Misallocation in a Model with Lumpy Investment*

Francesco Lippi
Luiss University and EIEF
francescolippi@gmail.com

Aleksei Oskolkov
Princeton University
alekseioskolkov@princeton.edu

February 12, 2026

Abstract

This paper studies capital misallocation in a tractable model with random fixed costs of adjustment. We identify the distribution of fixed costs and productivity shocks using the entire size distribution and frequency of investments and provide an efficient estimation method. We derive the measure of capital misallocation in the presence of fixed costs and show that it differs from the traditional metric based on the variance of marginal product of capital. The key feature of models with lumpy investments responsible for this is their non-linearity: the distribution of marginal product of capital is not log-normal even with normal shocks and non-degenerate even when shocks are small. We apply our method to 40 years of panel data on Italian firms and find misallocation costs about 0.5-2% of output. Fixed costs contribute about one half to traditional measures of TFPR dispersion, putting an upper bound on potential inefficiencies.

Key Words: Misallocation, Lumpy investment, Generalized hazard function

JEL Classification Numbers: D22, D25, E20, E22

*We benefited from discussions with Andrey Alexandrov, Fernando Alvarez, Isaac Bailey, Andres Blanco, Luca Gagliardone, Mikhail Golosov, John Haltiwanger, Hugo Hopenhayn, Ilse Lindenlaub, Guido Moscarini, Michael Peters, Richard Rogerson, and Robert Shimer. Artyom Lipin provided excellent research assistance. Lippi acknowledges financial support from the ERC grant: 101054421-DCS.

1 Introduction

Over the past decades, a voluminous literature has investigated the effect of frictions on the allocation of production factors, see e.g. Hopenhayn and Rogerson (1993). Inputs should ideally be allocated across firms to equalize their marginal revenue product. Misallocation arises when frictions, either primitive or related to market failures, prevents such equalization. Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) pioneered the modern strand of this literature using theory to map firm-level data on production and factor use to a measure of misallocation.

Restuccia and Rogerson (2017) classify studies of misallocation in two categories, "direct" and "indirect". The indirect approach measures dispersion of firm-level marginal product of capital, attributing such deviations to wedges that capture unmodeled frictions. The direct approach, instead, models a source of misallocation explicitly and identifies it in a structural model. Among the sources of distortions discussed in the literature are financial frictions, like in Edmond, Midrigan, and Xu (2015), capital adjustment costs, like in Cooper and Haltiwanger (2006), Bond and Van Reenen (2007), Asker, Collard-Wexler, and Loecker (2014), and information frictions, like in David, Hopenhayn, and Venkateswaran (2016) and David and Venkateswaran (2019). Importantly, capital misallocation in this terminology is not equivalent to inefficiency, since frictions are a primitive of the environment, and equilibria are mostly efficient given these primitives.

Our paper follows the direct approach and explores capital misallocation using a structural investment model with non-convex adjustment costs, as developed by Caballero and Engel (1999). The model describes firms' behavior by means of a "generalized hazard function" that encodes a smooth version of an sS policy and reproduces the empirical distribution of capital investments. Unlike simple fixed-cost environments, this setup generates a probabilistic decision rule: firms draw random adjustment costs and invest if the draw is sufficiently small.

Our analysis is motivated by firm-level data on investment: investments are infrequent and sizable.[1] The literature on capital misallocation has often focused on models with convex adjust-

[1] See Doms and Dunne (1998); Abel and Eberly (1994); Caballero et al. (1995) for an early documentation of the lumpy nature of investment and the ability of models with non-convex adjustment cost to account for it.

ment costs. These models conveniently summarize aggregate misallocation with the cross-sectional variance of the marginal product of capital, but they also predict that investments occur every period and are "small", which is at odds with firm-level data. Lumpy investment models challenge the logic behind this mapping from firm-level dispersion of productivity to misallocation. The mapping relies on one of two features: either firms' productivity shocks and "wedges" are jointly log-normal, or the size of the shocks and distortions are "small", and the mapping works as a second-order approximation.[2] In models with lumpy investment, however, the distribution of the firms' marginal product of capital is not log-normal, even if underlying shocks are normally distributed. Moreover, this non-normality survives even in the limit of vanishing shocks.

We provide two analytic results and several empirical applications. The first result gives an efficient procedure to identify the primitives of a Caballero and Engel (1999) model, where firm behavior is described by a generalized hazard function. Caballero and Engel (1999) map primitives, such as the distribution of the fixed costs and other structural parameters, to the distribution of observable investment sizes. We work in the setup developed by Baley and Blanco (2021) and solve the inverse problem: starting from the observed distribution of the size of investments, we recover the primitives that generate the data, in particular the distribution of the adjustment costs.

Second, we derive aggregate misallocation in the presence of lumpy investment. The formula combines moments of the distribution of marginal product of capital that amount to a generalized Jensen correction. This measure of dispersion generalizes variance and is related to "entropy", a well-known functional in the asset pricing literature. If marginal product of capital were log-normally distributed, the measure would collapse to the standard variance metric. We show that the distribution is non-degenerate and not log-normal even in the limit of small shocks, so the variance-based metric is not accurate. We illustrate the size of the approximation error due to using the variance-based metric analytically in the special case of a constant hazard function. We also follow Baley and Blanco (2021) and characterize the aggregate capital-to-productivity ratio.

[2] For instance, David, Hopenhayn, and Venkateswaran (2016) consider a model where firms face imperfect information and assume that firm-level fundamentals and idiosyncratic distortions are jointly log-normal. David and Venkateswaran (2019) consider a model with convex adjustment costs, in addition to imperfect information, and firm-specific "wedges" and characterize misallocation using an approximation with small shocks.

In the presence of fixed costs of investment, this ratio is not approximated by the average marginal product of capital, which is the case in models with convex adjustment costs or no frictions at all.

We elaborate on this result in two ways. First, we show that the variance-based measure of misallocation can be salvaged in the double limit where adjustment frictions vanish together with shocks. The relative rate of decay of adjustment frictions and shocks determines the approximation error for the variance-based metric of misallocation and the average-based metric of the aggregate capital-productivity ratio. Second, we adapt the model of David and Venkateswaran (2019) to the Baley and Blanco (2021) setup as a convex-cost benchmark and show that the variance metric works well in that environment. We explain how non-linearities generated by non-convex adjustment costs cause the breakdown of log-normality and, consequently, the variance result.

We estimate the model using a large panel of capital investment data by Italian manufacturing firms, from nine industries, over a 40-year period. To do that, we develop an efficient approximate method to reduce the dimensionality of the problem. The theory maps one infinitely-dimensional object (the distribution of investments) to another one (the distribution of fixed costs), using differential equations. Our method turns this problem into a system of linear equations instead.

We then present four quantitative applications. First, we quantify capital misallocation over the whole sample period. Output losses from adjustment frictions add up to around 1% of output. This lands on the lower end of the spectrum of estimates in the literature, closer to structural models like David and Venkateswaran (2019), who find misallocation costs of $4-5\%$ of output, than to reduced-form results in the literature, which can go up to 50%. Comparing our measure of capital misallocation with the variance-based measure, we find that the bias of the latter is significant but not too large, about 10%. Second, we split the sample into subperiods, 1983-2003 and 2003-2023, and track the time evolution of misallocation. We find a decline of at least 30%, across all sectors, largely as a result of a fall in productivity growth. Intuitively, if the dynamics of the ideal capital stock are attenuated, firms are closer to their desired capital on average.[3]

Third, we revisit a question posed by Asker et al. (2014) and explore how much of measured

[3] This effect is akin to the price wedges caused by inflation in New Keynesian models, see e.g., Nakamura et al. (2018); Adam et al. (2024); Miyahara et al. (2026).

misallocation is attributed to fixed costs as opposed to reduced-form wedges, a stand-in for dynamic inefficiency. We use traditional methods à la Hsieh and Klenow (2009) to estimate the dispersion of total factor productivity of revenues (TFPR), and compare this metric with our model-based measure of misallocation. Our misallocation metric captures a substantial share of the TFPR dispersion, about one half, suggesting that a large part of traditionally measured misallocation is due to fixed costs as opposed to dynamic inefficiencies, a finding that is consistent with the results of Asker et al. (2014).

Finally, we quantify the importance of using a fully fledged generalized hazard model against the alternative of using a simple two-sided Calvo model with different arrival rates of positive and negative adjustment opportunities, a simplified version of Baley and Blanco (2021). We show that the two-sided model tends to substantially overestimate misallocation and develop an econometric test that rejects the simple two-side hazard function for most industries.

The paper is organized is follows. Section 2 sets up the model, Section 3 lays out our results on misallocation, Section 4 discusses identification, Section 5 describes the estimation results, and Section 6 presents our quantitative applications. We continuously engage with the literature throughout the paper and place a detailed review in Appendix A.

2 Setup

There is only good good in this economy, used for both investment and consumption, with its price normalized to one. Firms use capital and labor to produce: k_t units of capital and l_t units of labor combine into $f_t = \hat{z}_t^{(1-\alpha)\zeta} k_t^{\alpha\zeta} l_t^{1-\zeta}$ units of final good. Here $\hat{z}_t$ is productivity. Firms face no frictions in choosing labor and solve the following value-added maximization problem at every t:

$$\hat{F}(\hat{z}_t, k_t) = \max_{l_t} \ \hat{z}_t^{(1-\alpha)\zeta} k_t^{\alpha\zeta} l_t^{1-\zeta} - w_t l_t$$

After maximization, their value added is given by $F(z_t, k_t) = z_t^{1-\alpha} k_t^{\alpha}$, where

$$z_t \equiv \hat{z}_t \cdot \left(\frac{\zeta^\zeta (1-\zeta)^{1-\zeta}}{w_t^{1-\zeta}} \right)^{\frac{1}{\zeta(1-\alpha)}}$$

From now on, we will work with value added after the optimal choice of labor. Our notion of productivity subsumes wages. Its law of motion is $d\log(z_t) = \mu dt + \sigma dW_t$. When uncontrolled, capital evolves according to $d\log(k_t) = -\delta dt$. As in Caballero and Engel (1999) firms face random opportunities to adjust capital stock. When an ivestment opportunity arrives, it can be taken at a fixed cost ψz_t. With a Poisson intensity γ_d, firms get an opportunity to adjust down and draw an adjustment cost ψ. This cost is distributed with a cumulative distribution function $G_d(\cdot)$ on interval $[0, \psi_d]$. Analogously, an opportunity for adjusting up arrives with a Poisson intensity γ_u. Upward adjustment costs are distributed according to $G_u(\cdot)$ on interval $[0, \psi_u]$.[4]

Policy. We conjecture the following policy. Conditional on adjusting, firms always choose $y^* z_t$ as their new level of capital. If $k_t > y^* z_t$ and an opportunity arrives to adjust down, the firms do it if the cost draw ψ is low enough, satisfying $\psi \leq \boldsymbol{\psi}_d(k_t/z_t)$. If $k_t < y^* z_t$ and an opportunity arrives to adjust up, the firms do it if $\psi \leq \boldsymbol{\psi}_u(k_t/z_t)$. Here $\boldsymbol{\psi}_d(\cdot)$ and $\boldsymbol{\psi}_u(\cdot)$ are the downward and upward adjustment cutoff functions.

We guess and verify that the the capital-to-productivity ratio $y \equiv k/z$ fully determines the firm's value and its adjustment decisions (see Appendix C for details). Specifically, the downward adjustment cutoff function $\boldsymbol{\psi}_d(\cdot)$ maps $[y^*, \infty)$ to $[0, \psi_d]$. The upward adjustment cutoff function $\boldsymbol{\psi}_u(\cdot)$ maps $[0, y^*]$ to $[0, \psi_u]$. Let $i \in \{u, d\}$ be a binary indicator that determines where the firm is relative to the optimal point. If $i = u$, the firm is below the optimal capital and would like to adjust upwards. If $i = d$, it wishes to disinvest. Denoting $\rho \equiv r - \mu - \sigma^2/2$ and $\nu \equiv r + \delta$, the Hamilton-Jacobi-Bellman equation in a stationary environment is

$$\rho v(y) = \underbrace{y^\alpha - \nu y}_{\text{net revenue}} + (\rho - \nu) y v'(y) + \frac{\sigma^2}{2} y^2 v''(y) + \underbrace{\sum_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\} dG_i(\psi)}_{\text{adjustment option value}} \quad (1)$$

Net revenue $y^\alpha - \nu y$ is net of depreciation δ and opportunity cost of holding capital r. In the option value term, $\mathbb{1}_i$ with $i \in \{u, d\}$ indicates that the firm is in the region where it wants to adjust up or down. The optimality condition for the choice of y^* is $v'(y^*) = 0$. The cutoff functions $\boldsymbol{\psi}_i(\cdot)$ are given by $\boldsymbol{\psi}_i(y) = v(y^*) - v(y)$.

[4] In Appendix C, we show that this setup is equivalent to one where firms rent capital instead of owning.

Define the generalized hazard function as the intensity of adjustment given the state y:

$$\lambda(y) = \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i(\boldsymbol{\psi}_i(y)) \tag{2}$$

This object encodes the fundamental decisions of the firm and determines the steady-state distribution of capital-to-productivity ratio y, which is key for computing total output and other aggregates. We also use is it as a bridge between the model primitives and observable objects such as the capital adjustments recorded in the balance sheet. In Section 4, we show how to recover $\lambda(\cdot)$ from the data. The remainder of the present section focuses on using a given $\lambda(\cdot)$ together with the structure of the firm's problem, to recover (μ, σ) and $(\gamma_i, G_i(\cdot))_{i=u,d}$.[5]

Recovering the primitives. Equation (1) and equation (2) establish how the primitives of the model $(r, \alpha, \delta, \mu, \sigma)$ and $(\kappa_i, G_i(\cdot))_{i=u,d}$ determine the generalized hazard function $\lambda(\cdot)$. Obtaining the inverse mapping is challenging: equation (1) is non-linear, and $\lambda(\cdot)$ does not enter it directly. To deal with this, we differentiate equation (1) with respect to y and work with the marginal value $v'(y)$ instead of $v(y)$. We also change the variables to $y \mapsto x(y) \equiv \log(y)$ for convenience. To place our results closer to the literature, in particular David et al. (2016) and David and Venkateswaran (2019), we refer to the variable x as log-ARPK since it is proportional to the average revenue product of capital in logs: $ARPK_t \equiv F(z_t, k_t)/k_t = (z_t/k_t)^{1-\alpha}$, and $\log ARPK_t = (\alpha - 1)x_t$. We will work with the marginal value function $U(x(y)) \equiv v'(y)$ and with the generalized hazard function $\Lambda(x(y)) \equiv \lambda(y)$ defined over log ratios. For these functions, we establish the following:

Proposition 1. *The functions $U(\cdot)$ and $\Lambda(\cdot)$ solve the following system:*

$$(\nu + \Lambda(x))U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu + \delta)U'(x) + \frac{\sigma^2}{2}U''(x) \tag{3}$$

$$\Lambda(x) = \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i \left(\int_x^{x^*} U(s) e^s \, ds \right) \tag{4}$$

where x^* is defined as the unique root of $U(\cdot)$: $U(x^*) = 0$.

Given $\Lambda(\cdot)$, equation (3) is a linear differential equation for $U(\cdot)$. With $\Lambda(\cdot)$ on hand and suitable boundary conditions that we describe in the proof, solving for $U(\cdot)$ numerically is feasible.

[5] Of course, $\lambda(\cdot)$ cannot be read from the data directly since productivity z that enters the underlying capital-to-productivity ratios $y \equiv k/z$ is not observable. Section 4 deals with this problem in detail.

In some cases, the solution is essentially analytical. Appendix D shows a closed-form solution for equation (3) under the assumption that the distributions $\{G_u(\cdot), G_d(\cdot)\}$ are collections of mass points, making $\Lambda(\cdot)$ piece-wise constant. We use this case in our empirical application in Section 5.

3 Aggregation and the steady state

Aggregation of firm decisions hinges on the generalized hazard function $\Lambda(\cdot)$. We will consider quasi-stationary environments in which this function and the distribution of log-ARPK is fixed over time. This does not mean that the economy as a whole is stationary: productivity z diverges over time because it follows a geometric Brownian motion, and capital stock k keeps up with productivity on average. Despite this, our assumed production function and the fact that fixed costs are proportional to z guarantee that the log-ARPK has a stationary distribution. The generalized hazard function $\Lambda(\cdot)$ is a key object that determines it.

Let $\phi(\cdot)$ denote the ergodic density of x, where the uncontrolled x follows a diffusion with drift $-(\mu+\delta)$ and volatility σ. If $\Lambda(\cdot)$ is the generalized hazard function for investment, in a steady state $\phi(\cdot)$ and $\Lambda(\cdot)$ satisfy a Kolmogorov forward equation

$$0 = (\mu+\delta)\phi'(x) + \frac{\sigma^2}{2}\phi''(x) - \phi(x)\Lambda(x) \tag{5}$$

The first two terms reflect the diffusion of individual firms across the distribution between adjustments. Log-ARPK change because the underlying productivity changes with a drift μ and a volatility σ, and capital depreciates at a rate δ. The last term reflects investment happening with a state-dependent intensity $\Lambda(x)$: upon investing, a firm disappears from the set of firms with log-ARPK equal to $(\alpha-1)x$ and joins those with a log-ARPK equal to $(\alpha-1)x^*$. At x^* itself, equation (5) does not hold: firms arrive at this point continually, so $\phi(\cdot)$ has a kink there.

We show how to identify $\phi(\cdot)$ and $\Lambda(\cdot)$ from the data in Section 4. In the remainder of this section we study capital misallocation treating $\phi(\cdot)$ and $\Lambda(\cdot)$ as known objects.

3.1 Measuring misallocation

We focus on a static notion of misallocation. Given the joint distribution of capital and productivity, we define the maximum level of output potentially achievable through reallocation of capital between firms. We call misallocation the (log) difference between this maximum level and actual output. This notion of misallocation holds aggregate capital fixed and only describes the costs of physical frictions as opposed to potential market failures.

Formally, let $g(k, z)$ be the density of firms with capital k and productivity z. This joint density does not converge to a stationary distribution, so we make $g(\cdot)$ an argument of aggregate statistics shown below. Aggregate output Y of an economy with a joint distribution g is

$$Y(g) = \int \int z^{1-\alpha} k^{\alpha} g(k, z) dk dz$$

The maximum output is

$$\begin{aligned} \hat{Y}(g) = \max_{\hat{k}(\cdot)} & \int \int z^{1-\alpha} \hat{k}(z)^{\alpha} g(k, z) dk dz \\ \text{s.t.} & \int \int \hat{k}(z) g(k, z) dk dz = \int \int k g(k, z) dk dz \end{aligned}$$

By definition, $\hat{Y}(g)$ is the maximal level of output given the capital stock in this economy. It is determined by the marginal density of z: the thought experiment that generates this quantity freely allocates capital between firms with different productivity, only respecting the constraint on the total capital stock. Using $Y(g)$ and $\hat{Y}(g)$, we define misallocation as

$$\mathcal{M}(g) \equiv \log \hat{Y}(g) - \log Y(g)$$

Characterizing the maximum level of output is straightforward: the marginal product of capital must be equalized across firms. This implies that the hypothetical optimal capital allocation $\hat{k}(z)$ is linear in z, so for a firm with productivity z it equals $\hat{k}(z) = zK(g)/Z(g)$, where we define $Z(g) \equiv \int \int z g(k, z) dk dz$ as the average productivity and $K(g) \equiv \int \int k g(k, z) dk dz$ as the total capital stock. The distribution of ARPK is degenerate and the maximum output is

$$\hat{Y}(g) = Z(g)^{1-\alpha} K(g)^{\alpha}$$

Misallocation is then equal to

$$\mathcal{M}(g) = -\log \int\int \left(\frac{z}{Z(g)}\right)^{1-\alpha} \left(\frac{k}{K(g)}\right)^{\alpha} g(k,z)\, dkdz \tag{6}$$

Computing this object is challenging because it depends on $g(k,z)$. The literature has developed approximations that do not rely on tracking the entire distribution $g(k,z)$ and only require knowing the density $\phi(x)$. This density is stationary, which simplifies computation.

The most popular approach to this reduction of dimensionality assumes that g is either exactly log-normal, as in the frictionless model of Hsieh and Klenow (2009) with log-normal wedges, or approximately log-normal due to small shocks, as in the model of David and Venkateswaran (2019) with convex adjustment costs. Under this assumption, misallocation is related to the variance of log-ARPK. Take equation (6) and replace the double integral with the expectation operator:

$$\mathcal{M}(g) = (1-\alpha)\log(\mathbb{E}[z]) + \alpha \log(\mathbb{E}[k]) - \log(\mathbb{E}[z^{1-\alpha}k^{\alpha}])$$

Now imagine that (z,k) are jointly log-normal, perhaps exactly, like due to wedges in Hsieh and Klenow (2009), or approximately, like in the limit of small shocks in David and Venkateswaran (2019) and Baley and Blanco (2021). Then the log aggregate productivity and the capital stock only depend on the first and second moments of $\log(z)$ and $\log(k)$:

$$\log(\mathbb{E}[z]) = \mathbb{E}[\log(z)] + \frac{1}{2}\mathbb{V}[\log(z)]$$
$$\log(\mathbb{E}[k]) = \mathbb{E}[\log(k)] + \frac{1}{2}\mathbb{V}[\log(k)]$$

The log of aggregate output additionally depends on their covariance:

$$\begin{aligned}\log(\mathbb{E}[z^{1-\alpha}k^{\alpha}]) = (1-\alpha)\mathbb{E}[\log(z)] + \alpha\mathbb{E}[\log(k)] + \frac{(1-\alpha)^2}{2}\mathbb{V}[\log(z)] + \frac{\alpha^2}{2}\mathbb{V}[\log(k)] \\ + (1-\alpha)\alpha \cdot \mathbb{C}[\log(z), \log(k)]\end{aligned} \tag{7}$$

These equations imply (after simple algebra) that

$$\mathcal{M}(g) = \frac{(1-\alpha)\alpha}{2}\mathbb{V}[\log(k) - \log(z)] = \frac{(1-\alpha)\alpha}{2}\mathbb{V}[x] \tag{8}$$

The first moments of $\log(k)$ and $\log(z)$ drop out, while the second moments of $\log(k)$ and $\log(z)$

and their covariance combine into $\mathbb{V}[\log(k) - \log(z)]$. This coincidence is due to log-normality: the second moments of $\log(k)$ and $\log(z)$ enter the expressions with coefficients that make all other terms in equation (8) exactly offset each other. Equation (8) is convenient if log-ARPK have a stationary density, and their variance can easily be computed.

