters satisfy the following model restrictions (where we suppress the conditioning on age for conciseness):

$$\tau_- = \frac{\sum_{t=1}^{T} \mathbb{E}\Big[\int \mathbb{1}\Big\{i_{jt+1} \leq \sum_{k=0}^{K} a_{k1}^I f_k(\cdot)\Big\} g_j(\eta_j^T; \cdot) d\eta_i^T\Big]}{\sum_{t=1}^{T} \mathbb{E}\Big[\int \Big(i_{jt+1} - \sum_{k=0}^{K} a_{k1}^I f_k(\cdot)\Big) \mathbb{1}\Big\{i_{jt+1} \leq \sum_{k=0}^{K} a_{k1}^I f_k(\cdot)\Big\} g_j(\eta_j^T; \cdot) d\eta_i^T\Big]}$$

and

$$\tau_+ = \frac{\sum_{t=1}^{T} \mathbb{E}\Big[\int \mathbb{1}\Big\{i_{jt+1} \geq \sum_{k=0}^{K} a_{kL}^I f_k(\cdot)\Big\} g_j(\eta_j^T; \cdot) d\eta_i^T\Big]}{\sum_{t=1}^{T} \mathbb{E}\Big[\int \Big(i_{jt+1} - \sum_{k=0}^{K} a_{kL}^I f_k(\cdot)\Big) \mathbb{1}\Big\{i_{jt+1} \geq \sum_{k=0}^{K} a_{kL}^I f_k(\cdot)\Big\} g_j(\eta_j^T; \cdot) d\eta_i^T\Big]}$$

where $f_k(\cdot) = f_k(k_{jt}, \eta_{jt}, \varepsilon_{jt})$.

F.3 Likelihood function

The likelihood function, omitting the conditioning on firm age for conciseness, is:

$$f(i_{jt+1}, k_{jt}, \eta_{jt}, \varepsilon_{jt}; \theta, \mu) = \prod_{t=1}^{T} f(i_{jt+1}|k_{jt}, \eta_{jt}, \varepsilon_{jt}; \mu) \prod_{t=1}^{t} f(z_{jt}|\eta_{jt}; \theta) \prod_{t=2}^{T} f(\eta_{jt}|\eta_{jt-1}; \theta) f(\eta_{j1}; \theta).$$

The likelihood function can be written in closed form, using similar techniques as in Arellano et al. (2017).

F.4 Estimation algorithm

Start at $\widehat{\mu}^{(0)}$. Iterate on $s = 1, 2, \ldots$ the following two steps:

1. *Stochastic E step:* Draw $\eta^{(m)} = (\eta_{j1}^{(m)}, \ldots, \eta_{jT}^{(m)})$ from the posterior density of the distribution of η given the data. The complete form of the likelihood for the posterior density is the following:

$$f(\eta_j^T | i_{jt+1}, k_{jt}, \varepsilon_{jt}; \widehat{\theta}, \mu^{(s)}) \propto \prod_{t=1}^{T} f(i_{jt+1}|k_{jt}, \eta_{jt}, \varepsilon_{jt}) f(z_{jt}|\eta_{jt}) \times \prod_{t=2}^{T} f(\eta_{jt}|\eta_{jt-1}) f(\eta_{j1}) \tag{50}$$

2. *M step:* Compute, for $l = 1, \ldots, L$:

$$\min_{a_0^I(\tau_l), \ldots, a_K^I(\tau_l)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \rho_{\tau_l} \left(i_{jt+1} - \sum_{k=0}^{K} a_k^I(\tau_l) f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}, age_{it+1}) - \mathbf{X}'_{jt}\beta(\tau_l) \right) \tag{51}$$

and for the tail parameters:

$$\tau_- = \frac{\sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{m=1}^{M} \mathbb{1} \left\{ i_{jt+1} \leq \sum_{k=0}^{K} a_{k1}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}) \right\}}{\sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{m=1}^{M} (i_{jt+1} - \sum_{k=0}^{K} a_{k1}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)})) \mathbb{1} \left\{ i_{jt+1} \leq \sum_{k=0}^{K} a_{k1}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}) \right\}}$$