In addition to equation (8), which holds conditional on aggregate capital stock, this capital stock itself has a simple representation with log-normality. Simple log-normal algebra implies

$$\log K(g) = \log Z(g) + \mathbb{E}[x] + \frac{1}{2}\mathbb{V}[x] + \mathbb{C}[\log(z), x] \tag{9}$$

The justification for the log-normality assumption often relies on the small shock approximation, $\sigma \longrightarrow 0$. Equation (8) and equation (9) also hold as a second-order approximation if higher-order moments of log-ARPK vanish faster than the variance. In equation (9), the last two terms are also assumed to vanish faster than the first two, so in practice the literature uses an approximation $\log K(g) \approx \log Z(g) + \mathbb{E}[x]$, relying on the average log-ARPK as a good summary of the aggregate capital-to-productivity ratio.

In contrast, in our model with fixed costs, log-normality does not obtain even in the limit of small idiosyncratic shocks, nor do higher-order moments of log-ARPK vanish in the limit. The reason is that firms cannot adjust at will and allow their log-ARPK to drift away from the optimal level for a while. Even when they do get an opportunity to adjust, small enough deviations from the optimum do not always warrant paying the fixed cost. At the same time, without idiosyncratic shocks, the capital-productivity ratio simply drifts in the same direction for all firms until they reset it to $x = x^*$. As a result, the distribution of log-ARPK in the limit $\sigma \longrightarrow 0$ is both non-degenerate and substantially different from log-normal. We demonstrate this in the following proposition.

Proposition 2. *Suppose that the generalized hazard function $\Lambda(\cdot;\sigma)$ converges uniformly to $\Lambda_0(\cdot)$ as $\sigma \longrightarrow 0$, where $\Lambda_0(\cdot)$ is bounded and $\inf_x \Lambda_0(x) > 0$. Suppose that $\mathbb{E}[e^{2\alpha x}|\sigma]$ exists for all $\sigma \in (0, \overline{\sigma})$ for some $\overline{\sigma} > 0$ and let $\mu + \delta > 0$. The stationary distribution of x then weakly converges to the distribution with a density $\phi_0(\cdot)$, where $\phi_0(x) = 0$ for $x > x^*$, and for $x \leq x^*$,*

$$\phi_0(x) \propto \exp\left(-\frac{1}{\mu+\delta}\int_x^{x^*} \Lambda_0(s)ds\right)$$

For all fixed $t < \infty$, misallocation and capital stock converge to the following limits as $\sigma \longrightarrow 0$:

$$\mathcal{M}(g) \longrightarrow \alpha \mathbb{J}_0[e^x] - \mathbb{J}_0[e^{\alpha x}]$$

$$\log K(g) \longrightarrow \log Z(g) + \mathbb{E}_0[x] + \mathbb{J}_0[e^x]$$

where $\mathbb{J}[x] \equiv \log(\mathbb{E}[x]) - \mathbb{E}[\log(x)]$, and $\mathbb{E}_0[\cdot]$ and $\mathbb{J}_0[\cdot]$ are taken at the limiting density $\phi_0(\cdot)$.

The limiting density has two salient properties. First, it is only positive to the left of x^*. Without shocks, firms never find themselves with $x > x^*$: they land exactly at x^* when they invest, and then the drift takes them to the left of this point until the next adjustment. Second, the density is monotone wherever it is positive: as firms drift away from x^* at a constant rate, their population decreases, since some of them adjust their log-ARPK and go back to x^* in process.

Misallocation converges to a combination of two Jensen's correction terms $\alpha \mathbb{J}_0[e^x]$ and $\mathbb{J}_0[e^{\alpha x}]$. If x were normally distributed, these terms would be equal to $\alpha \mathbb{V}[x]/2$ and $\alpha^2 \mathbb{V}[x]/2$, restoring the familiar formula $\mathcal{M} = \alpha(1-\alpha)\mathbb{V}[x]/2$. If higher-order moments of x vanished, and did it faster than the variance, then $\alpha \mathbb{E}[x] + \alpha \mathbb{V}[x]/2$ and $\alpha \mathbb{E}[x] + \alpha^2 \mathbb{V}[x]/2$ would be the second-order approximations to $\alpha \mathbb{J}_0[e^x]$ and $\mathbb{J}_0[e^{\alpha x}]$, respectively. In that case, $\alpha(1-\alpha)\mathbb{V}[x]/2$ would be the leading term in the Taylor expansion for $\mathcal{M}$. It is deviations from normality and the fact that higher-order moments of the distribution of log ARPK are finite even in the limit of small shoocks that prevent this from happening.

Another observation is that $\mathbb{E}[x]$ is not a good statistic for the aggregate capital-to-productivity ratio $\log K(g) - \log Z(g)$. As long as the Jensen correction terms do not vanish as $\sigma \longrightarrow 0$, using the $\mathbb{E}[x]$ instead of $\log(\mathbb{E}[e^x])$ introduces a non-trivial mistake in capital stock aggregation. If x were normally distributed, this error would be proportional to its variance, and if the variance vanished in the limit of small shocks, then $\log K(g) - \log Z(g) \approx \mathbb{E}[x]$ would work approximately. Proposition 2 shows that neither is the case.

Importantly, it is still sufficient to use the marginal distributions of x and z to compute all three statistics in Proposition 2. The reason is that the covariance between x and z does not show up in the zeroth order because small shocks make z deterministic.

Special case. In the special case with a constant generalized hazard function Λ, there are closed-form solution for moments of log-ARPK, and the connection between the failure of approximate log-normality and the non-degenerate limiting distribution as $\sigma \longrightarrow 0$ is more clear:

Corollary 1. *Fix x^* and $\mu + \delta > 0$. Let $\Lambda_0(\cdot) \equiv \lambda > 0$ and maintain the assumptions from Proposition 2. For all fixed $t < \infty$, the moments of x converge to the following limits as $\sigma \longrightarrow 0$:*

$$\begin{aligned} \mathbb{E}[x] &\longrightarrow E \equiv x^* - \frac{\mu + \delta}{\lambda} \\ \mathbb{V}[x] &\longrightarrow V \equiv \frac{(\mu + \delta)^2}{\lambda^2} \end{aligned} \tag{10}$$

Misallocation and capital stock converge to

$$\begin{aligned} \mathcal{M}(g) &\longrightarrow \log\left(1 + \alpha\sqrt{V}\right) - \alpha \log\left(1 + \sqrt{V}\right) \\ \log K(g) &\longrightarrow \log Z(g) + E + \sqrt{V} - \log\left(1 + \sqrt{V}\right) \end{aligned}$$

The limiting expression for misallocation $\mathcal{M}$ is an increasing function of V. An implication is that misallocation decreases in λ and increases in $\mu + \delta$. The former is not surprising. The latter might seem surprising, but has a simple explanation: with constant adjustment opportunities, faster drift increases the dispersion of capital-to-productivity ratios and exacerbates output losses. This result echoes results from the literature on sticky prices, where, all else equal, higher levels of inflation lead to higher markup dispersion and increase output losses (see Cavallo, Lippi, and Miyahara (2023) and Adam, Alexandrov, and Weber (2024) for examples of this mapping).

Restoring the approximation. Corollary 1 suggests is a way to restore the variance approximation equation (8). It is especially clear from the corollary that if the variance of log-ARPK converged to zero in the limit of small shocks, misallocation would be well approximated by $\alpha(1 - \alpha)V/2$, in the sense that $\alpha(1 - \alpha)V/2$ would become the leading term in the Taylor expansion of $\mathcal{M}$. Equation (10) suggests a way to achieve this: if adjustment frictions vanished together with idiosyncratic shocks, meaning $\lambda \longrightarrow \infty$ as $\sigma \longrightarrow 0$, then V would converge to zero. This would restore $\alpha(1 - \alpha)V/2$ to the position of the leading term in the Taylor expansion for $\mathcal{M}$, and x^*, which is also the expected log ratio x without frictions, would be the leading term in the Taylor expansion for $\log K(g) - \log Z(g)$. We develop a general version of this result.

PROPOSITION 3. *Fix $\Lambda(x;\sigma) = \Lambda_0(x)/\kappa$ and let $\min_x \Lambda_0(x) > 0$. Assume that $\mathbb{E}[e^x]$ exists. Let $\sigma \longrightarrow 0$ and $\kappa \longrightarrow 0$ so that $\kappa = O(\sigma^b)$ for some $b \in [0,2]$. For all fixed $t < \infty$,*

$$\mathbb{V}[x] = O(\sigma^{2b})$$

Misallocation and aggregate capital stock are

$$\mathcal{M}(g) = \frac{\alpha(1-\alpha)}{2}\mathbb{V}[x] + O\left(\sigma^{\min\{1+2b,3b\}}\right)$$
$$\log K(g) = \log Z(g) + \mathbb{E}[x] + O\left(\sigma^{\min\{1+b,2b\}}\right)$$

This result shows that the properties usually obtained by assuming exact log-normality or quickly decaying moments of log-ARPK are restored in the double limit of small shocks and small frictions. If adjustment frictions vanish together with shocks, the variance of log-ARPK vanishes too, and $\alpha(1-\alpha)\mathbb{V}[x]/2$ becomes the leading term in the expression for misallocation. The average log-ARPK becomes a good summary of aggregate capital: $\mathbb{E}[x]$ becomes the leading term in $\log K(g) - \log Z(g)$. This is particularly useful in light of the results obtained by Baley and Blanco (2021), who develop analytical characterization of the cumulative impulse responses of aggregate capital stock using $\mathbb{E}[x]$ as an approximation.

In principle, the rate of decay in $\mathbb{V}[x]$ can be different depending on the speed of divergence of the generalized hazard function. For instance, if $b = 1$, then $\Lambda(x)$ diverges as $1/\sigma$, and we obtain the classical case $\mathbb{V}[x] \sim \sigma^2$. Misallocation is then well approximated by $\alpha(1-\alpha)\mathbb{V}[x]/2$, and the next term in the expression for $\mathcal{M}(g)$ is of order σ^3. In a less typical case of $b = 1/2$, the orders are $\Lambda(x) \sim 1/\sqrt{\sigma}$ and $\mathbb{V}[x] \sim \sigma$. The variance of log-ARPK is still a good approximation for misallocation, and the next term in the expression for misallocation is then of order $\sigma^{3/2}$. If $b = 0$, adjustment frictions do not vanish, and misallocation loses its connection to $\mathbb{V}[x]$, while $\mathbb{E}[x]$ stops being a good approximation for aggregate capital.

A final remark here is that varying σ and $\Lambda(\,\cdot\,;\sigma)$ together in a particular way does not hold constant the primitives of the model $\{\gamma_i, G_i(\cdot)\}_{i=u,d}$. Preserving the shape of $\Lambda(\,\cdot\,;\sigma)$ as $\sigma \longrightarrow 0$

requires changing the distributions of random adjustment costs in the background, although we do not explicitly compute them as we take the limit.

3.2 Relationship to convex adjustment costs.

We now briefly point out that the reason why misallocation in our model is not well approximated by the variance of log-ARPK is the lumpiness of investment and not the adjustment friction itself. To that end, we explain the relationship between our model with random fixed costs and models with convex adjustment costs, such as that of David and Venkateswaran (2019). We use a simplified version of David and Venkateswaran (2019) without wedges and information frictions.

The main takeaway is that, unlike fixed costs, convex adjustment costs do not lead to a non-degenerate distribution of log-ARPK in the limit of small idiosyncratic shocks. With the distribution of log-ARPK converging to a degenerate measure, higher-order moments vanish faster than lower-order ones, so $\mathbb{E}[x]$ is the leading term in the expression for the aggregate capital-to-productivity ratio, while $\alpha(1-\alpha)\mathbb{V}[x]/2$ is the leading term in the expression for misallocation.

The intuition for why convex costs do not create a non-degenerate distribution of log-ARPK without idiosyncratic shocks is simple. Convexity makes firms dislike time variation in adjustment, so they decide to smooth investment over time as much as possible. When $\sigma \longrightarrow 0$, firms simply choose to offset the drift in log-ARPK all the time, maintaining $x = x^*$.

To make this concrete, take firms that produce $y = z^{1-\alpha}k^{\alpha}$ and invest subject to a quadratic investment cost. Denote the investment rate by i. The evolution of capital and productivity is $d\log(k) = (i-\delta)dt$ and $d\log(z) = \mu dt + \sigma dW$. The recursive problem of the firm is

$$rV(k,z) = \max_i \; z^{1-\alpha}k^{\alpha} + (i-\delta)kV_k(k,z) - ik - \frac{\varphi k}{2}i^2 + \left[\mu + \frac{\sigma^2}{2}\right] zV_z(k,z) + \frac{\sigma^2 z^2}{2}V_{zz}(k,z)$$

The optimal investment rate is $i = \varphi^{-1}(V_k(k,z)-1)$. Make a change of variables to the capital-productivity ratio: $y \equiv k/z$. Consider a different value function $v(\cdot)$ given by $V(k,z) = zv(k/z)+k$. Implementing this change of variables and denoting $\rho = r - \mu - \sigma^2/2$ and $\nu = r + \delta$, as before,

$$\rho v(y) = y^{\alpha} - \nu y + (\rho - \nu)yv'(y) + \frac{\sigma^2 y^2}{2}v''(y) + \frac{[v'(y)]^2 y}{2\varphi} \qquad \text{(convex costs)}$$

Compare this to the value with fixed costs:

$$\rho v(y) = y^{\alpha} - \nu y + (\rho - \nu) y v'(y) + \frac{\sigma^2 y^2}{2} v''(y) + \mathcal{H}(v(y^*) - v(y)) \qquad \text{(fixed costs)}$$

Here $\mathcal{H}(\cdot)$ is the option value of capital stock adjustment:

$$\mathcal{H}(v(y^*) - v(y)) = \sum_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\} dG_i(\psi)$$

The two HJB equations only differ in the last, adjustment-related, terms that introduce non-linearities. With convex costs, only the local shape of v determines the speed of adjustment and the value of the adjustment option, while with fixed costs, the option value of adjusting depends on the global properties of the value.

The key difference between the models is in the Kolmogorov forward equations that determine their stationary distributions. To compare the stationary distributions, change the variables to $x \equiv \log(y)$. Optimal investments rate is $\imath(x) = \varphi^{-1} v'(e^x)$. The distribution ϕ solves the following:

$$[\imath(x)\phi(x)]' = (\mu + \delta)\phi'(x) + \frac{\sigma^2}{2}\phi''(x) \qquad \text{(convex costs)}$$

Compare this to

$$\Lambda(x)\phi(x) = (\mu + \delta)\phi'(x) + \frac{\sigma^2}{2}\phi''(x) \qquad \text{(fixed costs)}$$

Heuristically, setting $\sigma = 0$ in the equation for the fixed cost case leads to a first-order differential equation. This is because the term $\Lambda(x)\phi(x)$ on the left reflects discrete transitions of firms when they reset their ARPK. Investment takes the form of occasional jumps. In the convex cost case, setting $\sigma = 0$ leads to $[\imath(x)\phi(x)]' = (\mu + \delta)\phi'(x)$, which implies $\imath(x)\phi(x) = (\mu + \delta)\phi(x)$, so $\phi(x)$ must be zero everywhere except the point where $\imath(x) = \mu + \delta$. In contrast to the lumpy case, in the convex case investment is a form of drift, and it has to offset other sources of drift to induce stationarity. Intuitively, if firms do not exactly offset the drift in their log-ARPK with investment, they move around the distribution. Since without shocks all firms are moving in the same direction, this contradicts stationarity. We formalize the result in the following proposition.

PROPOSITION 4. *Suppose the investment function $\imath(\cdot; \sigma)$ converges uniformly to $\imath_0(\cdot)$ as $\sigma \longrightarrow 0$,* *where* $\inf_x \imath_0'(x) > 0$. *The stationary measure of firms weakly converges to the measure fully concentrated at* $x = x_0^*$ *given by* $\imath_0(x_0^*) = \mu + \delta$. *Misallocation and aggregate capital stock are*

$$\mathcal{M}(g) = \frac{\alpha(1-\alpha)}{2}\mathbb{V}[x] + O(\sigma^3)$$
$$\log K(g) = \log Z(g) + \mathbb{E}[x] + O(\sigma^2)$$

Here $\mathbb{V}[x] = O(\sigma^2)$ *and* $\mathbb{E}[x - x_0^*] = O(\sigma)$.

A final remark to make about the relationship between convex cost and fixed cost model is that the former preserve approximate log-normality because they are amenable to linearization. Indeed, David and Venkateswaran (2019) log-linearize the model first, taking advantage of the smoothness of adjustment costs, and then use log-normality of the underlying shocks.[6] Fundamental non-linearities of the fixed cost models introduce a disconnect between the distribution of exogenous shocks and endogeneous states. Computing misallocation and other aggregate statistics around the limit with no idiosyncratic shocks requires additional care.

4 Mapping the model to the data

We now explain the connection between $\Lambda(\cdot)$ and observable objects. This is the second part of our mapping between the data and model primitives, where $\Lambda(\cdot)$ is the bridge between observable objects and parameters of the fixed costs. The connection works through aggregation of adjustment by all firms in the economy, which is summarized by the Kolmogorov forward equation (5).

Productivity is not observable, so $\{\phi(\cdot), \Lambda(\cdot)\}$ cannot be simply read from the data. We use the observed distribution and frequency of investments to recover them. Upon observing an adjustment, we record its size in logs, $\Delta \log(K) = \ln(K_+) - \ln(K_-)$, where K_- and K_+ is capital stock right before and right after the change. Since z has continuous sample paths, $\Delta \log(K) = \Delta x$.

Let the distribution of recorded changes in log capital stock be $Q(\cdot) : \mathbb{R} \mapsto [0, 1]$. Denote the

[6] The literature using convex adjustment costs in DSGE models goes back to at least Lucas and Prescott (1971). Recent contributions, like Ottonello and Winberry (2020), move production of investment goods to a separate sector of the economy and equip firms in that sector with a concave production function, which makes aggregate dynamics the same as in the case with convex adjustment costs.

corresponding density by $q(\cdot)$. Then, letting N be the frequency of adjustments,

$$\phi(x)\Lambda(x) = Nq(x^* - x)$$

The number of adjustments of the size $x^* - x$ per unit of time is equal to the number of firms who have a log ARPK of $(\alpha - 1)x$ times the probability per unit of time for such firms to adjust. Crucial here is that all firms choose exactly x^* when they invest or disinvest.

While $\phi(\cdot)$ and $\Lambda(\cdot)$ are not separately observable, their product counts adjustments among the firms with log ARPK x, and so can be mapped to the density of adjustments of size $x^* - x$. Importantly, $\phi(\cdot)$ and $\Lambda(\cdot)$ are not independent objects that could be assumed to be arbitrary for estimation or calibration: $\Lambda(\cdot)$ determines the steady-state density $\phi(\cdot)$ through equation (5), and the observed distribution of investments jointly determines them both. In Section 4, we use this fact to identify $\phi(\cdot)$ and $\Lambda(\cdot)$ from the observed distribution of investment sizes and its frequency.

It is convenient to re-center $\phi(\cdot)$ and $\Lambda(\cdot)$ around x^*. Define re-centered functions $f(\cdot)$ and $L(\cdot)$ by $f(x) = \phi(x + x^*)$ and $L(x) = \Lambda(x + x^*)$, and call $x - x^*$ capital gaps. We have

$$\begin{aligned} f(x)L(x) &= (\mu + \delta)f'(x) + \frac{\sigma^2}{2}f''(x) \\ f(x)L(x) &= Nq(-x) \end{aligned}$$

Replacing $f(x)L(x)$ in the first equation with a known function $Nq(-x)$ turns it into a linear ordinary differential equation for $f(\cdot)$, which can be solved by repeated integration. With $f(\cdot)$ on hand, one can recover $L(\cdot)$ from $f(x)L(x) = Nq(-x)$.

While theoretically feasible, this direct approach is likely to run into practical issues. Specifically, the full density $q(\cdot)$ is usually not accessible to the econometrician, who is forced to use histograms instead. In addition to noise in the estimate of the density, which is particularly unstable in the tails and around the kink at x^*, numerical integration of the ODE can take very long, reducing the number of parameters that could be estimated in a reasonable time span. We next propose an indirect approach that leverages the histogram structure of the data and circumvents issues related to numerical integration by using closed-form solutions.

4.1 Recovering (L, f) in practice

In practice, estimation procedures use histograms, which pool observation within bins and provide limited information on the tails. We propose a method to recover (L, f) under a functional form assumption that maximizes computational convenience taking into account these data limitations. Specifically, we assume that $L(\cdot)$ is constant within each bin of the observed adjustment histogram. Since the information on the functional form of $Q(\cdot)$ within bins is lost anyway, we choose a data generating process that makes computations fast and easily scales with the number of bins.

Formally, let positive investments be binned into $\mathbb{u}$ bins with the mass of H_j in each bin. Since any investment observation x corresponds to a gap $-x$ before adjustment, these negative gaps fall into $\mathbb{u}$ bins X_j with boundaries $\{x_j\}_{-\mathbb{u} \leq j \leq 0}$, where $x_0 = 0$ and $x_{-\mathbb{u}} = -\infty$. Accordingly, let the positive gaps corresponding to negative investment be pooled into $\mathbb{d}$ bins X_j with boundaries $\{x_j\}_{0 \leq j \leq \mathbb{d}}$, where $x_0 = 0$ and $x_{\mathbb{d}} = \infty$.

Assumption 1. *The re-centered generalized hazard function $L(\cdot)$ is given by $L(x) = \lambda_j$ for $x \in X_j$, where $X_j = (x_j, x_{j+1}]$ for $-\mathbb{u} \leq j \leq -1$ and $X_j = [x_{j-1}, x_j)$ for $1 \leq j \leq \mathbb{d}$.*

Under this assumption, the model has the following parameters: the drift of capital gaps μ, the volatility of productivity shocks σ, and $\mathbb{u} + \mathbb{d}$ hazard levels $\boldsymbol{\lambda} = \{\lambda_j\}_{-\mathbb{u} \leq j \leq \mathbb{d}, j \neq 0}$. We denote the set of parameters by $\mathcal{P} = (\mu, \sigma^2, \boldsymbol{\lambda})$.[7]

The data provide a frequency of investments N and a histogram $\boldsymbol{Q} = \{Q_j\}_{-\mathbb{d} \leq j \leq \mathbb{u}, j \neq 0}$ of investment sizes. The recorded histogram $\boldsymbol{Q}$ contains measurement error and might differ from the true histogram generated by the model. To formalize this, we denote the true data by $\mathcal{D} = (N, \boldsymbol{H})$, where the histogram $\boldsymbol{H} = \{H_j\}_{-\mathbb{d} \leq j \leq \mathbb{u}, j \neq 0}$ is potentially different from the observed $\boldsymbol{Q}$. We next characterize the mapping $\mathcal{P} \mapsto \mathcal{D}$.

[7] We estimate the deprecation rate directly using depreciation reported by firms.

4.2 Mapping parameters to the data

Under Assumption 1, the solution to the homogeneous equation (5) is a linear combination of two exponential functions on each segment X_j:

$$f_j(x) = \eta_{1,j} e^{\xi_{1,j} x} + \eta_{2,j} e^{\xi_{2,j} x} \tag{11}$$

The distribution of gaps over any X_j is given by $f(x) = f_j(x)$. The powers are easily computed:

$$\{\xi_{1,j}, \xi_{2,j}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2 \lambda_j}}{\sigma^2} \tag{12}$$

Denote the vectors of exponents by $\boldsymbol{\xi}_1 = \{\xi_{1,j}\}_{-\mathbb{u} \le j \le \mathbb{d}, j \neq 0}$ and $\boldsymbol{\xi}_2 = \{\xi_{2,j}\}_{-\mathbb{u} \le j \le \mathbb{d}, j \neq 0}$. The coefficients $\boldsymbol{\eta}_1 = \{\eta_{1,j}\}_{-\mathbb{u} \le j \le \mathbb{d}, j \neq 0}$ and $\boldsymbol{\eta}_2 = \{\eta_{2,j}\}_{-\mathbb{u} \le j \le \mathbb{d}, j \neq 0}$ satisfy the following continuity conditions: the probability density must be continuous, including at zero, and differentiable at the boundaries between all segments except for the junction at $x = 0$. The jump in the first derivative of $f(\cdot)$ at zero is due to the "reinjection" of firms: they adjust capital discretely, continually arriving at zero. The size of the jump in $f'(\cdot)$ is

$$\lim_{x \to -0} f'_{-1}(x) - \lim_{x \to +0} f'_1(x) = \frac{2N}{\sigma^2} \tag{13}$$

This can be shown by integrating equation (5) over the real line and using the statistical fact that $q(-x)N = L(x) f(x)$. In addition, $f(\cdot)$ must not explode at infinity, and it must integrate to one. These conditions provide $2(\mathbb{u}+\mathbb{d})$ linear equations to solve for $2(\mathbb{u}+\mathbb{d})$ unknowns $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. The next proposition uses this fact to establish the mapping from parameters to the data:

Proposition 5. *Fix $\mathcal{P}$, a $(\mathbb{u}+\mathbb{d})$-dimensional non-negative vector $\boldsymbol{\lambda}$ and a pair (μ, σ^2). The density of log ARPK is given by equation (11), where the coefficients $\boldsymbol{\xi}_1(\mathcal{P})$ and $\boldsymbol{\xi}_2(\mathcal{P})$ are given by equation (12), and $\boldsymbol{\eta}_1(\mathcal{P})$ and $\boldsymbol{\eta}_2(\mathcal{P})$ solve a $2(\mathbb{u}+\mathbb{d})$-dimensional linear system. The true data $\mathcal{D} = (N, \boldsymbol{H})$ are given by the functions $N = n(\mathcal{P})$ and $H_{-j} = h_j(\mathcal{P})$ for $-\mathbb{u} \le j \le \mathbb{d}$ with $j \neq 0$:*

$$n(\mathcal{P}) := \frac{\sigma^2}{2}(\eta_{1,-1}\xi_{1,-1} + \eta_{2,-1}\xi_{2,-1} - \eta_{1,1}\xi_{1,1} - \eta_{2,1}\xi_{2,1}) \tag{14}$$

$$h_j(\mathcal{P}) := \frac{1}{N}\left[\frac{\lambda_j \eta_{1,j}}{\xi_{1,j}}(e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}) + \frac{\lambda_j \eta_{2,j}}{\xi_{2,j}}(e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j})\right] \tag{15}$$

We show how to construct the linear system for $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ in the proof. As a corollary, we note a homogeneity property:

Corollary 2. *The coefficients $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, $\boldsymbol{\xi}_1$, and $\boldsymbol{\xi}_2$ and the histogram $\boldsymbol{H}$ are homogeneous of degree zero in $\mathcal{P} = (\mu, \sigma^2, \boldsymbol{\lambda})$. The frequency N is homogeneous of degree one.*

This property means that the drift and variance of the underlying process and the adjustment hazard are only pinned down in levels by the observed frequency. Scaling them all together scales the frequency without changing the histogram. In practice, it means that we can set $N = 1$ in all equations to estimate the full set of parameters up to a common constant without using the time dimension of the data and then scale the estimates using the observed frequency. Another option is to fix σ^2 or μ, optimize over all other parameters, and then scale the estimates by the ratio of the observed frequency to that in equation (14).

4.3 Estimation

The vector $\boldsymbol{\lambda}$ can be estimated by minimizing the error in equation (15) under one restriction: the menu cost model implies that $L(x)$ is non-decreasing for $x > 0$ and non-increasing for $x < 0$.

In practice, this implies that the estimates $\hat{\boldsymbol{\lambda}}$ should satisfy $\hat{\lambda}_{j+1} \geq \hat{\lambda}_j$ for $j > 0$ and $\hat{\lambda}_{j-1} \geq \hat{\lambda}_j$ for $j < 0$. Together with $(\hat{\mu}, \hat{\sigma}^2)$, the total set of resulting estimates $\hat{\mathcal{P}} = (\hat{\mu}, \hat{\sigma}^2, \hat{\boldsymbol{\lambda}})$ solves

$$\hat{\mathcal{P}} = \arg\ \min_{\mathcal{P}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{16}$$
$$\text{s.t. } \lambda_{j+1} \geq \lambda_j \text{ for } j > 0,\ \lambda_{j-1} \geq \lambda_j \text{ for } j < 0,\ n(\mathcal{P}) = N$$

Here the true data $\boldsymbol{H}(\mathcal{P})$ and $n(\mathcal{P})$ are given by equation (15), and $(\boldsymbol{Q}, N)$ are the recorded data. In practice, we first solve the relaxed version of the problem in equation (16) without the condition $n(\mathcal{P}) = N$ by fixing σ^2 and maximizing over $(\mu, \boldsymbol{\lambda})$. By the homogeneity property in Corollary 2, we can then divide $\mathcal{P} = (\mu, \sigma^2, \boldsymbol{\lambda})$ by $n(\mathcal{P})/N$ to obtain the solution to the full problem.

We make a brief note on what the constraints on the monotonicity of $\hat{\boldsymbol{\lambda}}$ imply for the shape of the adjustment histogram $\mathcal{H}(\hat{\mathcal{P}})$. Proposition 7 in Appendix C.2 shows that, when $\boldsymbol{\lambda}$ is close to constant, making it slightly increasing in distance from x^* leads to a thicker adjustment histogram. An intuitive implication is that in practice, the monotonicity constraint is likely to bind when the tails of the empirical histogram $\mathcal{Q}$ are too thin or decrease too fast. In this case, the shape of the histogram $\mathcal{Q}$ indicates forces unaccounted for by our model. A pervasive issue that might lead to estimating potentially non-monotone generalized hazard functions is unobserved heterogeneity.[8] Alternatively, Blanco and Baley (2024) estimate a model with partial investment irreversibility generated by a spread between buying and selling prices of investment goods. In their model, firms have different reset points x^* depending on the sign of adjustment. Blanco and Baley (2024) show that a restricted version of their model, where the spread is eliminated, leads to non-monotone estimates of the generalized hazard function, and incorporating the spread solves this issue.

4.4 Recovering adjustment costs

To recover the primitive distributions of adjustment costs $G_i(\cdot)$ and arrival intensities γ_i, we need the true hazard Λ instead of the re-centered version $L(x) = \Lambda(x - x^*)$. The challenge is that x^* is not observable and we need to solve for it. Our procedure relies on the optimality condition for x^*: the marginal value function $\mathbb{u}$ satisfies $\mathbb{u}(x^*) = 0$. We can guess x^* to obtain marginal value $\mathbb{u}$ from equation (3) and then update the guess based on this optimality condition until convergence.

Specifically, having $L(\cdot)$ and a guess of x^*, we solve

$$(\nu + L(x + x^*))U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu + \delta)U'(x) + \frac{\sigma^2}{2}U''(x)$$

using the procedure described in Appendix D. This procedure takes advantage of the fact that $L(\cdot)$ is piece-wise constant and turns solving a differential equation into solving a linear system. We then update the guess of x^* by finding the point at which $U(x^*) = 0$.

Piece-wise constant generalized hazard functions correspond to piece-wise constant distributions $G_i(\cdot)$ and hence discrete sets of adjustment cost ψ. Take upward adjustments first. They happen when log ARPK is less than the reset point, $x < 0$. It is straightforward to compute the arrival intensity $\gamma_u = \lim_{x\to-\infty} \Lambda(x)$. To find the values of ψ with positive mass, use

$$\Lambda(x) = \gamma_u G_u \left(\int_x^{x^*} U(t)e^t dt \right)$$

[8] See equation (22) in Alvarez, Borovičková, and Shimer (2023), and the related references therein.

The values of x at which $\Lambda(\cdot)$ jumps map into values of ψ at which $G_u(\cdot)$ jumps: for any such x,

$$\psi(x) = \int_x^{x^*} U(t)e^t dt$$

The size of jumps in $G_u(\cdot)$ corresponds to the size of jumps in $\Lambda(\cdot)$ scaled by γ_u. The case of downward adjustments is treated in the same way.

5 A calibration on Italian manufacturing investment data

We fit our model to panel data on Italian manufacturing firms drawn from the Company Accounts Data Service (CADS, *Centrale dei Bilanci* in Italian), which collects annual balance-sheet information and other items on a sample of over 45,000 Italian limited-liability firms and listed corporations, over a 40 year period, from 1983 to 2023. The data contain balance sheet information on the firm's assets, investment in tangible and non-tangible assets, as well as disinvestments. The data also contain information about employment (total number of employees) and provides a detailed description of the firms' demographic characteristics such as the year of foundation, location, type of organization and ownership status, and the industry and region in which each firm operates. A subset of these data, covering the 1983 to 2004 period, was used by Guiso and Schivardi (2007) and Guiso et al. (2017). The firms included in the CADS database are borrowers of leading banks in Italy. The focus on this sample of firms skews the sample towards larger firms that are considered credit-worthy (as firms in default are not included in the database).[9]

We calculate gross investment and disinvestment rates separately. This is in contrast to Baley and Blanco (2021), who use net investment. Having records of both positive and negative adjustments allows us to identify $\{\gamma_u, G_u(\cdot)\}$ separately from $\{\gamma_d, G_d(\cdot)\}$ and estimate inactivity more precisely. With substantial degree of time aggregation, as is the case with our annual data, firms often both invest and disinvest within period, so lumping positive and negative investments together could inflate the observed share of inactivity and distort the estimates of costs.

[9] Since most leading banks are in the Northern part of the country, the sample has more firms headquartered in the North than in the South. Despite this, a comparison between sample and population moments, done by Guiso and Schivardi (2007), suggests that the CADS is not too far from being representative of the whole population.

Table 1: Summary statistics by industry

Industry	# Firms	Observations	Years	# Workers
Mining and Quarrying	2700	23961	8.9	128
Chemicals	12579	131216	10.4	74
Metals and Machinery	36992	367749	9.9	67
Food and Beverages	40295	406595	10.1	38
Construction	18882	146490	7.8	28
Retail	58074	536840	9.2	34
Transportation	8941	86716	9.7	134
Insurance	11329	91008	8.0	69
Health and Beauty	6850	75863	11.1	118
Total	196,642	186,6438	9.5	54

We normalize investment I and disinvestment D by the stock of the illiquid assets, given by the total assets less financial and other liquid assets. This is our measure of the firm's capital. In the main text, we will refer to the illiquid assets simply as "assets", A. Importantly, A is recorded at the end of the period, when capital stock adjustments have been made. Capital stock before investment or disinvestment is then $A - N$, where $N = I - D$ is net investment.

Our variable of interest is the log change in assets resulting from investment or disinvestment:

$$\Delta x_I \equiv \ln\left(\frac{A - N + I}{A - N}\right) \text{ and } \Delta x_D \equiv \ln\left(\frac{A - N - D}{A - N}\right)$$

Table 2 reports summary information sector-by-sector for the sample, using a 9-sector classification.

The data are characterized by a significant presence of inaction. In a typical year, about 14% of the firms are inactive. We follow Cooper and Haltiwanger (2006) and Baley and Blanco (2021) and consider all investments with an absolute size smaller than 1% of the firm's capital "zero investment". Inactive firms are those with both Δx_I and Δx_D below the threshold in a given year.

The average size of net investment (relative to assets) is about 30%. Most of adjustments has a positive sign: the fraction of disinvestments is small in all sectors, pointing to the presence of drift and the possibility of asymmetries in the adjustment costs. There is significant variation across industries, both in the typical size of investments and in the prevalence of inaction.

The prevalence of inaction in either upward or downward adjustments on the firm level does

Table 2: summary statistics of investment by industry

Industry	Inactive	mean(Δx)	std(Δx)	$\mathbb{P}\{\Delta x < 0\}$
Mining and Quarrying	0.24	0.32	0.61	0.07
Chemicals	0.1	0.25	0.46	0.15
Metals and Machinery	0.09	0.3	0.52	0.14
Food and Beverages	0.12	0.26	0.48	0.15
Construction	0.18	0.33	0.63	0.17
Retail	0.16	0.34	0.6	0.13
Transportation	0.18	0.39	0.6	0.07
Insurance	0.17	0.47	0.68	0.09
Health and Beauty	0.13	0.32	0.55	0.12
Total	0.14	0.32	0.56	0.14

Note: Investment is considered zero if the investment to capital ratio is less than 1% in absolute value. Disinvestments are treated in the same way. *Inactive* is the share of firm-year pairs with both zero investment and disinvestment. The mean and standard deviation are computed for non-zero investments.

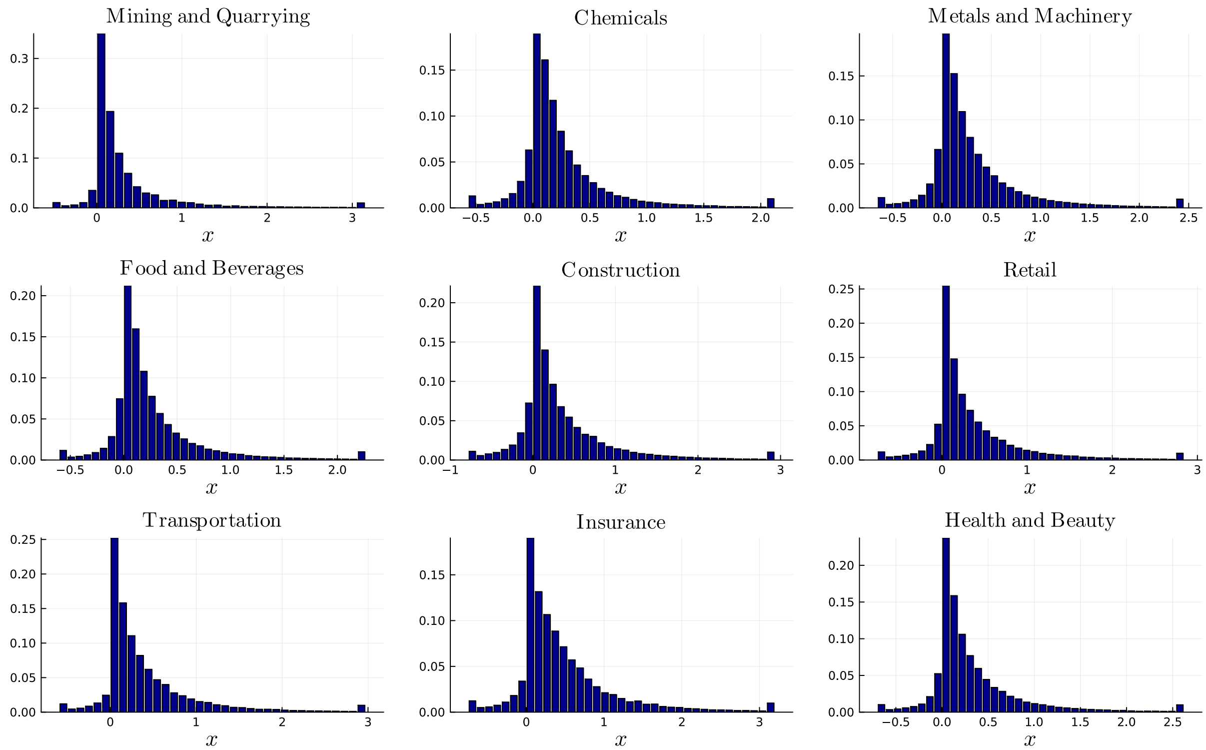
not appear to be related to other observables. Define, for every firm i, the following quantities: up_i as the share of years in which this firm was observed with a non-zero upward investment (relative to its total tenure in the sample), $down_i$ as the share of years it was observed disinvesting, and $inactive_i$ as the share of years it was inactive. Define also rev_i, emp_i, and $time_i$ to be firm i's average revenue, average employment, and the total number of years in the sample. Table 3 shows pairwise correlations of these firm-level time averages for a particular sector "Metals & Machinery".

Table 3: Correlations between firm-level observables and activity measures (Metals & Machinery).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.10	-0.94	0.04	0.05	0.15
$down_i$	0.10		-0.21	0.13	0.18	0.07
$inactive_i$	-0.94	-0.21		-0.04	-0.05	-0.14
rev_i	0.04	0.13	-0.04		0.80	0.04
emp_i	0.05	0.18	-0.05	0.80		0.04
$time_i$	0.15	0.07	-0.14	0.04	0.04	
mean	0.87	0.14	0.11	18,584	67	9.94
std	0.22	0.22	0.20	160,070	445	7.91

Importantly, the correlation between inactivity and firm size, as measured by revenue or employment, is close to zero. Firms do appear to be less inactive on average if they stay in the sample

Figure 1: Distribution of non-zero investments.

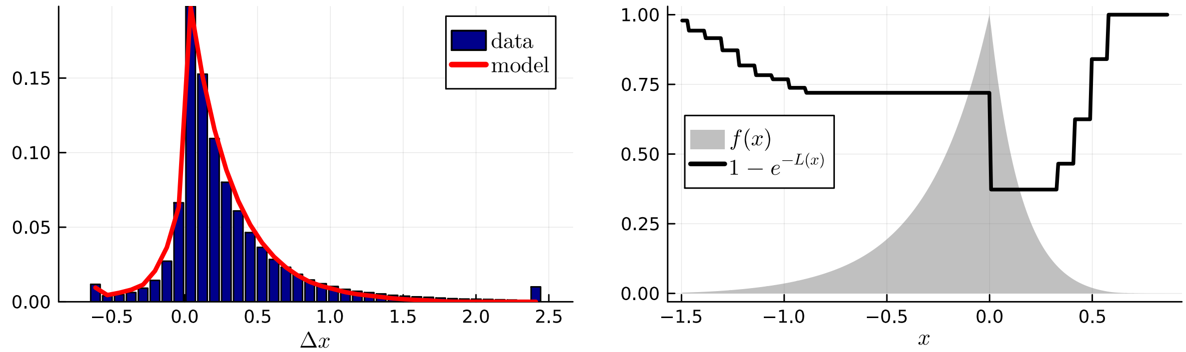


for longer, which might be indicative of right-censoring of inactivity spells. This correlation is not strong either, as firms on average spend around 10 years in the sample, and inactivity spells are rarely longer than one year. Results are similar for other sectors. We present tables for those sectors in Appendix I.

The histograms in Figure 1 describe the distribution of the size of the (non-zero) investments in each of the 9 industries considered. This distribution corresponds to the theoretical measure $Q(\cdot)$ described by the equation $q(-x)N = L(x)f(x)$. The leftmost and rightmost bins in each graph contain the entire tails of the observed distribution.

5.1 Estimating the model

The model is parametrized by $\mathcal{P} = (\mu, \sigma^2, \boldsymbol{\lambda})$ and (r, α, δ). We estimate the first set of parameters, $\mathcal{P}$, using the distribution and frequency of investments. In the second set, we estimate δ for every sector directly from firm reports and set $r = 5\%$. We estimate sector-specific elasticities α in the



(a) Data and the full model.

(b) Recovered generalized hazard.

Figure 2: The distribution of investments and log ARPK and the generalized hazard function implied by the model for "Metals aand Machinery".

following way. Recall the production function: for every firm j, output is $y_{jt} = \hat{z}_{jt}^{(1-\alpha)\zeta} k_{jt}^{\alpha\zeta} l_{jt}^{1-\zeta}$. We project output on labor and capital to estimate input elasticities and deduce α:

$$\log y_{jt} = \beta_0 + \beta_{cap} \log k_{jt} + \beta_{lab} \log l_{jt} + \epsilon_{jt}$$

The elasticities correspond to the primitive coefficients as follows: $1 - \zeta = \beta_{lab}$ and $\alpha\zeta = \beta_{cap}$, so $\alpha = \beta_{cap}/(1 - \beta_{lab})$. We later use the residuals of this projection as log TFPR to compare our results to the approach in Hsieh and Klenow (2009) and Calligaris, Del Gatto, Hassan, Ottaviano, and Schivardi (2018), who use the same data, in Section 6.

Steady-state objects. Having estimated model parameters, we reconstruct the steady-state distributions of investments and log ARPK, as well as the generalized hazard function. Figure 2 shows the empirical histogram against the one generated by the model for one sector from our data, "Metals & Machinery". Appendix E shows the corresponding figures for all other sectors.