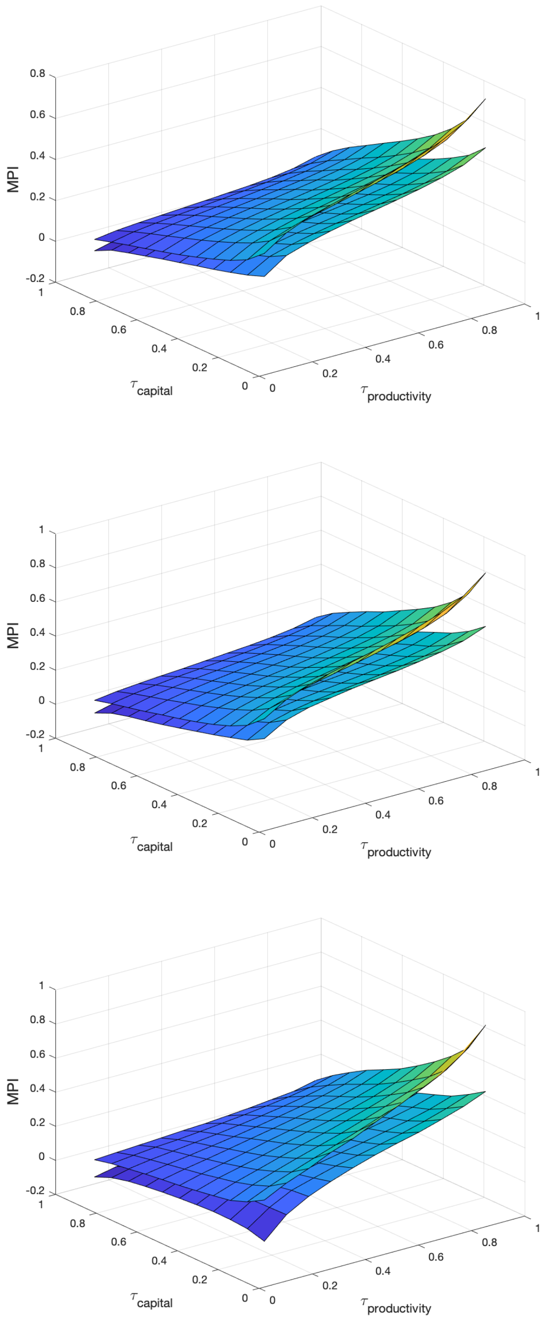
and

$$\tau_+ = \frac{\sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{m=1}^{M} \mathbb{1} \left\{ i_{jt+1} \geq \sum_{k=0}^{K} a_{kL}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}) \right\}}{\sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{m=1}^{M} (i_{jt+1} - \sum_{k=0}^{K} a_{kL}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)})) \mathbb{1} \left\{ i_{jt+1} \geq \sum_{k=0}^{K} a_{kL}^I f_k(k_{jt}, \eta_{jt}^{(m)}, \varepsilon_{jt}^{(m)}) \right\}}$$

In practice, we start the algorithm with different choices for $\widehat{\mu}^{(0)}$. For example, we estimate the investment rate on capital, productivity and age via quantile regressions. We proceed similarly to set other starting parameter values. We experimented with several other choices and selected the parameter values corresponding to the highest average log-likelihood over iterations. We observed some effect of starting values on estimates of tail parameters, although most results were stable.

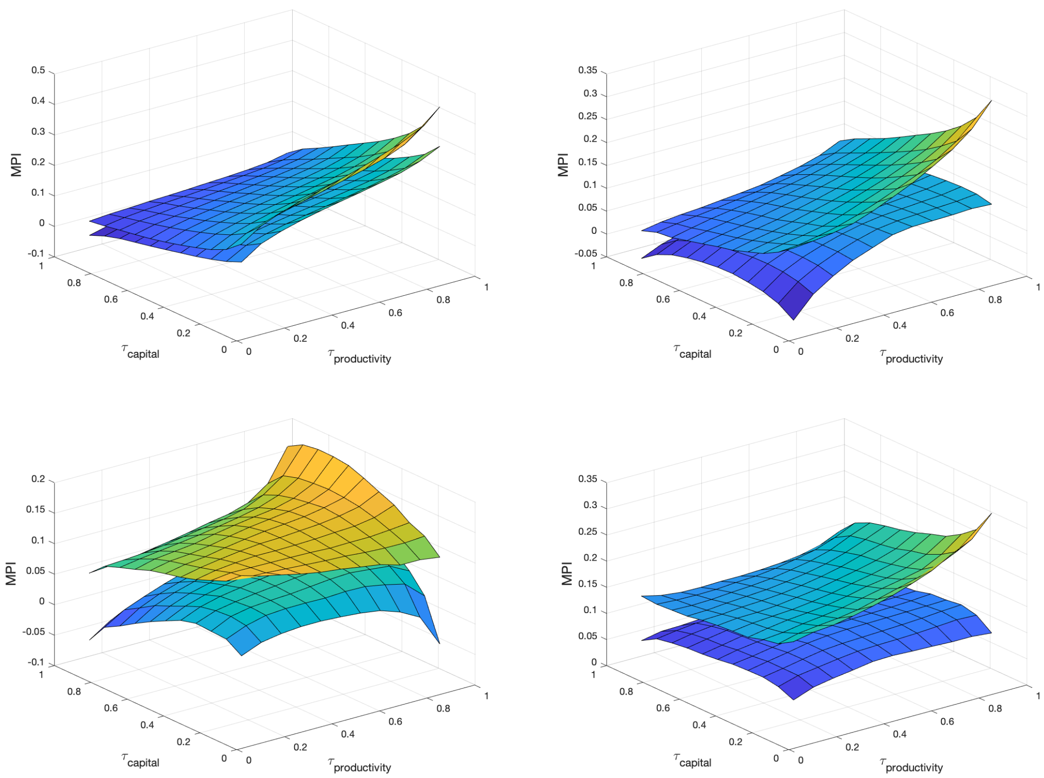
G Additional results on empirical investment policy functions