Figure 3 shows the recovered distributions of fixed costs for "Metal & Machinery". Panel (a) shows the cumulative distribution functions $G_u(\cdot)$ and $G_d(\cdot)$. We express adjustment costs ψ in percent of $e^{\alpha x^*} - (r + \delta)e^{x^*}$, the instantaneous profits at the optimum x^*. This is the maximum attainable level of profits conditional on the environment. Firms would earn this if their log ARPK was always set to x^*. Panels (b) and (c) on Figure 3 show the arrival frequencies of different adjustment cost values ψ. These figures are histograms corresponding to $G_u(\cdot)$ and $G_d(\cdot)$.

Table 4: frequency of opportunities and costs, drawn and paid, in percent of annual profit

Industry	γ^u	$\mathbb{E}[\psi^u]$	$\mathbb{E}[\psi^u\|\text{paid}]$	γ^d	$\mathbb{E}[\psi^d]$	$\mathbb{E}[\psi^d\|\text{paid}]$	$\mu+\delta$	σ
Mining and Quarrying	21.3	8.5%	0.1%	4.2	4.5%	1.1%	0.14	0.21
Chemicals	36.1	16.4%	0.2%	10.5	7.7%	0.8%	0.19	0.27
Metals and Machinery	43.5	15.3%	0.2%	11.5	7.1%	0.7%	0.24	0.3
Food and Beverages	66.4	12.1%	0.2%	11.8	6.4%	0.5%	0.19	0.26
Construction	50.6	7.8%	0.2%	10.7	5.6%	0.7%	0.19	0.34
Retail	23.7	4.0%	0.1%	9.3	4.0%	0.7%	0.17	0.32
Transportation	72.6	11.6%	0.3%	0.2	1.1%	0.3%	0.22	0.38
Insurance	1.1	0.0%	0.0%	7.8	8.7%	1.8%	0.36	0.4
Health and Beauty	26.2	7.6%	0.2%	9.8	5.7%	0.8%	0.2	0.3

Note: annual profit at the optimum is $e^{\alpha x^*} - (r+\delta)e^{x^*}$

The corresponding figures for all sectors are in Appendix F.

The average adjustment costs for positive investments is equal to 10% of this profit, and these opportunities arrive with an annual intensity of $\gamma_u = 44$. The opportunity for a negative adjustment arrives with an annual intensity of $\gamma_d = 12$, and the average cost is 4% of maximal profits. In this particular sector, the asymmetry between positive and negative investment opportunities comes from differences in both the frequency of arrival and the size of costs.

Table 4 shows the same quantities for other sectors. In most of them, the difference between positive and negative adjustment costs is similar to "Metals & Machinery": opportunities for upward adjustments are more costly but arrive more frequently. A notable exception is "Insurance", where the model recovers rare but costless opportunities for positive investment.

Besides the fundamental distributions of fixed costs drawn by firms, we can also compute the distributions of costs actually paid. If $g_i(\psi)$ is the probability to draw ψ, the probability $\hat{g}_i(\psi)$ to pay it is proportional to $g_i(\psi)\mathbb{P}\{v(x^*) - v(x) \geq \psi\}$. The cost is only paid by those firms for which the value gain is sufficiently large. The average cost paid for a positive investment in "Metal & Machinery" is just 0.1% of the maximum profit, which is 100 times lower than the average cost drawn. For negative investments, the average paid is 0.4% of the maximum profit, a 10-fold decrease relative to the average arriving cost.

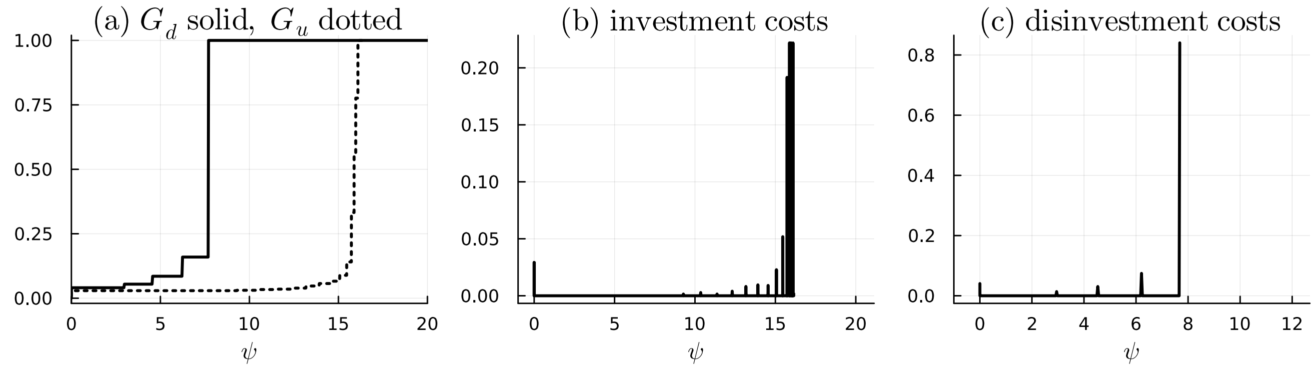


Figure 3: Recovered distributions of adjustment cost ψ in "Metals & Machinery". Left panel: cumulative distribution functions G_u (dotted) and G_d (solid) for costs of positive and negative adjustment. Center and right panels: frequencies of costs of positive and negative adjustments. Costs expressed in percent of instantaneous profits at optimum $e^{\alpha x^*} - (r + \delta)e^{x^*}$.

6 Applications

This section presents four quantitative applications of our model. We first present the approximate measure of misallocation derived in Proposition 2 and compare it to the commonly used variance-based measure $(1 - \alpha)\alpha/2 \cdot \mathbb{V}[x]$. This exercise evaluates the difference between the two approximations and the gains from using the correct one. We then split the sample into subperiods 1983-2003 and 2003-2023 and present evidence on dynamics of misallocation across sectors.

Third, we relate to a discussion initiated by Asker et al. (2014), exploring whether measured misallocation, namely cross sectional dispersion in the marginal productivity of capital, is consistent with dynamic efficiency. Finally, we compare our results on misallocation with results obtained in a more restrictive model, one where the generalized hazard function is constant given the sign of adjustment. This model, used e.g. by Baley and Blanco (2021), corresponds to a two-sided Calvo model with different arrival rates of positive and negative adjustment opportunities. We estimate the gains from using a fully flexible model and conduct econometric tests of the two-sided one.

6.1 Comparing measures of misallocation.

Table 5 shows the two approximations for capital misallocation: the true limit characterized in Proposition 2 and the measure based on the variance of log ARPK. Both measures are multiplied by

Table 5: two approximations for misallocation: $\mathcal{M}$ and $(1-\alpha)\alpha/2 \cdot \mathbb{V}[x]$, both multiplied by 100 to show misallocation in percent of the total capital stock.

Industry	$\mathcal{M}$	$\alpha(1-\alpha)/2 \cdot \mathbb{V}[x]$	difference
Mining and Quarrying	0.48	0.51	7.0%
	(0.02)	(0.02)	
Chemicals	0.65	0.72	11.0%
	(0.01)	(0.01)	
Metals and Machinery	0.88	1.00	14.0%
	(0.01)	(0.01)	
Food and Beverages	0.63	0.70	10.0%
	(0.01)	(0.01)	
Construction	0.92	1.00	9.0%
	(0.01)	(0.02)	
Retail	0.77	0.82	6.0%
	(0.01)	(0.01)	
Transportation	1.54	1.49	-3.0%
	(0.1)	(0.06)	
Insurance	1.98	2.63	32.0%
	(0.08)	(0.16)	
Health and Beauty	0.76	0.83	9.0%
	(0.02)	(0.02)	

100, so the units in Table 5 are percent of the total capital stock. The third column shows that, on average across sectors, differences amount to 10%. The variance-based measure almost universally overestimates misallocation. These differences indicate substantial deviations from log-normality and show that using the right approximation from Proposition 2 matters.

6.2 Misallocation over time

We next assess the dynamics of misallocation over time. To this end, we split the sample into two halves, 1983-2003 and 2003-2023, and repeat the estimation procedure for both of them. Table 6 presents the estimates for all nine sectors. We multiply $\mathcal{M}$ by 100 to present the numbers in percent of the total output. Looking across time the two halves of the sample, estimated misallocation decreases on average by around one third between 1983-2003 and 2003-2023. The rate of this decrease is similar across sectors, mostly lying within the $[30\%, 40\%]$ range.

What can this large decrease in misallocation be attributed to? Table 8 in Appendix G presents estimates of drift in capital-to-productivity ratios $\mu + \delta$ and their volatility σ across subsamples. Both drift and volatility decrease in all sectors. The fact that the decline in volatility leads to

Table 6: Misallocation across periods in percent of total capital stock ($100 \cdot \mathcal{M}_t$).

Industry	full sample	1983–2003	2003–2023	% difference
Mining and Quarrying	0.48	0.77	0.55	-28.6%
	(0.02)	(0.06)	(0.05)	
Chemicals	0.65	0.95	0.50	-47.4%
	(0.01)	(0.04)	(0.01)	
Metals and Machinery	0.88	1.15	0.71	-38.3%
	(0.01)	(0.04)	(0.01)	
Food and Beverages	0.63	0.83	0.51	-38.6%
	(0.01)	(0.04)	(0.01)	
Construction	0.92	1.29	0.81	-37.2%
	(0.01)	(0.03)	(0.01)	
Retail	0.77	1.03	0.72	-30.1%
	(0.01)	(0.01)	(0.01)	
Transportation	1.54	2.17	1.34	-38.2%
	(0.10)	(0.16)	(0.12)	
Insurance	1.98	2.91	1.93	-33.7%
	(0.08)	(0.18)	(0.07)	
Health and Beauty	0.76	1.11	0.70	-36.9%
	(0.02)	(0.06)	(0.02)	

lower misallocation is intuitive. The fact that the declining drift has the same effect is reminiscent of results in the literature on pricing frictions. In sticky price models, markup dispersion also increases in the level of inflation. In the special case with a constant generalized hazard function, monotonicity of $\mathcal{M}$ in $\mu + \delta$ is especially clear, as we show in Corollary 1.

6.3 Capital misallocation and dynamic efficiency

The capital misallocation literature often uses an indirect approach, focusing on productivity measures such as total factor productivity of revenues (TFPR). The variance of TFPR maps to output losses from primitive wedges that the econometrician includes in the model. These wedges receive a normative interpretation: they reflect unmodeled disturbances, often non-economic in nature, that lead to inefficiencies. The logic behind this mapping from TFPR dispersion to losses is traditionally static, and an "efficient" counterfactual is computed by instantly reallocating capital across firms. Asker et al. (2014) are the first to note that dispersion of TFPR (hence inefficiency) should not be fully attributed to static wedges in the presence of adjustment costs.

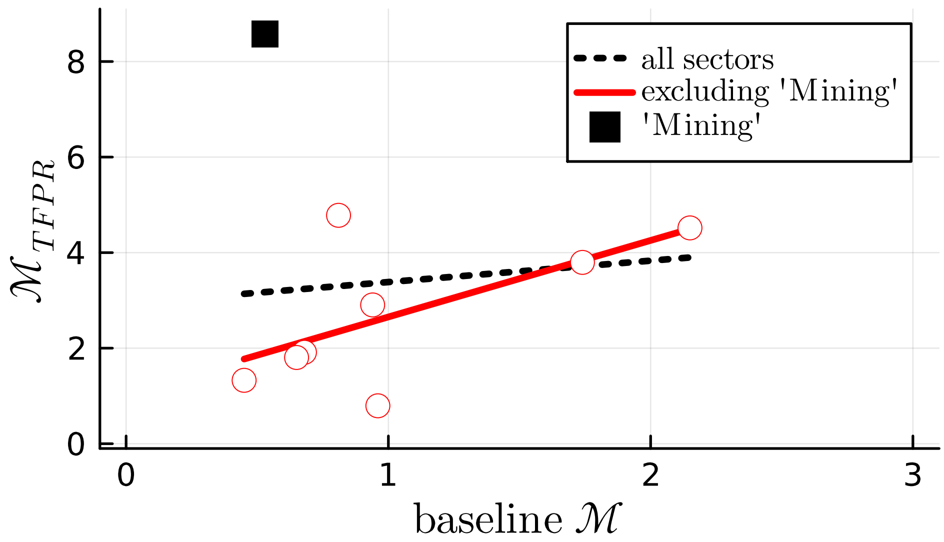


Figure 4: Comparison of $100 \cdot \mathcal{M}$ between our baseline and the Hsieh and Klenow (2009) and Calligaris et al. (2018) method using TFPR.

Our model with adjustments costs is a case in point of the warning raised by Asker et al. (2014). Indeed, in the presence of fixed costs a constrained efficient equilibrium will feature dispersion in the marginal product of capital. In this section we investigate how much of the reduced-form misallocation, as measured by the the variance of TFPR, is accounted for by the output losses from adjustment costs in our model. To that end, we compare our estimates of misallocation to the estimates of the variance of TFPR obtained earlier by Calligaris, Del Gatto, Hassan, Ottaviano, and Schivardi (2018) on the same data. Calligaris et al. (2018) use the approach pioneered by Hsieh and Klenow (2009). They measure log TFPR on the firm level and compute its population variance. To measure log TFPR, Calligaris et al. (2018) assume a Cobb-Douglas production function and estimate input elasticities using data on revenues, capital, and wage bill. We repeat the simplest version of their exercise to compare the two approaches.

Figure 4 compares our estimates of $\mathcal{M}$ to those obtained using the method in Calligaris et al. (2018), denoted by $\mathcal{M}_{TFPR}$. Our approach estimates lower degree of misallocation, although the estimates have the same order of magnitude and align in the cross-section of sectors. Excluding

"Mining", an outlier sector, the correlation across sectors for both $\mathcal{M}$ is 0.63, and the regression coefficient is 1.60 (0.81). Altogether the analysis suggests that a non-negligible part of the measured cross-section "misallocation" is consistent with the dispersion of productivity triggered by the presence of adjustment costs, and hence that caution should be exercised when inferring the presence of large macro inefficiencies based on reduced form measures of "misallocation".

6.4 Misallocation in a two-sided constant hazard model

The model we estimate is rich in parameters, accommodating generalized hazard function encoded in potentially high-dimensional vectors $\boldsymbol{\lambda}$. How important is this flexibility?

To investigate this, we fit a two-sided distribution model with a generalized hazard function that is constant on positive and negative half-lines but potentially asymmetric around zero. This is a minimally augmented variant of the Calvo model. Baley and Blanco (2021) use a version of this model, also giving the firms an opportunity to always pay a given fixed cost à la Golosov and Lucas (2007). Instead of equation (16), we solve

$$\hat{\mathcal{P}}_{\text{two-sided}} = \arg \min_{\{\mu,\sigma,\lambda_u,\lambda_d\}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{17}$$
$$\text{s.t. } \lambda_j = \lambda_d \text{ for } j > 0,\ \lambda_j = \lambda_u \text{ for } j < 0,\ n(\mathcal{P}) = N$$

Optimization is over four numbers $(\lambda_u, \lambda_d, \mu, \sigma)$, which substantively speeds up computation. In Appendix C.3, we show that the number of parameters is further reduced to three once we estimate the unrestricted generalized hazard function in each sector.

The two-sided model has a particularly simple closed-form solution for the density of log ARPK. It is a single exponential function on either side of $x = 0$, as opposed to a sum of two exponential functions, which automatically sets half of the coefficients $\boldsymbol{\xi}_1(\mathcal{P})$ and $\boldsymbol{\xi}_2(\mathcal{P})$ to zero, and the other half are all equal to the same number, following from continuity and differentiability.

Proposition 6. *Consider a two-sided distribution model and suppose it is parameterized by $\mathcal{P} = (\mu, \sigma^2, \lambda_u, \lambda_d)$, with a hazard of positive adjustments $\lambda_j = \lambda_u$ for $j < 0$ and that of negative* *adjustments* $\lambda_j = \lambda_d$ *for* $j > 0$*. The coefficients* $\boldsymbol{\eta}_1(\mathcal{P})$ *and* $\boldsymbol{\eta}_2(\mathcal{P})$ *are*

$$\eta_{1,i} = \eta_{2,j} = \frac{\left(\sqrt{\mu^2 + 2\sigma^2\lambda_u} - \mu\right)\left(\sqrt{\mu^2 + 2\sigma^2\lambda_d} + \mu\right)}{\sigma^2\left(\sqrt{\mu^2 + 2\sigma^2\lambda_u} + \sqrt{\mu^2 + 2\sigma^2\lambda_d}\right)}$$

and $\eta_{2,i} = \eta_{1,j} = 0$ *for all* $i < 0$ *and* $j > 0$*. The frequency* $n(\mathcal{P})$ *is*

$$n(\mathcal{P}) = \frac{\left(\sqrt{\mu^2 + 2\sigma^2\lambda_u} - \mu\right)\left(\sqrt{\mu^2 + 2\sigma^2\lambda_d} + \mu\right)}{2\sigma^2}$$

If the model is further restricted to a single hazard $\lambda_u = \lambda_d = \lambda$*, then* $n(\mathcal{P}) = \lambda$ *and*

$$\eta_{1,i} = \eta_{2,j} = \frac{\lambda}{\sqrt{\mu^2 + 2\sigma^2\lambda}} \text{ for } i < 0, j > 0$$

with $\eta_{2,i} = \eta_{1,j} = 0$ *for* $i < 0, j > 0$*.*

Table 7 compares the measure of misallocation in the full and the two-sided models. The two-sided model overestimates misallocation. The reason is that a flat generalized hazard can capture the overall adjustment intensity well if λ_u and λ_d are set at the right level, but it induces fatter tails in the distribution of x since $\Lambda(x)$ does not increase as x departs from x^*. Firms with large gaps do not adjust more frequently than firms with small ones, and the distribution of gaps is more dispersed.

Finally, in Appendix H, we describe a test we run to try and reject the two-sided model against our fully flexible benchmark. We use semi-parametric bootstrap. Figure A.17 and Table 9 show that our test rejects the two-sided model in 5 sectors out of 9 at the 5% level.

7 Conclusion

This paper examines capital misallocation in a framework that explicitly accounts for the lumpy and asymmetric nature of investment. We build on the generalized hazard function framework of Caballero and Engel (1999) and establish a novel procedure to identify the distribution of random fixed costs and the volatility of productivity shocks from the observed distribution of investments.

We derive the measure of aggregate misallocation for environments with non-convex adjustment costs. This measure combines moments of the cross-sectional distribution of the marginal product

Table 7: misallocation across models in percent of total capital stock ($100 \cdot \mathcal{M}$).

Industry	model with GHF	two-sided	% difference
Mining and Quarrying	0.53	1.22	130.2%
	(0.02)	(0.02)	
Chemicals	0.45	0.61	35.6%
	(0.01)	(0.01)	
Metals and Machinery	0.96	1.2	25.0%
	(0.01)	(0.01)	
Food and Beverages	0.68	0.85	25.0%
	(0.01)	(0.0)	
Construction	0.94	1.6	70.2%
	(0.02)	(0.01)	
Retail	0.65	1.35	107.7%
	(0.01)	(0.01)	
Transportation	1.74	2.12	21.8%
	(0.1)	(0.04)	
Insurance	2.15	3.07	42.8%
	(0.09)	(0.02)	
Health and Beauty	0.81	1.32	63.0%
	(0.02)	(0.01)	

of capital and generalizes variance, only collapsing to variance in the log-normal case. We show that this distribution remains non-degenerate and deviates from log-normality even in the limit of small shocks, a property that distinguishes lumpy investment models from those with convex costs. The traditional variance-based measure of misallocation is not the correct limit.

Our empirical application to a 40-year panel of Italian firms shows that, across industries, misallocation hovers between 0.5 to 2 percent of total output. We find that the standard variance-based measure yields overall similar magnitudes, typically overshooting by a margin of approximately 10%. This suggests that standard benchmarks seem robust to distributional non-normalities.

Finally, our quantitative results show that fixed costs account for a large portion of the productivity dispersion observed in most industries. This finding is reminiscent of Asker et al. (2014): much of the observed cross-sectional dispersion in marginal products is consistent with dynamic efficiency in the presence of physical investment rigidities. It suggests that what appears as static misallocation in reduced-form data may reflect the optimal, forward-looking, behavior of firms responding to idiosyncratic shocks in the presence of adjustment costs.

References

A. B. Abel and J. C. Eberly. A unified model of investment under uncertainty. The American Economic Review, 84(5):1369–1384, 1994. ISSN 00028282.

K. Adam, A. Alexandrov, and H. Weber. Inflation distorts relative prices: Theory and evidence. Number 23/2024. Deutsche Bundesbank Discussion Paper, 2024.

F. Alvarez, K. Borovičková, and R. Shimer. Decomposing duration dependence in a stopping time model. The Review of Economic Studies, 91(6):3151–3189, 12 2023. ISSN 0034-6527.

F. E. Alvarez, F. Lippi, and A. Oskolkov. The macroeconomics of sticky prices with generalized hazard functions. The Quarterly Journal of Economics, 137(2):989–1038, May 2022.

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

I. Baley and A. Blanco. Aggregate dynamics in lumpy economies. Econometrica, 89(3):1235–1264, 2021.

A. Blanco and I. Baley. The macroeconomics of irreversibility. 2024.

S. Bond and J. Van Reenen. Chapter 65 microeconometric models of investment and employment. volume 6 of Handbook of Econometrics, pages 4417–4498. Elsevier, 2007.

R. J. Caballero and E. M. R. A. Engel. Explaining investment dynamics in u.s. manufacturing: A generalized (s,s) approach. Econometrica, 67(4):783–826, July 1999.

R. J. Caballero, E. M. R. A. Engel, and J. C. Haltiwanger. Plant-Level Adjustment and Aggregate Investment Dynamics. Brookings Papers on Economic Activity, 26(2):1–54, 1995.

S. Calligaris, M. Del Gatto, F. Hassan, G. I. Ottaviano, and F. Schivardi. The productivity puzzle and misallocation: an italian perspective. Economic Policy, 33(96):635–684, 2018.