Figure 18: Graphs of the MPI with respect to productivity, nonparametric bootstraps.



Note: These graphs provide the nonparametric bootstrap for the marginal propensities to invest with respect to productivity z_{jt}. Top: Data. Middle: Simulated data from the model under the nonlinear productivity process. Bottom: Simulated data from the model under the canonical productivity process.

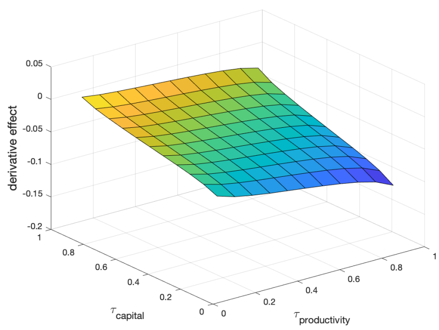
Figure 19: Graphs of the MPI with respect to persistent and transitory productivities, nonparametric bootstraps.



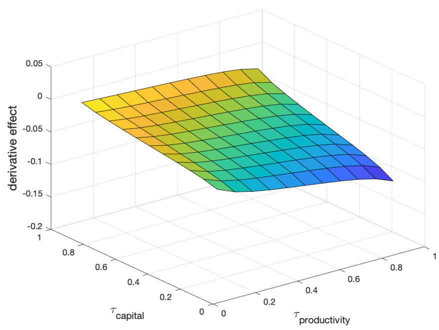
Note: These graphs provide the non-parametric bootstrap for the marginal propensities to invest with respect to persistent productivity η_{jt} and transitory productivity ε_{jt}. Top left: η_{jt}, nonlinear process. Top right: ε_{jt}, nonlinear process. Bottom left: η_{jt}, canonical process. Bottom right: ε_{jt}, canonical process.

Figure 20: Marginal propensities to invest with respect to capital

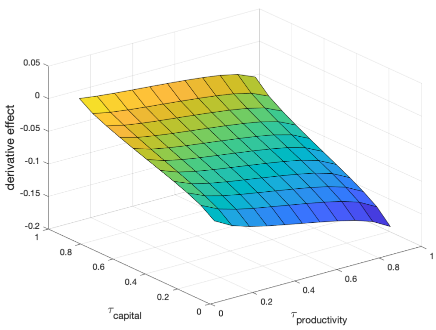
(a) Response to capital, data



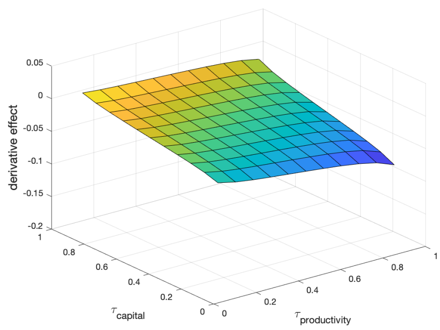
(b) Response to capital, simulated model



(c) Response to capital, non-linear model

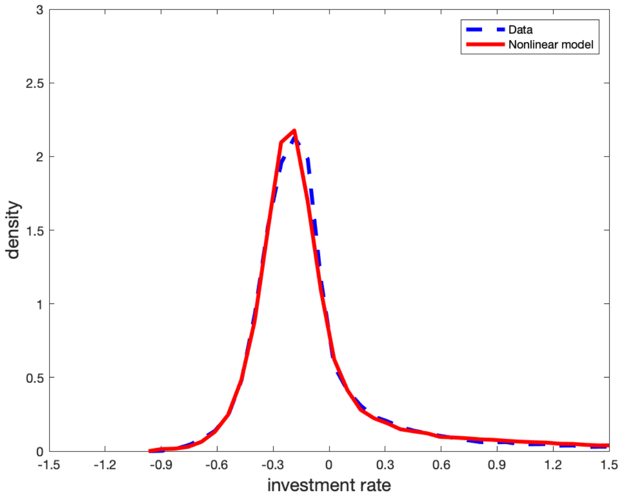


(d) Response to capital, canonical model



Note: The above figures show the average derivative effect of capital on firm investment, evaluated at different percentiles of capital and productivity, and averaged across age. Top left: investment responses from the data. Top right: investment responses to capital based on simulated data from the nonlinear model. Bottom left: investment responses to capital, non-linear model. Bottom left: investment responses to capital, canonical model.