G. A. Calvo. Staggered prices in a utility-maximizing framework. Journal of monetary Economics, 12(3):383–398, 1983.

A. Cavallo, F. Lippi, and K. Miyahara. Inflation and misallocation in New Keynesian models. EIEF, Einaudi Institute for Economics and Finance, 2023.

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

J. M. David and V. Venkateswaran. The sources of capital misallocation. American Economic Review, 109(7):2531–2567, 2019.

J. M. David, H. A. Hopenhayn, and V. Venkateswaran. Information, misallocation, and aggregate productivity. The Quarterly Journal of Economics, 131(2):943–1005, 02 2016.

M. Doms and T. Dunne. Capital adjustment patterns in manufacturing plants. Review of Economic Dynamics, 1(2):409–429, 1998. ISSN 1094-2025.

C. Edmond, V. Midrigan, and D. Y. Xu. Competition, markups, and the gains from international trade. American Economic Review, 105(10):3183–3221, October 2015. doi: 10.1257/aer.20120549. URL https://www.aeaweb.org/articles?id=10.1257/aer.20120549.

M. Golosov and R. E. Lucas. Menu costs and phillips curves. Journal of Political Economy, 115 (2):171–199, 2007.

L. Guiso and F. Schivardi. Spillovers in industrial districts. The Economic Journal, 117(516): 68–93, 2007.

L. Guiso, C. Lai, and M. Nirei. An empirical study of interaction-based aggregate investment fluctuations. The Japanese Economic Review, 68(2):137–157, 2017.

H. Hopenhayn and R. Rogerson. Job turnover and policy evaluation: A general equilibrium analysis. Journal of Political Economy, 101(5):915–38, 1993.

C.-T. Hsieh and P. J. Klenow. Misallocation and manufacturing tfp in china and india*. The Quarterly Journal of Economics, 124(4):1403–1448, 11 2009.

R. E. Lucas and E. C. Prescott. Investment under uncertainty. Econometrica: Journal of the Econometric Society, pages 659–681, 1971.

K. Miyahara, A. Cavallo, and F. Lippi. Uncovering the Costs of High Inflation. mimeo, 2026.

E. Nakamura, J. Steinsson, P. Sun, and D. Villar. The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation. The Quarterly Journal of Economics, 133(4):1933–1980, 2018.

P. Ottonello and T. Winberry. Financial heterogeneity and the investment channel of monetary policy. Econometrica, 88(6):2473–2502, 2020.

D. Restuccia and R. Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. Review of Economic Dynamics, 11(4):707–720, 2008. ISSN 1094-2025.

D. Restuccia and R. Rogerson. The causes and costs of misallocation. Journal of Economic Perspectives, 31(3):151–74, August 2017. doi: 10.1257/jep.31.3.151. URL https://www.aeaweb.org/articles?id=10.1257/jep.31.3.151.

A Details on related literature

Our paper relates to a set of recent papers that attempt to uncover the frictions that underlie observed lumpy behavior, such as Baley and Blanco (2021); Alvarez, Lippi, and Oskolkov (2022); David and Venkateswaran (2019); Asker, Collard-Wexler, and Loecker (2014), and to the seminal analysis on capital misallocation by Hsieh and Klenow (2009). We briefly review these contributions below and highlight the main novel elements that are conveyed by our analysis.

The recent paper by Baley and Blanco (2021) studies a lumpy investment model where the fixed cost is either zero, with some probability, or else equal to a constant (possibly different for investment vs disinvestments). The main objective of that paper is to derive a mapping that connects the cumulative impulse response of the economy to a set of observable steady-state moments, in the spirit of a "sufficient statistic approach". Their paper presents an empirical application that, among other things, quantifies the shape of the adjustment costs from the observed size distribution of investments based on a panel of Chilean data. One difference compared to our paper is that while their main empirical application restricts the distribution of adjustment costs to have two values, our model allows for a general shape of this distribution. We show in the empirical application that the investment data reject the simple two-side hazard function for most industries.

The paper by Alvarez, Lippi, and Oskolkov (2022) investigates infrequent price adjustments through the lenses of a menu-cost model and has several elements in common with the problem studied here, mainly in its attempt to uncover the fundamental primitives of a model with "lumpy" price changes. A main methodological difference is that pricing behavior is quite symmetric, i.e. the shape of the positive adjustments is similar to the shape of the negative adjustments. While symmetry provides a reasonable approximation for price-setting behavior in low-inflation countries, such an assumption is clearly violated by the investment data. Thus, a methodological novelty of this paper is to solve for the inverse mapping in a problem where the distribution of adjustments is not symmetric.

Our paper closely relates to the literature that investigates and quantifies the capital misallocation. The seminal contribution in this area is Hsieh and Klenow (2009) who use production microdata (revenues, labor, and capital) to infer firm-specific distortions, without taking an explicit stance on their nature. The core idea is that in the absence of these distortions, revenue productivity (TFPR) should be equalized across firms within narrowly defined industries. The observed dispersion in TFPR can then be used to quantify such distortions and to calculate the potential gains in aggregate productivity that could be achieved if resources were reallocated to equalize marginal products across firms to a benchmark level (e.g., observed in the United States).

Our analysis relates to David and Venkateswaran (2019), who develop a methodology to disentangle the sources of capital misallocation, defined as dispersion in the average revenue product of capital (*arpk*). The authors augment a standard general equilibrium model of firm dynamics to include capital adjustment costs, informational frictions (imperfect knowledge about firm-level fundamentals), and other firm-specific factors, capturing unobserved heterogeneity in markups and/or production technologies, financial frictions, or institutional/policy-related distortions. Their empirical strategy measures the contribution of each force to the observed *arpk* dispersion, using a set of moments from firm-level investment and value-added data. An application to Chinese manufacturing firms reveals that while adjustment and informational frictions are significant, they explain only a modest fraction of the productivity dispersion. A substantial portion stems from other firm-specific factors, particularly a component correlated with productivity and a fixed effect.

Lastly, our work closely relates to the analysis in Asker et al. (2014) which gives center stage to non-convex adjustment costs as a possible cause behind capital misallocation. One difference is that we derive the theory-consistent measure of misallocation, showing that the intuitive variance-based measure is correct only in special cases. Another difference concerns the quantitative application, where we use a flexible GHF model that allows us to match the whole distribution of the size of adjustments (as opposed to a single moment concerning its dispersion), as well as the frequency of price adjustments. These differences lead us to estimate much smaller frictions for capital adjustment, about 0.1% of profits per adjustment, compared to the very high value of near 10% per adjustment estimated by their paper. In terms of results, our analysis is largely consistent with the one by Asker et al. (2014) and suggests that caution should be exercised when inferring the presence of large macro inefficiencies based on reduced form measures of "misallocation".

B Proofs

Proof. (of Proposition 2). As a preliminary step, consider some $\sigma \in (o, \overline{\sigma}]$ and $\mathbb{E}[e^{jx}]$ for some $j \in [0, 2\alpha]$. Specifically, take equation (5) and multiply both sides by e^{jx}:

$$\Lambda(x)e^{jx}\phi(x) = \mu e^{jx}\phi'(x) + \frac{\sigma^2}{2}e^{jx}\phi''(x) \tag{A.1}$$

Integrate both sides:

$$\begin{aligned}\mathbb{E}[\Lambda(x)e^{jx}] &= -\frac{\mu}{j}\mathbb{E}[e^{jx}] + \frac{\sigma^2}{2}\left[\lim_{x\to x^*-0}\phi'(x) - \lim_{x\to x^*+0}\phi'(x)\right]e^{jx^*} + \frac{\sigma^2}{2j^2}\mathbb{E}[e^{jx^*}] \\ &= \mathbb{E}[\Lambda(x)]e^{jx^*} + \left(\frac{\sigma^2}{2j^2} - \frac{\mu}{j}\right)\mathbb{E}[e^{jx}] \end{aligned} \tag{A.2}$$

The first line uses integration by parts, and the second line replaces the discontinuity in $\phi'(\cdot)$ at x^* by integrating the original equation (5) over the real line. Integration by parts relies on the fact that $e^{kx}\phi'(x)$ and $e^{jx}\phi(x)$ vanish at infinity, which is implied by the existence of $\mathbb{E}[e^{jx}]$ for all $j \leq 2$. Rearranging,

$$\left(\frac{\mu}{j} - \frac{\sigma^2}{2j^2}\right)\mathbb{E}[e^{jx}] + \mathbb{E}[\Lambda(x)e^{jx}] = \mathbb{E}[\Lambda(x)]e^{jx^*} \tag{A.3}$$

Take $\hat{\sigma} = \min\{\overline{\sigma}, \sqrt{\mu j}\}$. Since $\Lambda(x) \leq \overline{\lambda}$ and $\mathbb{E}[\Lambda(x)e^{jx}] \geq 0$, for all $\sigma \in [0, \hat{\sigma})$,

$$\mathbb{E}[e^{jx}] \leq \frac{2j\overline{\lambda}}{\mu}e^{jx^*} \tag{A.4}$$

Hence, for all $j \in [0, 2\alpha]$, $\mathbb{E}[e^{jx}]$ is uniformly bounded on $[0, \hat{\sigma})$.

The second step is to arrive at the expressions for the maximal and actual output

$$\log Y(g) \longrightarrow \log\mathbb{E}[z] + \alpha\hat{\mathbb{E}}[x] + \hat{\mathbb{J}}[e^{\alpha x}] \tag{A.5}$$

$$\log\hat{Y}(g) \longrightarrow \log\mathbb{E}[z] + \alpha\hat{\mathbb{E}}[x] + \alpha\hat{\mathbb{J}}[e^{x}] \tag{A.6}$$

First, consider Y_t:

$$\log Y_t = \log \mathbb{E}[z_t e^{\alpha x_t}] = \log \mathbb{E}[z_t] + \log \left(\mathbb{E}[e^{\alpha x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{\alpha x_t} \right] \right) \tag{A.7}$$

Take the last term. By the Cauchy-Schwartz inequality,

$$\left| \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{\alpha x_t} \right] \right| \leq \frac{\sqrt{\mathbb{V}[z_t]}}{\mathbb{E}[z_t]} \cdot \sqrt{\mathbb{E}[e^{2\alpha x_t}]} = \frac{\sqrt{e^{\sigma^2 t}(e^{\sigma^2 t} - 1)}}{e^{\mu t + \sigma^2 t/2}} \cdot \sqrt{\mathbb{E}[e^{2\alpha x_t}]} \tag{A.8}$$

Since t is fixed and $\mathbb{E}[e^{2\alpha x_t}]$ is uniformly bounded, this term approaches 0 as $\sigma \longrightarrow 0$. The other term, $\mathbb{E}[e^{\alpha x_t}]$, approaches $\hat{\mathbb{E}}[e^{\alpha x_t}]$, since the moment generating function evaluated at $j = \alpha$ is continuous in σ on the segment that includes 0. Hence,

$$\log Y_t \longrightarrow \log \mathbb{E}[z_t] + \log \hat{\mathbb{E}}[e^{\alpha x_t}] = \log \mathbb{E}[z_t] + \alpha \hat{\mathbb{E}}[x_t] + \hat{\mathbb{J}}[e^{\alpha x_t}] \tag{A.9}$$

Analogously, take

$$\begin{aligned} \log \hat{Y}_t &= (1-\alpha) \log \mathbb{E}[z_t] + \alpha \log \mathbb{E}[z_t e^{x_t}] \\ &= \log \mathbb{E}[z_t] + \alpha \log \left(\mathbb{E}[e^{x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{x_t} \right] \right) \end{aligned} \tag{A.10}$$

Applying the same procedure, we get

$$\log \hat{Y}_t \longrightarrow \log \mathbb{E}[z_t] + \alpha \log \hat{\mathbb{E}}[e^{x_t}] = \log \mathbb{E}[z_t] + \alpha \hat{\mathbb{E}}[x_t] + \alpha \hat{\mathbb{J}}[e^{x_t}] \tag{A.11}$$

For K_t, we get

$$\log K_t = \frac{\log \hat{Y}_t - (1-\alpha) \log \mathbb{E}[z_t]}{\alpha} \longrightarrow \log \mathbb{E}[z_t] + \hat{\mathbb{E}}[x_t] + \hat{\mathbb{J}}[e^{x_t}] \tag{A.12}$$

Misallocation then converges to

$$\mathcal{M}_t \longrightarrow \alpha \hat{\mathbb{J}}[e^{x_t}] - \hat{\mathbb{J}}[e^{\alpha x_t}] \tag{A.13}$$

Finally, to obtain the limiting density, consider equation (5) for $\sigma = 0$:

$$\Lambda(x) \hat{\phi}(x) = (\mu + \delta) \hat{\phi}'(x) \tag{A.14}$$

Rearranging, $[\log \hat{\phi}(x)]' = \Lambda(x)/(\mu + \delta)$. Without shocks, no firms are to the right of x^*. Hence,

$$\hat{\phi}(x) = \hat{a} \exp \left(-\frac{1}{\mu + \delta} \int_x^{x^*} \Lambda(t) dt \right) \tag{A.15}$$

Here $\hat{a}$ insures that $\hat{\phi}(\cdot)$ integrates to one:

$$\frac{1}{\hat{a}} = \int_{-\infty}^{x^*} \exp\left(-\frac{1}{\mu+\delta}\int_x^{x^*} \Lambda(t)dt\right) dx \tag{A.16}$$

This completes the proof. □

Proof. (of Corollary 1). In addition to the results on capital stock and misallocation, we will also derive the limits for the actual and maximum output:

$$\log Y(g) \longrightarrow \log \mathbb{E}[z] + \alpha E + \alpha\sqrt{V} - \log\left(1+\alpha\sqrt{V}\right) \tag{A.17}$$

$$\log \hat{Y}(g) \longrightarrow \log \mathbb{E}[z] + \alpha E + \alpha\sqrt{V} - \alpha\log\left(1+\sqrt{V}\right) \tag{A.18}$$

Start with the limiting distribution. Since $\Lambda(x) \equiv \lambda$,

$$\hat{\phi}(x) = \frac{\exp\left(-\frac{\lambda(x^*-x)}{\mu+\delta}\right)}{\int_{-\infty}^{x^*}\exp\left(-\frac{\lambda(x^*-x)}{\mu+\delta}\right)dx} = \frac{\lambda}{\mu+\delta}\exp\left(-\frac{\lambda(x^*-x)}{\mu+\delta}\right) \tag{A.19}$$

The expectation and variance corresponding to this distribution are

$$V \equiv \hat{\mathbb{V}}[x] = \left(\frac{\mu+\delta}{\lambda}\right)^2 \tag{A.20}$$

$$E \equiv \hat{\mathbb{E}}[x] = x^* - \frac{\mu+\delta}{\lambda} = x^* - \sqrt{V} \tag{A.21}$$

The limiting Jensen correction is

$$\hat{\mathbb{J}}[e^x] = \log\hat{\mathbb{E}}[e^x] - \hat{\mathbb{E}}[x] = \log\left(\frac{\lambda}{\lambda+\mu+\delta}\right) + \frac{\mu+\delta}{\lambda} = \sqrt{V} - \log\left(1+\sqrt{V}\right) \tag{A.22}$$

$$\hat{\mathbb{J}}[e^{\alpha x}] = \log\hat{\mathbb{E}}[e^{\alpha x}] - \alpha\hat{\mathbb{E}}[x] = \log\left(\frac{\lambda}{\lambda+\alpha(\mu+\delta)}\right) + \frac{\alpha(\mu+\delta)}{\lambda} = \alpha\sqrt{V} - \log\left(1+\alpha\sqrt{V}\right) \tag{A.23}$$

Hence,

$$\log K_t \longrightarrow \log\mathbb{E}[z] + x^* - \log\left(1+\sqrt{V}\right) \tag{A.24}$$

$$\log Y_t \longrightarrow \log\mathbb{E}[z] + \alpha x^* - \log\left(1+\alpha\sqrt{V}\right) \tag{A.25}$$

$$\log \hat{Y}_t \longrightarrow \log\mathbb{E}[z] + \alpha x^* - \alpha\log\left(1+\sqrt{V}\right) \tag{A.26}$$

$$\mathcal{M}_t \longrightarrow \log\left(1+\alpha\sqrt{V}\right) - \alpha\log\left(1+\sqrt{V}\right) \tag{A.27}$$

This completes the proof. □

Proof. (of Proposition 3). Plug $\Lambda(x) = \hat{\Lambda}(x)/\kappa$ into equation (5):

$$\hat{\Lambda}(x)\phi(x) = (\mu + \delta)\kappa\phi'(x) + \frac{\kappa\sigma^2}{2}\phi''(x) \tag{A.28}$$

Define ℓ_k and e_k as follows:

$$\ell_k = \int (x - x^*)^k \hat{\Lambda}(x)\phi(x)dx \tag{A.29}$$

$$e_k = \int (x - x^*)^k \phi(x)dx \tag{A.30}$$

Integrating equation (A.28) with $(x - x^*)^k$ for $k \geq 2$,

$$\ell_k = -\kappa(\mu + \delta)ke_{k-1} + \frac{\kappa\sigma^2}{2}k(k-1)e_{k-2} \tag{A.31}$$

For even $k \geq 0$, $\ell_k/\lambda \geq e_k \geq 0$. For odd $k \geq 1$, use Cauchy-Schwartz inequality to obtain

$$|e_k|^2 \leq e_{k+1}e_{k-1} \leq \frac{\ell_{k+1}e_{k-1}}{\lambda} = \frac{\kappa\sigma^2}{2\lambda}k(k+1)e_{k-1}^2 - \frac{\kappa(\mu+\delta)(k+1)}{\lambda}e_k e_{k-1} \tag{A.32}$$

Using $e_{k-1} \geq 0$,

$$\left|\frac{e_k}{\sigma^b e_{k-1}}\right|^2 \leq \sigma^{2-b}k(k+1)\frac{\kappa\sigma^{-b}}{2\lambda} - (\mu+\delta)(k+1)\frac{\kappa\sigma^{-b}}{\lambda}\frac{e_k}{\sigma^b e_{k-1}} \tag{A.33}$$

Since $\kappa\sigma^{-b}$ has a positive limit and $b \leq 2$, $e_k/(\sigma^b e_{k-1})$ is bounded as a function of σ. This implies that $e_k = O(\sigma^b e_{k-1})$ if k is odd. For even $k \geq 2$,

$$e_k \leq -\frac{\kappa(\mu+\delta)}{\lambda}ke_{k-1} + \frac{\kappa\sigma^2}{2\lambda}k(k-1)e_{k-2} \tag{A.34}$$

If $e_{k-2} \neq 0$, this implies

$$\frac{e_k}{\sigma^{2b}e_{k-2}} \leq -k\frac{\kappa\sigma^{-b}(\mu+\delta)}{\lambda}\frac{e_{k-1}}{\sigma^b e_{k-2}} + \sigma^{2-b}k(k-1)\frac{\kappa\sigma^{-b}}{2\lambda} \tag{A.35}$$

As already established, $e_{k-1}/(\sigma^b e_{k-2})$ is bounded, so $\kappa\sigma^{-b}$ and σ^{2-b} having a finite limit implies that $e_k/(\sigma^{2b}e_{k-2})$ is bounded too. This means that $e_k = O(\sigma^{2b}e_{k-2})$.

Together with $e_0 = 1$, the fact that $e_k = O(\sigma^b e_{k-1})$ for odd $k \geq 1$ and $e_k = O(\sigma^{2b}e_{k-2})$ for even $k \geq 2$ implies that $e_k = O(\sigma^{kb})$ for all k. In particular, $\mathbb{V}[x_t] = O(\sigma^{2b})$. If $\mathbb{E}[e^{x_t}]$ exists, then

$$\begin{aligned}\mathcal{M}_t &= \alpha\log\left(\mathbb{E}[e^{x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]}e^{x_t}\right]\right) - \log\left(\mathbb{E}[e^{\alpha x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]}e^{\alpha x_t}\right]\right)\\ &= \alpha\log\left(\mathbb{E}[e^{\hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\hat{x}_t}\right]\right) - \log\left(\mathbb{E}[e^{\alpha\hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\alpha\hat{x}_t}\right]\right)\end{aligned} \tag{A.36}$$

Here $\hat{z}_t \equiv (z_t - \mathbb{E}[z_t])/\mathbb{E}[z_t]$ and $\hat{x}_t \equiv x_t - x^*$. Expanding $e^{\hat{x}_t}$ and $e^{\alpha\hat{x}_t}$,

$$\mathcal{M}_t = \alpha \log\left(1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[\hat{x}_t^k(1+\hat{z}_t)]}{k!}\right) - \log\left(1 + \sum_{k=1}^{\infty} \frac{\alpha^k \mathbb{E}[\hat{x}_t^k(1+\hat{z}_t)]}{k!}\right) \tag{A.37}$$

Applying the Cauchy-Schwartz inequality,

$$\left|\mathbb{E}[\hat{x}_t^k \hat{z}_t]\right| \leq \sqrt{\mathbb{V}[\hat{z}_t]} \cdot \sqrt{\mathbb{E}[\hat{x}_t^{2k}]} = O\left(\sigma^{1+kb}\right) \tag{A.38}$$

Hence,

$$\begin{aligned}
\mathcal{M}_t &= \alpha \log\left(1 + \mathbb{E}[\hat{x}_t] + \mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{1}{2}\mathbb{E}[\hat{x}_t^2] + O(\sigma^{1+2b})\right) \\
&\quad - \log\left(1 + \alpha\mathbb{E}[\hat{x}_t] + \alpha\mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{\alpha^2}{2}\mathbb{E}[\hat{x}_t^2] + O(\sigma^{1+2b})\right) \\
&= \frac{\alpha(1-\alpha)}{2}\mathbb{E}[\hat{x}_t^2] - \frac{\alpha(1-\alpha)}{2}(\mathbb{E}[\hat{x}_t])^2 + \sum_{k=3}^{\infty}\frac{(-1)^k(\alpha-\alpha^k)}{k!}(\mathbb{E}[\hat{x}_t])^k + O(\sigma^{1+2b}) \\
&= \frac{\alpha(1-\alpha)}{2}\mathbb{V}[\hat{x}_t] + O(\sigma^{3b}) + O(\sigma^{1+2b}) = \frac{\alpha(1-\alpha)}{2}\mathbb{V}[\hat{x}_t] + O\left(\sigma^{\min\{3b,1+2b\}}\right)
\end{aligned} \tag{A.39}$$

We will additionally derive the limits for the actual and maximal output:

$$\log Y(g) = \log \mathbb{E}[z] + \alpha\mathbb{E}[x] + O\left(\sigma^{\min\{1+b,2b\}}\right) \tag{A.40}$$

$$\log \hat{Y}(g) = \log \mathbb{E}[z] + \alpha\mathbb{E}[x] + O\left(\sigma^{\min\{1+b,2b\}}\right) \tag{A.41}$$

First, consider actual output Y_t:

$$\begin{aligned}
\log Y_t &= \log(\mathbb{E}[z_t e^{\alpha x_t}]) = \log \mathbb{E}[z_t] + \log\left(\mathbb{E}[e^{\alpha x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]}e^{\alpha x_t}\right]\right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(\mathbb{E}[e^{\alpha\hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\alpha\hat{x}_t}\right]\right)
\end{aligned} \tag{A.42}$$

Expanding $e^{\alpha\hat{x}_t}$,

$$\begin{aligned}
\log Y_t &= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(1 + \sum_{k=1}^{\infty}\frac{\alpha^k\mathbb{E}[\hat{x}_t^k(1+\hat{z}_t)]}{k!}\right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(1 + \alpha\mathbb{E}[\hat{x}_t] + \alpha\mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{\alpha^2}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^{1+2b}\right)\right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \alpha\mathbb{E}[\hat{x}_t] + O\left(\sigma^{\max 1+b,2b}\right) = \alpha\mathbb{E}[x_t] + O\left(\sigma^{\max 1+b,2b}\right)
\end{aligned} \tag{A.43}$$

Similarly, for $\hat{Y}_t$,

$$\begin{aligned}
\log \hat{Y}_t &= (1-\alpha)\log \mathbb{E}[z_t] + \alpha \log(\mathbb{E}[z_t e^{x_t}]) = \log \mathbb{E}[z_t] + \alpha \log \left(\mathbb{E}[e^{x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{x_t} \right] \right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log \left(\mathbb{E}[e^{\hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\hat{x}_t} \right] \right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log \left(1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[\hat{x}_t^k (1+\hat{z}_t)]}{k!} \right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log \left(1 + \mathbb{E}[\hat{x}_t] + \mathbb{E}[\hat{x}_t \hat{z}_t] + \frac{1}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^{1+2b}\right) \right) \\
&= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \mathbb{E}[\hat{x}_t] + O\left(\sigma^{\max 1+b, 2b}\right) = \log \mathbb{E}[z_t] + \alpha \mathbb{E}[x_t] + O\left(\sigma^{\max 1+b, 2b}\right) \quad \text{(A.44)}
\end{aligned}$$

Finally, for K_t,

$$\log K_t = \frac{\log \hat{Y}_t - (1-\alpha)\log \mathbb{E}[z_t]}{\alpha} = \log \mathbb{E}[z_t] + \mathbb{E}[x_t] + O\left(\sigma^{\max 1+b, 2b}\right) \quad \text{(A.45)}$$

This completes the proof. □

Proof. (of Proposition 4). We will proceed in three steps. Denote by $\bar{x}(\sigma)$ the point where $\imath(\bar{x}(\sigma);\sigma) = \mu + \delta$. First, we will prove that the stationary measures defined for positive σ weakly converge to the measure fully concentrated in $\overline{x}$, where $\overline{\imath}(\overline{x}) = \mu + \delta$, as $\sigma \longrightarrow 0$. Second, we will prove that $\mathbb{E}[(x - \overline{x}(\sigma))^k] = O(\sigma^k)$ for all integer $k \geq 0$. Third, we will apply this to the measure of misallocation and aggregate capital stock.

Step 1 (stationary density). Denoting by $\phi(\cdot;\sigma)$ the stationary density for a fixed σ, integrate the KFE:

$$\phi(x;\sigma) = \psi(\sigma) \exp\left\{ \frac{2\theta(x;\sigma)}{\sigma^2} \right\} \quad \text{(A.46)}$$

Here $\psi(\sigma)$ ensures that $\phi(\cdot;\sigma)$ integrates to one and the function $\theta(\cdot;\sigma)$ is

$$\theta(x;\sigma) = \int_{\bar{x}(\sigma)}^{x} (\imath(t;\sigma) - \mu - \delta) dt \quad \text{(A.47)}$$

Note that this function is increasing for $x < \bar{x}(\sigma)$ and decreasing for $x > \bar{x}(\sigma)$.

Fix $a < \bar{x} < b$. Since $\imath(\cdot;\sigma)$ converges uniformly to $\overline{\imath}(\cdot)$, the root of $\imath(x;\sigma) - \mu - \delta$ converges to that of $\overline{\imath}(x) - \mu - \delta$: $\bar{x}(\sigma) \longrightarrow \bar{x}$. Hence, there exists a σ_1 such that for all $\sigma < \sigma_1$, $b - \bar{x}(\sigma) > (b - \bar{x})/2$ and $\bar{x}(\sigma) - a > (\bar{x} - a)/2$. For $\sigma \in [0, \sigma_1]$, the associated measure $M(\cdot;\sigma)$ satisfies

$$M([\bar{x}(\sigma), b];\sigma) \geq \psi(\sigma) \exp\left\{ \frac{2\theta(b;\sigma)}{\sigma^2} \right\} \cdot \frac{b - \bar{x}}{2} \quad \text{(A.48)}$$

$$M([a, \bar{x}(\sigma)];\sigma) \geq \psi(\sigma) \exp\left\{ \frac{2\theta(a;\sigma)}{\sigma^2} \right\} \cdot \frac{\bar{x} - a}{2} \quad \text{(A.49)}$$

For all $x > b$, since $\imath(\cdot;\sigma)$ is decreasing

$$\phi(x;\sigma) < \psi(\sigma)\exp\left\{\frac{2\theta(b;\sigma)}{\sigma^2}\right\}\cdot\exp\left\{\frac{2(x-b)(\imath(b;\sigma)-\mu-\delta)}{\sigma^2}\right\} \tag{A.50}$$

Similarly, for all $x < a$,

$$\phi(x;\sigma) < \psi(\sigma)\exp\left\{\frac{2\theta(a;\sigma)}{\sigma^2}\right\}\cdot\exp\left\{\frac{2(a-x)(\mu+\delta-\imath(a;\sigma))}{\sigma^2}\right\} \tag{A.51}$$

Fix $\varepsilon > 0$ such that $\bar{\imath}(b)-\varepsilon-\mu-\delta > 0$ and $\mu+\delta-\bar{\imath}(a)-\varepsilon > 0$. There exists a $\sigma_2 > 0$ such that for all $\sigma < \sigma_2$, $\imath(a;\sigma)-\mu-\delta > \bar{\imath}(a)-\varepsilon-\mu-\delta$ and $\mu+\delta-\imath(b;\sigma) > \mu+\delta-\bar{\imath}(b)-\varepsilon$. For $\sigma \in [0,\sigma_2]$,

$$M([b,\infty);\sigma) < \psi(\sigma)\exp\left\{\frac{2\theta(b;\sigma)}{\sigma^2}\right\}\cdot\frac{\sigma^2}{2(\mu+\delta-\bar{\imath}(b)-\varepsilon)} \tag{A.52}$$

$$M((-\infty,a];\sigma) < \psi(\sigma)\exp\left\{\frac{2\theta(a;\sigma)}{\sigma^2}\right\}\cdot\frac{\sigma^2}{2(\bar{\imath}(a)-\varepsilon-\mu-\delta)} \tag{A.53}$$

Hence, for $\sigma \in [0,\min\{\sigma_1,\sigma_2\}]$,

$$\frac{M([b,\infty);\sigma)}{M([\bar{x}(\sigma),b];\sigma)} \leq \frac{\sigma^2}{(\mu+\delta-\bar{\imath}(b)-\varepsilon)(b-\bar{x})} \tag{A.54}$$

$$\frac{M((-\infty,a];\sigma)}{M([a,\bar{x}(\sigma)];\sigma)} \leq \frac{\sigma^2}{(\bar{\imath}(a)-\varepsilon-\mu-\delta)(\bar{x}-a)} \tag{A.55}$$

This implies

$$\frac{M(\mathbb{R}/[a,b];\sigma)}{(1-M(\mathbb{R}/[a,b];\sigma))} \leq \sigma^2\max\left\{\frac{1}{(\bar{\imath}(b)-\varepsilon-\mu-\delta)(b-\bar{x})},\frac{1}{(\mu+\delta-\varepsilon-\bar{\imath}(a))(\bar{x}-a)}\right\} \tag{A.56}$$

As $\sigma \longrightarrow 0$, the measure of any open interval containing $\bar{x}$ approaches one. Hence, the measure of any open set containing $\bar{x}$ converges to one. By Portmanteau's lemma, this is equivalent to the weak convergence of $\mu(\cdot;\sigma)$ to $\mu(\cdot)$ that is concentrated in $\bar{x}$.

Step 2 (magnitude of moments). To prove the second part, fix an even $k \geq 2$ and consider a different measure $g_k(\cdot;\sigma)$ given by

$$g_k(x;\sigma) = \xi(\sigma)(x-\bar{x}(\sigma))^k\exp\left\{\frac{2\theta(x;\sigma)}{\sigma^2}\right\} \tag{A.57}$$

Here again $\xi(\sigma)$ ensures that this function integrates to one. This density function is equal to zero at $\bar{x}(\sigma)$ and has two peaks at $x_+(\sigma) > \bar{x}(\sigma)$ and $x_-(\sigma) < \bar{x}(\sigma)$ given by the solutions to the following equation over t:

$$(\mu+\delta-\imath(t;\sigma))(t-\bar{x}(\sigma)) = \frac{\sigma^2 k}{2} \tag{A.58}$$

Fix $a < \bar{x} < b$ again. Since $\imath(\cdot;\sigma)$ is strictly decreasing for all σ and converges uniformly to $\bar{\imath}(\cdot)$, which is also strictly decreasing, both $x_+(\sigma)$ and $x_-(\sigma)$ converge to $\bar{x}$ as $\sigma \longrightarrow 0$. Hence, there exists $\sigma_1 > 0$ such that for all $\sigma \in [0, \sigma_1]$ it holds that $(b - x_+(\sigma)) > (b - \bar{x})/2$ and $(x_-(\sigma) - a) > (\bar{x} - a)/2$. Hence, for the measure $M_k(\cdot; \sigma)$ associated with $g_k(\cdot; \sigma)$

$$M_k((a, \bar{x}(\sigma)]; \sigma) > g_k(a; \sigma) \cdot \frac{\bar{x} - a}{2} \tag{A.59}$$

$$M_k([\bar{x}(\sigma), b); \sigma) > g_k(b; \sigma) \cdot \frac{b - \bar{x}}{2} \tag{A.60}$$

Now consider a function $w(\cdot; \sigma)$ given by

$$w(x; \sigma) = \xi(\sigma) \exp\left\{ \frac{\theta(x; \sigma)}{\sigma^2} \right\} \tag{A.61}$$

Take the derivative of the ratio $\phi(x; \sigma)/w(x; \sigma)$:

$$\left[\frac{g_k(x; \sigma)}{w(x; \sigma)} \right]' = \left(\frac{(x - \bar{x}(\sigma))(\imath(x; \sigma) - \mu - \delta)}{\sigma^2} + k \right) (x - \bar{x}(\sigma))^{k-1} \exp\left\{ \frac{\theta(x; \sigma)}{\sigma^2} \right\} \tag{A.62}$$

This derivative is negative for all $x > \hat{x}_+(\sigma)$, where $\hat{x}_+(\sigma) > \bar{x}(\sigma)$ is the higher root of the following equation over t:

$$(t - \bar{x}(\sigma))(\mu + \delta - \imath(t; \sigma)) = \sigma^2 k \tag{A.63}$$

Similarly, the derivative of the ratio $g_k(x; \sigma)/w(x; \sigma)$ is positive for $x < \hat{x}_-(\sigma)$, where $\hat{x}_-(\sigma) < \bar{x}(\sigma)$ is the lower root of this equation. From the uniform convergence of $\imath(\cdot; \sigma)$ to $\bar{\imath}(\cdot)$ and the strict monotonicity of $\imath(\cdot; \sigma)$ and $\bar{\imath}(\cdot)$ it follows that there is a $\sigma_2 > 0$ such that $\hat{x}_+(\sigma) < b$ and $\hat{x}_-(\sigma) > a$ for all $\sigma \in [0, \sigma_2]$. For these $\sigma \in [0, \sigma_2]$,

$$g_k(x; \sigma) \leq w(x; \sigma) \frac{g_k(b; \sigma)}{w(b; \sigma)} \text{ for } x \geq b \tag{A.64}$$

$$g_k(x; \sigma) \leq w(x; \sigma) \frac{g_k(a; \sigma)}{w(a; \sigma)} \text{ for } x \leq a \tag{A.65}$$

For the associated measure $M_k(\cdot; \sigma)$, this implies that

$$M_k((-\infty, a]; \sigma) \leq g_k(a; \sigma) \frac{\sigma^2}{(\imath(a; \sigma) - \mu - \delta)} \tag{A.66}$$

$$M_k([b, \infty); \sigma) \leq g_k(b; \sigma) \frac{\sigma^2}{(\mu + \delta - \imath(b; \sigma))} \tag{A.67}$$

By the same argument as before, fix $\varepsilon > 0$ such that $\bar{\imath}(a) - \varepsilon - \mu - \delta > 0$ and $\mu + \delta - \bar{\imath}(b) - \varepsilon > 0$. There exists a $\sigma_3 > 0$ such that for all $\sigma < \sigma_3$, $\imath(a; \sigma) - \mu - \delta > \bar{\imath}(a) - \varepsilon - \mu - \delta$ and $\mu + \delta - \imath(b; \sigma) >$

$\mu + \delta - \bar{\imath}(b) - \varepsilon$. For $\sigma \in [0, \min\{\sigma_2, \sigma_3\}]$,

$$M_k((-\infty, a]; \sigma) \leq g_k(a; \sigma) \frac{\sigma^2}{(\bar{\imath}(a) - \varepsilon - \mu - \delta)} \tag{A.68}$$
$$M_k([b, \infty); \sigma) \leq g_k(b; \sigma) \frac{\sigma^2}{(\mu + \delta - \bar{\imath}(b) - \varepsilon)} \tag{A.69}$$

Hence, for all $\sigma \in [0, \min\{\sigma_1, \sigma_2, \sigma_3\}]$,

$$\frac{M_k((-\infty; a]; \sigma)}{M_k((a, \bar{x}(\sigma)]; \sigma)} \leq \frac{2\sigma^2}{(\bar{x} - a)(\bar{\imath}(a) - \varepsilon - \mu - \delta)} \tag{A.70}$$
$$\frac{M_k([b, \infty); \sigma)}{M_k([\bar{x}(\sigma), b); \sigma)} \leq \frac{2\sigma^2}{(b - \bar{x})(\mu + \delta - \bar{\imath}(b) - \varepsilon)} \tag{A.71}$$

By the same argument as before, this means that the measure of all open intervals containing $\bar{x}$ converges to one. Hence, the measure associated with $g_k(\cdot; \sigma)$ weakly converges to the measure concentrated at $x = \bar{x}$.

Now consider the following notation:

$$m_k(\sigma) = \mathbb{E}[(x - \bar{x}(\sigma))^k | \sigma] \tag{A.72}$$
$$e_k(\sigma) = \mathbb{E}[(\mu + \delta - \imath(x; \sigma))(x - \bar{x}(\sigma))^k | \sigma] \tag{A.73}$$

Integrating the KFE, we have, for all $k \geq 1$,

$$e_k = \frac{\sigma^2 k}{2} m_{k-1} \tag{A.74}$$

Take odd k. Fix an interval $(\bar{x}(\sigma) - c, \bar{x}(\sigma) + c)$. Since the derivative of $\imath(\cdot; \sigma)$ is bounded away from zero, there exists a number $\varsigma > 0$ such that $|\imath(x; \sigma) - \mu - \delta| \geq \varsigma(x - \bar{x}(\sigma))$ for all $x \in (\bar{x}(\sigma) - c, \bar{x}(\sigma) + c)$. Hence,

$$e_k \geq \varsigma \int_{\bar{x}(\sigma)-c}^{\bar{x}(\sigma)+c} (x - \bar{x}(\sigma))^{k+1} \phi(x; \sigma) dx = \varsigma M_{k+1}((\bar{x}(\sigma) - c, \bar{x}(\sigma) + c); \sigma) \cdot m_{k+1} \tag{A.75}$$

Here $M_{k+1}(\cdot; \sigma)$ is the measure associated with the density $g_{k+1}(\cdot; \sigma)$ from above. Since the set $(\bar{x}(\sigma) - c, \bar{x}(\sigma) + c)$ contains $\bar{x}$ for σ small enough, its measure converges to one. Hence, for σ small enough, $e_k \geq \varsigma m_{k+1}/2$, and hence $m_{k+1} \leq \sigma^2 \varsigma^{-1} m_{k-1}$ for all odd k. Since $m_0 = 1$, we have $m_k = O(\sigma^k)$ for all even $k \geq 0$. For odd $k \geq 1$, using Cauchy-Schwartz inequality leads to

$$|m_k| \leq \sqrt{m_2} \cdot \sqrt{m_{2k-2}} = O(\sigma^k) \tag{A.76}$$

Step 3 (misallocation and aggregate capital). Finally, consider misallocation: $\mathcal{M}_t = \log \hat{Y}_t =$

$\log Y_t$. First, consider actual output Y_t:

$$\begin{aligned}\log Y_t &= \log(\mathbb{E}[z_t e^{\alpha x_t}]) = \log \mathbb{E}[z_t] + \log\left(\mathbb{E}[e^{\alpha x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{\alpha x_t}\right]\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(\mathbb{E}[e^{\alpha \hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\alpha \hat{x}_t}\right]\right) \end{aligned} \tag{A.77}$$

Expanding $e^{\alpha \hat{x}_t}$,

$$\begin{aligned}\log Y_t &= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(1 + \sum_{k=1}^{\infty} \frac{\alpha^k \mathbb{E}[\hat{x}_t^k(1+\hat{z}_t)]}{k!}\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \log\left(1 + \alpha\mathbb{E}[\hat{x}_t] + \alpha\mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{\alpha^2}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^3\right)\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \alpha\mathbb{E}[\hat{x}_t] + \alpha\mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{\alpha^2}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^3\right) \end{aligned} \tag{A.78}$$

Similarly, for $\hat{Y}_t$,

$$\begin{aligned}\log \hat{Y}_t &= (1-\alpha)\log \mathbb{E}[z_t] + \alpha \log(\mathbb{E}[z_t e^{x_t}]) = \log \mathbb{E}[z_t] + \alpha \log\left(\mathbb{E}[e^{x_t}] + \mathbb{E}\left[\frac{z_t - \mathbb{E}[z_t]}{\mathbb{E}[z_t]} e^{x_t}\right]\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log\left(\mathbb{E}[e^{\hat{x}_t}] + \mathbb{E}\left[\hat{z}_t e^{\hat{x}_t}\right]\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log\left(1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[\hat{x}_t^k(1+\hat{z}_t)]}{k!}\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \alpha \log\left(1 + \mathbb{E}[\hat{x}_t] + \mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{1}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^3\right)\right) \\ &= \log \mathbb{E}[z_t] + \alpha x^* + \alpha\mathbb{E}[\hat{x}_t] + \alpha\mathbb{E}[\hat{x}_t\hat{z}_t] + \frac{\alpha}{2}\mathbb{E}[\hat{x}_t^2] + O\left(\sigma^3\right) \end{aligned} \tag{A.79}$$

For misallocation, we have

$$\mathcal{M}_t = \frac{\alpha(1-\alpha)}{2}\mathbb{V}[x_t] + O(\sigma^3) \tag{A.80}$$

Finally, for K_t,

$$\log K_t = \frac{\log \hat{Y}_t - (1-\alpha)\log \mathbb{E}[z_t]}{\alpha} = \log \mathbb{E}[z_t] + \mathbb{E}[x_t] + O\left(\sigma^2\right) \tag{A.81}$$

This completes the proof. □

Proof. (of Proposition 5). The coefficients $\boldsymbol{\eta}_1 = \{\eta_{1,j}\}_{-\mathrm{u} \le j \le \mathrm{d}, j \neq 0}$ and $\boldsymbol{\eta}_2 = \{\eta_{2,j}\}_{-\mathrm{u} \le j \le \mathrm{d}, j \neq 0}$

satisfy the following continuity conditions:

$$f_{j-1}(x_j) = f_j(x_j) \text{ for } j \in \{-\mathbb{u}+1, \dots -1\} \tag{A.82}$$
$$f_{j+1}(x_j) = f_j(x_j) \text{ for } j \in \{1, \dots \mathbb{d}-1\} \tag{A.83}$$
$$f_{-1}(0) = f_1(0) \tag{A.84}$$
$$f'_{j-1}(x_j) = f'_j(x_j) \text{ for } j \in \{-\mathbb{u}+1, \dots -1\} \tag{A.85}$$
$$f'_{j-1}(x_j) = f'_j(x_j) \text{ for } j \in \{1, \dots \mathbb{d}-1\} \tag{A.86}$$
$$\lim_{x\to\infty} f_{\mathbb{d}}(x) = \lim_{x\to-\infty} f_{-\mathbb{u}}(x) = 0 \tag{A.87}$$
$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{A.88}$$

The jump in the first derivative of $f(\cdot)$ at zero is due to the "reinjection" of firms: they adjust capital gaps discretely, continually arriving at zero. The size of the jump in $f'(\cdot)$ is

$$\lim_{x\to-0} f'_{-1}(x) - \lim_{x\to+0} f'_1(x) = \frac{2N}{\sigma^2} \tag{A.89}$$

This can be shown by integrating equation (5) over the real line and using the statistical fact that $q(-x)N = L(x)f(x)$.