Figure 21: Predicted vs. observed distribution of investment rates



Note: The above figure shows the predicted distribution of investment rates implied by the nonlinear model of firm investment and the observed distribution of investment rates from the data.

H Additional results on the model

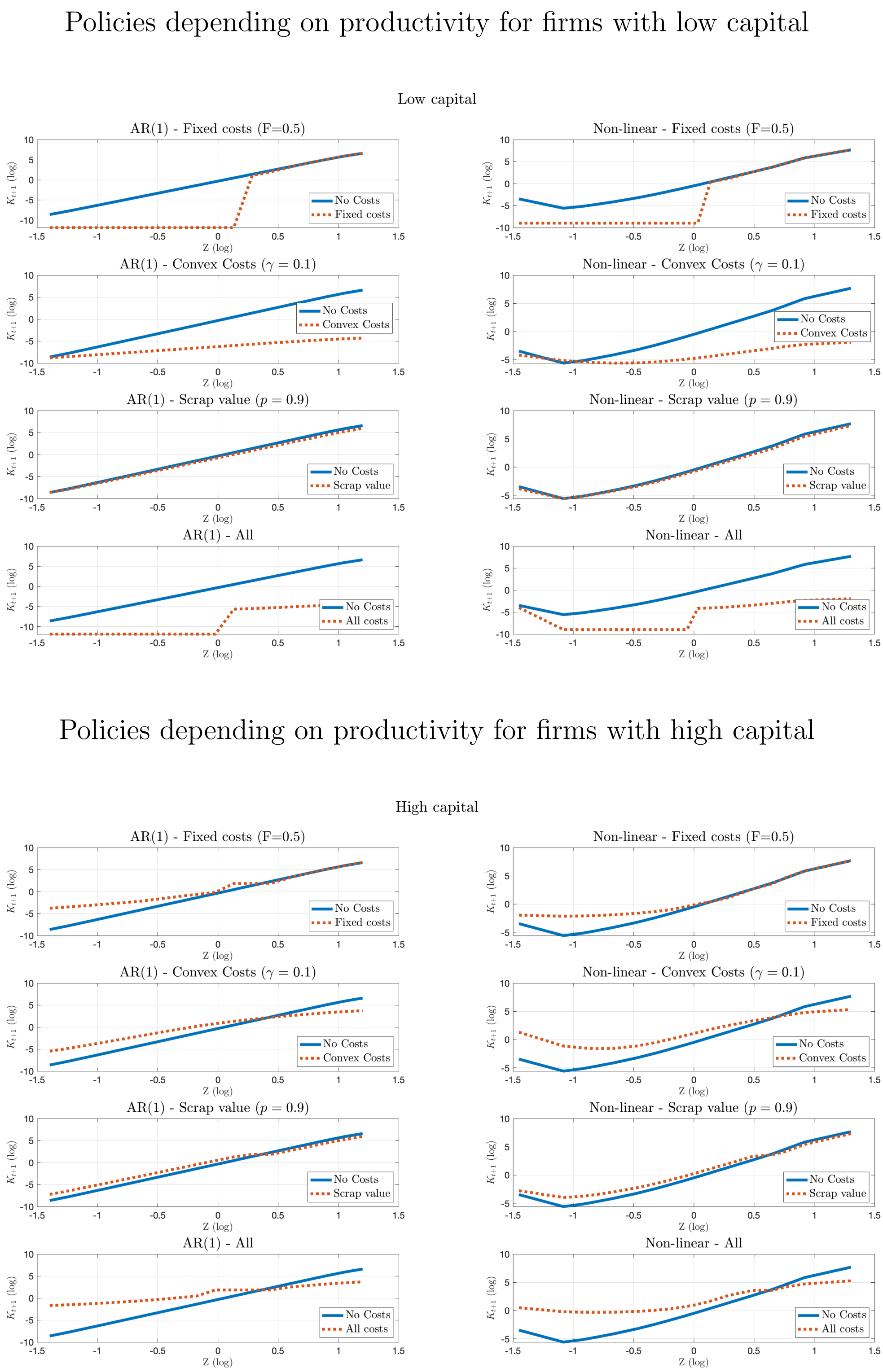


Figure 22: Policy functions in the model with no costs (blue) and with capital adjustment costs (orange) when firms' productivity follows an AR(1) process (left panels) and when firms' productivity follows the non-linear process (right panel).