The conditions in equation (A.82), equation (A.83), equation (A.84), equation (A.85), equation (A.86), equation (A.87), and equation (A.88) provide $2(\mathbb{u}+\mathbb{d})$ linear equations to solve for $2(\mathbb{u}+\mathbb{d})$ unknowns $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. Conditions (A.82)-(A.88) contain exactly $2(U+D)$ equations that are linear in $\{\eta_{1,j}, \eta_{2,j}\}$. Equation (A.88) can be rewritten as

$$1 = \sum_{j=-U}^{-1} \left(\frac{\eta_{1,j}}{\xi_{1,j}} (e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}) + \frac{\eta_{2,j}}{\xi_{2,j}} (e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j}) \right)$$
$$+ \sum_{j=1}^{D} \left(\frac{\eta_{1,j}}{\xi_{1,j}} (e^{\xi_{1,j}x_j} - e^{\xi_{1,j}x_{j-1}}) + \frac{\eta_{2,j}}{\xi_{2,j}} (e^{\xi_{2,j}x_j} - e^{\xi_{2,j}x_{j-1}}) \right) \tag{A.90}$$

Using this and equations (A.82)-(A.87), we can construct a $2(U+D) \times 2(U+D)$ matrix $\mathbb{A}$ and a $2(U+D) \times 1$ vector $\mathbf{b}$ as follows. First, for $-U+1 \leq j \leq -1$, set $k = j+U+1$ and

$$\mathbb{A}_{2k-1,2k-3} = -e^{\xi_{1,j-1}x_j} \tag{A.91}$$
$$\mathbb{A}_{2k-1,2k-2} = -e^{\xi_{2,j-1}x_j} \tag{A.92}$$
$$\mathbb{A}_{2k-1,2k-1} = e^{\xi_{1,j}x_j} \tag{A.93}$$
$$\mathbb{A}_{2k-1,2k} = e^{\xi_{2,j}x_j} \tag{A.94}$$
$$\mathbb{A}_{2k,2k-3} = -\xi_{1,j-1}e^{\xi_{1,j-1}x_j} \tag{A.95}$$
$$\mathbb{A}_{2k,2k-2} = -\xi_{2,j-1}e^{\xi_{2,j-1}x_j} \tag{A.96}$$
$$\mathbb{A}_{2k,2k-1} = \xi_{1,j}e^{\xi_{1,j}x_j} \tag{A.97}$$
$$\mathbb{A}_{2k,2k} = \xi_{2,j}e^{\xi_{2,j}x_j} \tag{A.98}$$

Next, for $1 \leq j \leq D-1$, set $k = j + U$ and

$$\mathbb{A}_{2k-1,2k-1} = e^{\xi_{1,j}x_j} \tag{A.99}$$
$$\mathbb{A}_{2k-1,2k} = e^{\xi_{2,j}x_j} \tag{A.100}$$
$$\mathbb{A}_{2k-1,2k+1} = -e^{\xi_{1,j+1}x_j} \tag{A.101}$$
$$\mathbb{A}_{2k-1,2k+2} = -e^{\xi_{2,j+1}x_j} \tag{A.102}$$
$$\mathbb{A}_{2k,2k-1} = \xi_{1,j}e^{\xi_{1,j}x_j} \tag{A.103}$$
$$\mathbb{A}_{2k,2k} = \xi_{2,j}e^{\xi_{2,j}x_j} \tag{A.104}$$
$$\mathbb{A}_{2k,2k+1} = -\xi_{1,j+1}e^{\xi_{1,j+1}x_j} \tag{A.105}$$
$$\mathbb{A}_{2k,2k+2} = -\xi_{2,j+1}e^{\xi_{2,j+1}x_j} \tag{A.106}$$

The remaining rows are $\{1, 2, 2(D+U)-1, 2(D+U)\}$. They encode (A.87), (A.84), and (A.88). For this, set $\mathbb{A}_{1,2} = 1$, $\mathbb{A}_{2(U+D),2(U+D)-1} = 1$, $(\mathbb{A}_{2,2U-1}, \mathbb{A}_{2,2U}, \mathbb{A}_{2,2U+1}, \mathbb{A}_{2,2U+2}) = (1, 1, -1, -1)$, and

$$\mathbb{A}_{2(U+D)-1,2k-1} = \frac{1}{\xi_{1,j}}(e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}\mathbb{1}\{j > -U\}),\ k = j+U+1,\ -U \leq j \leq -1 \tag{A.107}$$
$$\mathbb{A}_{2(U+D)-1,2k} = \frac{1}{\xi_{2,j}}(e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j}),\ k = j+U+1,\ -U \leq j \leq -1 \tag{A.108}$$
$$\mathbb{A}_{2(U+D)-1,2k-1} = \frac{1}{\xi_{1,j}}(e^{\xi_{1,j}x_j} - e^{\xi_{1,j}x_{j-1}}),\ k = j+U,\ 1 \leq j \leq D \tag{A.109}$$
$$\mathbb{A}_{2(U+D)-1,2k} = \frac{1}{\xi_{2,j}}(e^{\xi_{2,j}x_j}\mathbb{1}\{j < D\} - e^{\xi_{2,j}x_{j-1}}),\ k = j+U,\ 1 \leq j \leq D \tag{A.110}$$

All entries of the vector $\mathbf{b}$ are equal to zero, except for $\mathbf{b}_{2(U+D)-1}$, because this entry corresponds to the "integrating" row of $\mathbb{A}$. The coefficients η are recovered by solving for the vector $\boldsymbol{\eta}$ satisfying $\mathbb{A}\,\boldsymbol{\eta} = \mathbf{b}$ and setting $(\eta_{1,j}, \eta_{2,j}) = (\boldsymbol{\eta}_{2(j+U)+1}, \boldsymbol{\eta}_{2(j+U)+2})$ for $-U \leq j \leq -1$ and $(\eta_{1,j}, \eta_{2,j}) = (\boldsymbol{\eta}_{2(j+U)-1}, \boldsymbol{\eta}_{2(j+U)})$ for $1 \leq j \leq D$. $\square$

Proof. (of Corollary 2). It follows from inspecting equation (14) and equation (15).

Proof. (of Proposition 6). Recall that the density of gaps on a segment j is

$$\tilde{f}_j(x) = \eta_{1,j}e^{\xi_{1,j}x} + \eta_{2,j}e^{\xi_{2,j}x} \tag{A.111}$$

Consider a two-sided model. There are two segments in this case, $(-\infty, 0)$ with $\Lambda(x) = \lambda_u$ and $(-\infty, 0)$ with $\Lambda(x) = \lambda_d$. In the negative gap territory, the powers $\xi_{1,-1}$ and $\xi_{2,-1}$ are

$$\{\xi_{1,-1}, \xi_{2,-1}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u}}{\sigma^2} \tag{A.112}$$

This implies $\eta_{2,-1} = 0$ so that $f(\cdot)$ does not diverge at $-\infty$.

For positive gaps, the powers $\xi_{1,1}$ and $\xi_{2,1}$ are

$$\{\xi_{1,1}, \xi_{2,1}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d}}{\sigma^2} \tag{A.113}$$

This implies $\eta_{1,1} = 0$ so that $f(\cdot)$ does not diverge at ∞.

The remaining two coefficients are $\eta_{2,1}$ and $\eta_{1,-1}$. They are equal to each other, $\eta_{2,1} = \eta_{2,1} = \eta$, which is implied by the continuity at $x = 0$. To get the remaining condition on η, recall that $\tilde{f}(\cdot)$ should integrate to one over the real line:

$$\eta \left(\frac{1}{\xi_{1,-1}} - \frac{1}{\xi_{2,1}} \right) = 1 \tag{A.114}$$

Plugging,

$$\eta = \frac{\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d} + (\mu+\delta)\right)\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} - (\mu+\delta)\right)}{\sigma^2\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} + \sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d}\right)} \tag{A.115}$$

To get the aggregate frequency N, integrate the accounting identity $Nq(-x) = \tilde{f}(x)\tilde{\Lambda}(x)$:

$$N = \eta \left(\frac{\lambda_u}{\xi_{1,-1}} - \frac{\lambda_d}{\xi_{2,1}} \right) \tag{A.116}$$

Plugging η,

$$N = \frac{\lambda_u \xi_{2,1} - \lambda_d \xi_{1,-1}}{\xi_{2,1} - \xi_{1,-1}} \tag{A.117}$$

Now use the fact that $\xi_{2,1}$ and $\xi_{1,-1}$ satisfy the following quadratic equations:

$$\frac{\sigma^2}{2}\xi_{2,1}^2 + (\mu+\delta)\xi_{2,1} = \lambda_d \tag{A.118}$$

$$\frac{\sigma^2}{2}\xi_{1,-1}^2 + (\mu+\delta)\xi_{1,-1} = \lambda_u \tag{A.119}$$

Plugging λ_d and λ_u from these expressions,

$$N = -\frac{\sigma^2}{2}\xi_{2,1}\xi_{1,-1} = \frac{\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d} + (\mu+\delta)\right)\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} - (\mu+\delta)\right)}{2\sigma^2}$$

This is the statement of the proposition. $\square$

C Details of the model

In this section, we provide the details for the Hamilton-Jacobi-Bellman equation of the firms. We do it for a slightly more general version of the model that allows for costly investment even when the occasional Poisson opportunity has not arrived.

As in the main text, let $i \in \{u, d\}$ denote where the firm is relative to the optimal point. If $i = u$, the firm is below the optimal capital and would like to adjust upwards. If $i = d$, it will not adjust upwards but will disinvest should the opportunity arrive. The Bellman equation of the firm in the steady state is

$$rV(k,z) = z^{1-\alpha}k^{\alpha} - \delta k \partial_k V(k,z) + \left(\mu + \frac{\sigma^2}{2}\right) z\partial_z V(k,z) + \frac{\sigma^2}{2} z^2 \partial_{zz} V(k,z)$$
$$+ \sum\nolimits_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{V(y^*z,z) - y^*z - (V(k,z)-k) - \psi z, 0\} dG_i(\psi) \tag{A.120}$$

The optimality condition is $\partial_k V(y^*z, z) = 1$. The value-matching conditions at all cutoffs are

$$V(y^*z,z) - y^*z - V(k,z) + k = \boldsymbol{\psi}^i(k/z), \text{ for } i \in \{u,d\} \tag{A.121}$$

Suppose, for generality, that the firm always has a costly option to invest. It can pay $\boldsymbol{\psi}_u$ to adjust up and $\boldsymbol{\psi}_d$ to adjust down. This introduces ultimate cutoffs y^d and y^u, at which the firm adjusts even if the occasional Poisson opportunity has not arrived. Of course, these cutoffs are infinite when $\boldsymbol{\psi}_u$ and $\boldsymbol{\psi}_d$ are, in which case we are back to the benchmark model. Instead of these cutoffs, the boundary conditions on $V(k,z)$ at $k=0$ and $k=\infty$ are the following: the homogeneous part of the solution $V(k,z)$ goes to zero as $k \longrightarrow 0$ and $k \longrightarrow \infty$ for all fixed z.

If $\boldsymbol{\psi}_u$ and $\boldsymbol{\psi}_d$ are finite, the smooth-pasting conditions at these ultimate cutoffs are

$$\partial_z V(y^*z,z) - \partial_z V(y^iz,z) = \boldsymbol{\psi}_i, \text{ for } i \in \{u,d\} \tag{A.122}$$

These smooth-pasting conditions hold if the ultimate cutoffs are not zero and infinite, respectively. They are normally internal if the fixed cost the firm can always pay is finite. Otherwise, there is no decision to be taken, so the smooth-pasting conditions need not hold.

Now introduce a function $v(\cdot)$ given by

$$V(k,z) = zv(k/z) + k \tag{A.123}$$

The function $v(\cdot)$ measures the value of the firm net of the market value of its capital and adjusted for productivity. Note the following relations for its derivatives:

$$\partial_k V(k,z) = v'\left(\frac{k}{z}\right) + 1 \tag{A.124}$$
$$\partial_z V(k,z) = v\left(\frac{k}{z}\right) - \frac{k}{z} v'\left(\frac{k}{z}\right) \tag{A.125}$$
$$\partial_{zz} V(k,z) = \frac{k^2}{z^3} v''\left(\frac{k}{z}\right) \tag{A.126}$$

Plugging the relations for derivatives of $V(\cdot)$ into equation (A.120) and denoting $y = k/z$,

$$rzv(y) + rzy = zy^{\alpha} - \delta zyv'(y) - \delta zy + \left(\mu + \frac{\sigma^2}{2}\right)(zv(y) - zyv'(y)) + \frac{\sigma^2}{2} y^2 v''(y)$$
$$+ z\sum\nolimits_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\} dG_i(\psi) \tag{A.127}$$

Dividing everything by z and denoting $\rho = r - \mu - \sigma^2/2$ and $\nu = r + \delta$,

$$\rho v(y) = y^\alpha - \nu y + (\rho - \nu) y v'(y) + \frac{\sigma^2}{2} y^2 v''(y)$$
$$+ \sum\nolimits_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\} dG_i(\psi) \tag{A.128}$$

The optimality condition is $v'(y^*) = 0$. The value-matching and smooth-pasting are

$$v(y^*) - v(y) = \boldsymbol{\psi}^i \tag{A.129}$$
$$v(y^*) - y^* v'(y^*) - v(y^i) + y^i v'(y^i) = \boldsymbol{\psi}^i \tag{A.130}$$

for $i \in \{u, d\}$. Since $v'(y^*) = 0$, these two equations together imply $v'(y^i) = 0$ for both $i \in \{u, d\}$. Taking the derivative of equation (A.128) with respect to y and denoting $u(y) = v'(y)$,

$$\rho u(y) = \alpha y^{\alpha-1} - \nu + (\rho - \nu) u(y) + (\rho - \nu + \sigma^2) y u'(y) + \frac{\sigma^2}{2} y^2 u''(y)$$
$$= \sum_{i=u,d} \mathbb{1}_i \gamma_i \int \mathbb{1}\{v(y^*) - v(y) \geq \psi\} u(y) dG_i(\psi) \tag{A.131}$$

Rearranging,

$$\left(\nu + \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i\left(v(y^*) - v(y)\right)\right) u(y) = \alpha y^{\alpha-1} - \nu + (\rho - \nu + \sigma^2) y u'(y) + \frac{\sigma^2}{2} y^2 u''(y)$$

The generalized hazard function here is

$$\lambda(y) = \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i\left(v(y^*) - v(y)\right) \tag{A.132}$$

Plugging,

$$(\nu + \lambda(y)) u(y) = \alpha y^{\alpha-1} - \nu + (\rho - \nu + \sigma^2) y u'(y) + \frac{\sigma^2}{2} y^2 u''(y) \tag{A.133}$$

Boundary conditions for this equation are $u(y^u) = -\boldsymbol{\psi}^u$, $u(y^d) = -\boldsymbol{\psi}^d$, and $u(y^*) = 0$. Again, there are no smooth-pasting conditions $u(y^u) = -\boldsymbol{\psi}^u$ and $u(y^d) = -\boldsymbol{\psi}^d$ if the normal fixed cost is infinite. In this case, the firm does not take a decision that would call for smooth pasting. Instead, the homogeneous part of $u(\cdot)$ converges to zero as $y \longrightarrow 0$ or $y \longrightarrow \infty$.
Proof. (of Proposition 1). Take equation (A.133) and make a change of variable $y \mapsto x \equiv \log(y)$. Define $U(x) \equiv u(y(x))$ and $\Lambda(x) \equiv \lambda(y(x))$. With $u'(y) = U'(x)/y$ and $U''(x) = (U''(x) - U'(x))/y^2$, equation (A.133) transforms into

$$(\nu + \Lambda(x)) U(x) = \alpha e^{(\alpha-1)x} - \nu + \left(\rho - \nu + \frac{\sigma^2}{2}\right) U'(x) + \frac{\sigma^2}{2} U''(x) \tag{A.134}$$

Recalling that $\rho = r - \mu - \sigma^2/2$ and $\nu = r + \delta$,

$$(\nu + \Lambda(x))U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu+\delta)U'(x) + \frac{\sigma^2}{2}U''(x) \tag{A.135}$$

To arrive at the equation for $\Lambda(x)$, use equation (A.132):

$$\begin{aligned}\Lambda(x) &= \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i \left(v(e^{x^*}) - v(e^x)\right) = \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i \left(\int_{e^x}^{e^{x^*}} u(y)dy \right) \\ &= \sum_{i=u,d} \mathbb{1}_i \gamma_i G_i \left(\int_x^{x^*} U(t)e^t dt \right) \end{aligned} \tag{A.136}$$

This completes the proof. □

C.1 Renting capital instead of owning

In this subsection, we establish equivalence between the problems of a firm that owns capital and one that rents it at an interest rate $r+\delta$. Suppose a firm rents capital and faces the same adjustment costs. When there is no adjustment, capital simply depreciates at a rate δ, and the firm makes rental payments $(r+\delta)k_t$ per unit of time. When it decides to change the capital stock instead of simply letting it depreciate, it has to pay an adjustment cost ψz_t. The multiplier ψ is random.

Specifically, firms always have the option to pay fixed costs $\psi_d z_t$ or $\psi_u z_t$ and adjust downwards and upwards, respectively. With a Poisson intensity γ_d, they get an opportunity to draw a lower adjustment cost ψ that they can pay for adjusting down. This cost is distributed with a cumulative distribution function $G_d(\cdot)$ on $[0, \psi_d]$. For adjusting up, they get an opportunity to draw a lower cost with a Poisson intensity γ_u, and these costs are distributed according to $G_u(\cdot)$ on $[0, \psi_u]$.

The firms again follow a policy described by cutoffs. Conditional on adjusting, they always choose $K = k^* z_t$. When $k > y^* z_t$, firms only adjust down, and do this if and only if the corresponding cost reduction arrives and the new value drawn ψ satisfies $\psi \le \boldsymbol{\psi}^d(k/z)$. Here $\boldsymbol{\psi}^d(\cdot)$ is a cutoff function. When $k > y^* z_t$, firms only adjust up, and do this if and only if the corresponding cost reduction arrives and the new value drawn ψ satisfies $\psi \le \boldsymbol{\psi}^u(k/z)$. The function $\boldsymbol{\psi}^d(\cdot)$ maps $[y^*, y^d]$ to $[0, \psi^d]$, and $\boldsymbol{\psi}^u(\cdot)$ maps $[y^u, y^*]$ to $[0, \psi^u]$. The thresholds y^d and y^u correspond to values of capital at which the firms adjust even without a cost reduction.

The Bellman equation describing the value $\overline{V}(k,z)$ of such a firm is

$$\begin{aligned} r\overline{V}(k,z) &= z^{1-\alpha}k^\alpha - (r+\delta)k - \delta k \partial_k \overline{V}(k,z) + \left(\mu + \frac{\sigma^2}{2}\right) z \partial_z \overline{V}(k,z) + \frac{\sigma^2}{2} z^2 \partial_{zz} \overline{V}(k,z) \\ &+ \sum\nolimits_{i=u,d} \mathbb{1}_i \gamma_i \int \max\{\overline{V}(y^* z, z) - \overline{V}(k,z) - \psi z, 0\} dG_i(\psi) \end{aligned} \tag{A.137}$$

The optimality condition is $\partial_k \overline{V}(y^* z, z) = 0$. The value-matching conditions are

$$\overline{V}(y^* z, z) - \overline{V}(k,z) = \boldsymbol{\psi}^i(k/z)z, \text{ for } i \in \{u,d\} \tag{A.138}$$

The smooth-pasting conditions are

$$\partial_z \overline{V}(y^* z, z) - \partial_z \overline{V}(y^i z, z) = \psi^i, \text{ for } i \in \{u, d\} \tag{A.139}$$

Define a productivity-adjusted value function $v(\cdot)$ by

$$\overline{V}(k, z) = zv(k/z) \tag{A.140}$$

Note the following relations for derivatives:

$$\partial_k \overline{V}(k, z) = v'\left(\frac{k}{z}\right) \tag{A.141}$$

$$\partial_z \overline{V}(k, z) = v\left(\frac{k}{z}\right) - \frac{k}{z}v'\left(\frac{k}{z}\right) \tag{A.142}$$

$$\partial_{zz} \overline{V}(k, z) = \frac{k^2}{z^3}v''\left(\frac{k}{z}\right) \tag{A.143}$$

Plugging this into equation (A.137) and denoting $y = k/z$,

$$\begin{aligned} rzv(y) = zy^\alpha - (r+\delta)zy - \delta zyv'(y) + \left(\mu + \frac{\sigma^2}{2}\right)(v(y) - yv'(y)) + \frac{\sigma^2}{2}zy^2v''(y) \\ + z\sum\nolimits_{i=u,d} \mathbb{1}_i\gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\}dG_i(\psi) \end{aligned} \tag{A.144}$$

Dividing everything by z and denoting $\rho = r - \mu - \sigma^2/2$ and $\nu = r + \delta$,

$$\rho v(y) = y^\alpha - \nu y + (\rho - \nu)yv'(y) + \frac{\sigma^2}{2}y^2v''(y) + \sum_{i=u,d} \mathbb{1}_i\gamma_i \int \max\{v(y^*) - v(y) - \psi, 0\}dG_i(\psi)$$

This equation coincides with equation (1). Notice that $v'(y^*) = v'(y^u) = v'(y^d) = 0$, so all decisions are exactly the same as in the baseline.

C.2 Shape of the adjustment histogram

In general, it is not possible to characterize the shape of the histogram $\boldsymbol{H}(\hat{\mathcal{P}})$ locally: for any j, the size of H_{-j} depends on the entire vector $\hat{\boldsymbol{\lambda}}$, not just on λ_j and its neighbors. However, in some case it is possible to characterize the behavior of the ratio $H_{-j}/H_{-(j-1)}$ depending on $(\lambda_j, \lambda_{j-1})$. Specifically, consider $j = -\mathbb{u}$: the largest negative capital gaps (and the largest positive adjustments). Suppose that the constraint $\lambda_{-\mathbb{u}} \geq \lambda_{-(\mathbb{u}-1)}$ is binding. How does the shape of the tail of the investment histogram, as described by the ratio $H_{\mathbb{u}}/H_{\mathbb{u}-1}$, change when the constraint is relaxed on the margin, and $\lambda_{-\mathbb{u}} = \lambda_{-(\mathbb{u}-1)}$ turns into $\lambda_{-\mathbb{u}} > \lambda_{-(\mathbb{u}-1)}$? The following proposition shows that the tail becomes thicker.

Proposition 7. *Assume $\delta = 0$ without loss an fix $\mathcal{P} = (\mu, \sigma^2, \boldsymbol{\lambda})$, where the vector $\boldsymbol{\lambda}$ is such* *that* $\lambda_{-\mathbb{u}} = \lambda_{-(\mathbb{u}-1)}$*. Consider the histogram of adjustments* $\boldsymbol{H}(\mathcal{P})$ *generated by* $\mathcal{P}$*. It holds that*

$$\left.\frac{\partial(H_{\mathbb{u}}/H_{\mathbb{u}-1})}{\partial\lambda_{-\mathbb{u}}}\right|_{\lambda_{-\mathbb{u}}=\lambda_{-(\mathbb{u}-1)}} > 0$$

As the monotonicity constraint on $\boldsymbol{\lambda}$ relaxes, there are two offsetting effects. First, adjustment in the tail intensifies, raising $H_{\mathbb{u}}/H_{\mathbb{u}-\mathbb{1}}$. Second, because of more intensive adjustment, the quantity of firms still surviving with a capital gap in the tail decreases, lowering $H_{\mathbb{u}}/H_{\mathbb{u}-\mathbb{1}}$. Proposition 7 shows that the former dominates on the margin. Importantly, this statement only applies to the relative intensity of adjustment: everything else, including the frequency of investment and the levels of all $\{H_j\}$ change in this thought experiment too.

Proof. (of Proposition 7). Consider a generalized hazard function $\Lambda(\cdot)$ with $\Lambda(x) = \lambda_l$ on $(-\infty, x_{-\mathbb{u}+1}]$ and $\Lambda(x) = \lambda_r$ on $(x_{-\mathbb{u}+1}, x_{-\mathbb{u}+2}]$. The density of capital gaps $f(\cdot)$ around $x_{-\mathbb{u}+1}$ is

$$f(x) = \begin{cases} \eta_l e^{\xi_l x}, & x \leq x_{-\mathbb{u}+1} \\ \eta_{r,1} e^{\xi_{r,1} x} + \eta_{r,2} e^{\xi_{r,2} x}, & x \in [x_{-\mathbb{u}+1}, x_{-\mathbb{u}+2}] \end{cases} \tag{A.145}$$

Here

$$\xi_l = \frac{\sqrt{\mu^2 + 2\sigma^2\lambda_l} - \mu}{\sigma^2} \tag{A.146}$$

$$\xi_{r,1} = \frac{\sqrt{\mu^2 + 2\sigma^2\lambda_r} - \mu}{\sigma^2} \tag{A.147}$$

$$\xi_{r,2} = \frac{-\sqrt{\mu^2 + 2\sigma^2\lambda_r} - \mu}{\sigma^2} \tag{A.148}$$

Denote $\overline{x} = x_{-\mathbb{u}+1}$. At $x = \overline{x}$, the density must be continuous and differentiable:

$$\eta_l e^{\xi_l \overline{x}} = \eta_{r,1} e^{\xi_{r,1}\overline{x}} + \eta_{r,2} e^{\xi_{r,2}\overline{x}} \tag{A.149}$$

$$\xi_l \eta_l e^{\xi_l \overline{x}} = \xi_{r,1}\eta_{r,1} e^{\xi_{r,1}\overline{x}} + \xi_{r,2}\eta_{r,2} e^{\xi_{r,2}\overline{x}} \tag{A.150}$$

This implies

$$\eta_{r,1} = \eta_l \cdot \frac{\xi_{r,2} - \xi_l}{\xi_{r,2} - \xi_{r,1}} \cdot e^{(\xi_l - \xi_{r,1})\overline{x}} \tag{A.151}$$

$$\eta_{r,2} = \eta_l \cdot \frac{\xi_l - \xi_{r,1}}{\xi_{r,2} - \xi_{r,1}} \cdot e^{(\xi_l - \xi_{r,2})\overline{x}} \tag{A.152}$$

The histogram of capital gaps is given by

$$F_l = \frac{\eta_l}{\xi_l} e^{\xi_l \overline{x}} \tag{A.153}$$

$$F_r = \frac{\eta_{r,1}}{\xi_{r,1}} \left(1 - e^{\xi_{r,1}\overline{x}}\right) + \frac{\eta_{r,2}}{\xi_{r,2}} \left(1 - e^{\xi_{r,2}\overline{x}}\right) \tag{A.154}$$

The histogram of adjustments is

$$H_l = \frac{\lambda_l \eta_l}{N \xi_l} e^{\xi_l \overline{x}} \tag{A.155}$$

$$H_r = \frac{\lambda_r \eta_{r,1}}{N \xi_{r,1}} \left(1 - e^{\xi_{r,1}\overline{x}}\right) + \frac{\lambda_r \eta_{r,2}}{N \xi_{r,2}} \left(1 - e^{\xi_{r,2}\overline{x}}\right) \tag{A.156}$$

Here N is the adjustment frequency. Define the ratio $R = H_r/H_l$:

$$\begin{aligned} R(\lambda_l, \lambda_r) &= \frac{\lambda_r}{\lambda_l} \cdot \left[\frac{\xi_l}{\xi_{r,1}} \cdot \frac{\eta_{r,1}}{\eta_l} \cdot \frac{1 - e^{\xi_{r,1}\overline{x}}}{e^{\xi_l \overline{x}}} + \frac{\xi_l}{\xi_{r,2}} \cdot \frac{\eta_{r,2}}{\eta_l} \cdot \frac{1 - e^{\xi_{r,2}\overline{x}}}{e^{\xi_l \overline{x}}}\right] \\ &= \frac{\lambda_r}{\lambda_l} \cdot \left[\frac{\xi_l}{\xi_{r,1}} \cdot \frac{\xi_{r,2} - \xi_l}{\xi_{r,2} - \xi_{r,1}} \cdot \left(e^{-\xi_{r,1}\overline{x}} - 1\right) + \frac{\xi_l}{\xi_{r,2}} \cdot \frac{\xi_l - \xi_{r,1}}{\xi_{r,2} - \xi_{r,1}} \cdot \left(e^{-\xi_{r,2}\overline{x}} - 1\right)\right] \end{aligned} \tag{A.157}$$

The constrained benchmark is $\lambda_l = \lambda_r$. In this benchmark, $\xi_{r,1} = \xi_l$ and

$$R(\lambda_l, \lambda_r)\Big|_{\lambda_l = \lambda_r} = e^{-\xi_{r,1}\overline{x}} - 1 \tag{A.158}$$

Take the derivative of $R(\cdot)$ with respect to λ_l around $\lambda_l = \lambda_r$:

$$\begin{aligned} \frac{\partial R(\lambda_l, \lambda_r)}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} &= -\frac{1}{\lambda_r} \cdot \left(e^{-\xi_{r,1}\overline{x}} - 1\right) + \frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} \cdot \frac{1}{\xi_{r,1}} \cdot \left(e^{-\xi_{r,1}\overline{x}} - 1\right) \\ &\quad - \frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} \cdot \frac{1}{\xi_{r,2} - \xi_{r,1}} \cdot \left(e^{-\xi_{r,1}\overline{x}} - 1\right) \\ &\quad + \frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} \cdot \frac{\xi_{r,1}}{\xi_{r,2}} \cdot \frac{1}{\xi_{r,2} - \xi_{r,1}} \cdot \left(e^{-\xi_{r,2}\overline{x}} - 1\right) \end{aligned} \tag{A.159}$$

Rewriting,

$$\begin{aligned} \frac{\partial R(\lambda_l, \lambda_r)}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} &= \left[\frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} \cdot \frac{\xi_{r,2} - 2\xi_{r,1}}{\xi_{r,1}(\xi_{r,2} - \xi_{r,1})} - \frac{1}{\lambda_r}\right] \left(e^{-\xi_{r,1}\overline{x}} - 1\right) \\ &\quad + \frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} \cdot \frac{\xi_{r,1}}{\xi_{r,2}(\xi_{r,2} - \xi_{r,1})} \left(e^{-\xi_{r,2}\overline{x}} - 1\right) \end{aligned}$$

The second summand is negative because $\xi_{r,2} < 0$, $\xi_{r,1} > 0$, and $\overline{x} < 0$. Consider the first summand. First, $e^{-\xi_{r,1}\overline{x}} - 1 > 0$. Second,

$$\frac{\partial \xi_l}{\partial \lambda_l}\Big|_{\lambda_l = \lambda_r} = \frac{1}{\sqrt{\mu^2 + 2\sigma^2 \lambda_l}}\Big|_{\lambda_l = \lambda_r} = \frac{1}{\sqrt{\mu^2 + 2\sigma^2 \lambda_r}} \tag{A.160}$$

Denote $z = \sqrt{\mu^2 + 2\sigma^2\lambda_r}$. Plugging this into the expression in square brackets,

$$\frac{\partial \xi_l}{\partial \lambda_l}\frac{\xi_{r,2} - 2\xi_{r,1}}{\xi_{r,1}(\xi_{r,2} - \xi_{r,1})} - \frac{1}{\lambda_r} = \frac{\sigma^2(3z-\mu)}{2z^2(z-\mu)} - \frac{1}{\lambda_r} = \frac{\sigma^2(3z-\mu)}{2z^2(z-\mu)} - \frac{2\sigma^2}{z^2-\mu^2}$$
$$= \frac{\sigma^2}{2z^2(z^2-\mu^2)} \cdot ((3z-\mu)(z+\mu) - 4z^2) = -\frac{\sigma^2(z-\mu)}{2z^2(z+\mu)} < 0 \tag{A.161}$$

This proves that

$$\left.\frac{\partial R(\lambda_l, \lambda_r)}{\partial \lambda_l}\right|_{\lambda_l = \lambda_r} < 0 \tag{A.162}$$

This completes the proof. □

C.3 Degrees of freedom when switching between models.

We next make a note on the number of degrees of freedom that the econometrician has while switching between models. Estimating the model on the same data implies, in particular, that they will be fitted to the same frequency of adjustment. This establishes a connection between the recovered generalized hazard functions.

Proposition 8. *Consider two models that generate the same frequency of adjustments. Let the vectors $\boldsymbol{\lambda}_1 = \{\lambda_{j,1}\}$ and $\boldsymbol{\lambda}_2 = \{\lambda_{k,2}\}$ encode their respective generalized hazard functions, and let $\boldsymbol{H}_1 = \{H_{j,1}\}$ and $\boldsymbol{H}_2 = \{H_{k,2}\}$ be the histograms of adjustments they induce. It holds that*

$$\sum_j \frac{H_{j,1}}{\lambda_{j,1}} = \sum_k \frac{H_{k,2}}{\lambda_{k,2}}$$

In words, the harmonic average of the generalized hazard function, weighted by the histogram of adjustments, is the same across models. In particular, having estimated the full specification, we know the harmonic average of the two numbers λ_u and λ_d in the two-sided model, which leaves us with three parameters to estimate instead of four.

For a simpler example, consider estimating a pure Calvo (1983) model with generalized hazard $\Lambda(x) \equiv \lambda$ and a two-sided model with hazards (λ_u, λ_d). The proposition implies

$$\frac{1}{\lambda} = \frac{\mathbb{P}\{\text{positive investment}\}}{\lambda_u} + \frac{1 - \mathbb{P}\{\text{positive investment}\}}{\lambda_d}$$

Here λ_u is the upward adjustment hazard in the two-sided model, and λ_d is the downward one. Estimating any model also gives the econometrician the hazard in the Calvo model for free.

Proof. (of Proposition 8). Start with the accounting identity

$$Nq(-x) = \Lambda(x)f(x) \tag{A.163}$$

Take the j-th segment on which $\Lambda(x)$ is constant and integrate the over it:

$$\frac{H_j}{\lambda_j} = \frac{1}{N}\int_j f(x)dx \tag{A.164}$$

Now summing over all these segments,

$$\sum_j \frac{H_j}{\lambda_j} = \frac{1}{N} \tag{A.165}$$

Since N is the same in the two models by assumption, the harmonic averages of $\Lambda(x)$ weighted with adjustment probabilities on the left-hand side are the same too. □

D Recovering marginal value $U(\cdot)$

Recall that

$$\nu(x)U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu+\delta)U'(x) + \frac{\sigma^2}{2}U''(x) \tag{A.166}$$

Here a slight abuse of notation is $\nu(x) = \nu + \Lambda(x)$.

Now suppose $\Lambda(\cdot)$ is piece-wise constant. Specifically, let there be $\mathbb{u} + \mathbb{d} + 1$ nodes given by $\{\hat{x}_j\} = \{\hat{x}_j + x^*\}$ such that $\Lambda(\cdot)$ is constant on any $(\hat{x}_{j-1} + x^*, \hat{x}_j + x^*)$ for $1 \le j \le \mathbb{d}$, any $(\hat{x}_j + x^*, \hat{x}_{j+1} + x^*)$ for $-\mathbb{u} \le j \le -1$:

- $\Lambda(x) = \lambda_j$ on $(\hat{x}_{j-1} + x^*, \hat{x}_j + x^*)$ for $1 \le j \le \mathbb{d}$
- $\Lambda(x) = \lambda_j$ on $(\hat{x}_j + x^*, \hat{x}_{j+1} + x^*)$ for $-\mathbb{u} \le j \le -1$

The leftmost and rightmost nodes are infinite: $\hat{x}_{-\mathbb{u}} = -\infty$ and $\hat{x}_D = \infty$. Denote $\nu_j = \nu + \lambda_j$ and let $U_j(\cdot)$ be the part of $U(\cdot)$ defined on the segment j. The solution to equation (A.166) on any segment j is

$$U_j(x) = \eta_{1,j}e^{\xi_{1,j}x} + \eta_{2,j}e^{\xi_{2,j}x} + \theta_{1,j}e^{(\alpha-1)x} + \theta_{2,j} \tag{A.167}$$

The non-homogeneous part can be recovered immediately:

$$\theta_{1,j} = \frac{\alpha}{\nu_j - (1-\alpha)(\mu+\delta) - (1-\alpha)^2\sigma^2/2} \tag{A.168}$$

$$\theta_{2,j} = -\frac{\nu}{\nu_j} \tag{A.169}$$

The homogeneous part is a sum of two terms with four parameters per segment in total. Exponent parameters are given by

$$\{\xi_{1,j}, \xi_{2,j}\} = \frac{\mu + \delta \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2\nu_j}}{\sigma^2} \tag{A.170}$$

The weights $\eta_{1,j}$ and $\eta_{2,j}$, combining into $2(\mathbb{u}+\mathbb{d})$ unknowns, have to be recovered from continuity and differentiability conditions on $U(\cdot)$. Specifically, for all j corresponding to finite nodes, meaning $-\mathbb{u} < j < \mathbb{d}$ including $j = 0$,

$$U_{j-1}(\hat{x}_j + x^*) = U_j(\hat{x}_j + x^*) \tag{A.171}$$

$$U'_{j-1}(\hat{x}_j + x^*) = U'_j(\hat{x}_j + x^*) \tag{A.172}$$

This yields $2(\mathbb{u}+\mathbb{d}-1)$ conditions. Two other equations are $\eta_{1,-\mathbb{u}} = 0$ and $\eta_{2,\mathbb{d}} = 0$ ensuring that the homogeneous part of $U(\cdot)$ does not blow up at $-\infty$ and ∞. The non-homogeneous part, which represents the marginal instantaneous returns to capital, is infinite at $-\infty$, while the homogeneous part represents the marginal value of the real option coming from the future evolution of capital stock and is finite.

Given $U(\cdot)$, it is straightforward to recover $u(y) = U(\ln(y))$ and integrate $v(y) - v(y^*)$:

$$v(y) - v(y^*) = \int_{y^*}^{y} u(t)dt = \int_{x^*}^{x} U(t)e^t dt \tag{A.173}$$

This function of y will be necessary to recover G_i from $\lambda(y) = \gamma_i G_i(v(y) - v(y^*))$, where $i \in \{u, d\}$ indexes the direction of adjustment for which there is an opportunity.

D.1 Algorithm

The algorithm to recover $U(\cdot)$ given $L(\cdot)$ is a two-level iterative procedure. The inner part is, given $L(\cdot)$ and a guess for x^*, to recover $U(\cdot; x^*)$. The outer part is, given $U(\cdot; x^*)$, to update the guess for x^* based on the fact that $U(x^*; x^*) = 0$ at the true value.

Inner part. Given a guess $x^*_{(n)}$ for x^*, let $\hat{x}_j = x^*_{(n)} + x_j$. The conditions on $\eta_{1,j}$ and $\eta_{2,j}$ combine into a linear system of dimensionality $2\mathbb{u} + 2\mathbb{d}$. Let the vector $\boldsymbol{\eta}$ combine the unknowns in the following way: $\boldsymbol{\eta}_{2\mathbb{u}+2j+1} = \eta_{1,j}$ and $\boldsymbol{\eta}_{2\mathbb{u}+2j+2} = \eta_{2,j}$ for all j such that $-\mathbb{u} \le j \le -1$. For $1 \le j \le \mathbb{d}$, set $\boldsymbol{\eta}_{2\mathbb{u}+2j-1} = \eta_{1,j}$ and $\boldsymbol{\eta}_{2\mathbb{u}+2j} = \eta_{2,j}$.

Let A be a square matrix of size $(2\mathbb{u} + 2\mathbb{d}) \times (2\mathbb{u} + 2\mathbb{d})$ and b be a column vector of length $2\mathbb{u} + 2\mathbb{d}$. The $(2j + 2\mathbb{u} + 1)$-th rows of $\mathbb{A}$ and $\mathbf{b}$ represent equation (A.171) for $-\mathbb{u} \le j < -1$:

$$\begin{aligned} &\eta_{1,j}e^{\xi_{1,j}\hat{x}_{j+1}} + \eta_{2,j}e^{\xi_{2,j}\hat{x}_{j+1}} - \eta_{1,j+1}e^{\xi_{1,j+1}\hat{x}_{j+1}} - \eta_{2,j+1}e^{\xi_{2,j+1}\hat{x}_{j+1}} \\ &\qquad = (\theta_{1,j+1} - \theta_{1,j})e^{(\alpha-1)\hat{x}_{j+1}} + \theta_{2,j+1} - \theta_{2,j} \end{aligned} \tag{A.174}$$

The $(2j + 2\mathbb{u} + 2)$-th rows represent equation (A.172) for $-\mathbb{u} \le j < -1$:

$$\begin{aligned} &\eta_{1,j}\xi_{1,j}e^{\xi_{1,j}\hat{x}_{j+1}} + \eta_{2,j}\xi_{2,j}e^{\xi_{2,j}\hat{x}_{j+1}} - \eta_{1,j+1}\xi_{1,j+1}e^{\xi_{1,j+1}\hat{x}_{j+1}} - \eta_{2,j+1}\xi_{2,j+1}e^{\xi_{2,j+1}\hat{x}_{j+1}} \\ &\qquad = (\theta_{1,j+1} - \theta_{1,j})(1-\alpha)e^{(\alpha-1)\hat{x}_{j+1}} \end{aligned} \tag{A.175}$$

Then, the rows $2\mathbb{u} - 1$ takes care of continuity at $\hat{x}_0 = x^*_{(n)}$:

$$\begin{aligned} &\eta_{1,-1}e^{\xi_{1,-1}\hat{x}_0} + \eta_{2,-1}e^{\xi_{2,-1}\hat{x}_0} - \eta_{1,1}e^{\xi_{1,1}\hat{x}_0} - \eta_{2,1}e^{\xi_{2,1}\hat{x}_0} \\ &\qquad = (\theta_{1,1} - \theta_{1,-1})e^{(\alpha-1)\hat{x}_0} + \theta_{2,1} - \theta_{2,-1} \end{aligned} \tag{A.176}$$

The row $2\mathbb{u}$ takes care of differentiability at $\hat{x}_0 = x^*$:

$$\begin{aligned}\eta_{1,-1}\xi_{1,-1}e^{\xi_{1,-1}\hat{x}_0} + \eta_{2,-1}\xi_{2,-1}e^{\xi_{2,-1}\hat{x}_0} - \eta_{1,1}\xi_{1,1}e^{\xi_{1,1}\hat{x}_0} - \eta_{2,1}\xi_{2,1}e^{\xi_{2,1}\hat{x}_0}& \\ = (\theta_{1,1} - \theta_{1,-1})(1-\alpha)e^{(\alpha-1)\hat{x}_0}&\end{aligned} \tag{A.177}$$

Then, the $(2j + 2\mathbb{u} - 3)$-th rows of $\mathbb{A}$ and $\mathbf{b}$ represent equation (A.171) for $1 < j \leq \mathbb{d}$:

$$\begin{aligned}\eta_{1,j}e^{\xi_{1,j}\hat{x}_{j-1}} + \eta_{2,j}e^{\xi_{2,j}\hat{x}_{j-1}} - \eta_{1,j-1}e^{\xi_{1,j-1}\hat{x}_{j-1}} - \eta_{2,j-1}e^{\xi_{2,j-1}\hat{x}_{j-1}}& \\ = (\theta_{1,j-1} - \theta_{1,j})e^{(\alpha-1)\hat{x}_{j-1}} + \theta_{2,j-1} - \theta_{2,j}&\end{aligned} \tag{A.178}$$

The $(2j + 2\mathbb{u} - 2)$-th rows represent equation (A.172) for $1 < j \leq \mathbb{d}$:

$$\begin{aligned}\eta_{1,j}\xi_{1,j}e^{\xi_{1,j}\hat{x}_{j-1}} + \eta_{2,j}\xi_{2,j}e^{\xi_{2,j}\hat{x}_{j-1}} - \eta_{1,j-1}\xi_{1,j-1}e^{\xi_{1,j-1}\hat{x}_{j-1}} - \eta_{2,j-1}\xi_{2,j-1}e^{\xi_{2,j-1}\hat{x}_{j-1}}& \\ = (\theta_{1,j-1} - \theta_{1,j})(1-\alpha)e^{(\alpha-1)\hat{x}_{j-1}}&\end{aligned} \tag{A.179}$$

For the matrix $\mathbb{A}$ this means that, for $-\mathbb{u} \leq j \leq -1$,

$$\mathbb{A}_{2j+2\mathbb{u}+1,2j+2\mathbb{u}+1} = e^{\xi_{1,j}\hat{x}_{j+1}} \tag{A.180}$$
$$\mathbb{A}_{2j+2\mathbb{u}+1,2j+2\mathbb{u}+2} = e^{\xi_{2,j}\hat{x}_{j+1}} \tag{A.181}$$
$$\mathbb{A}_{2j+2\mathbb{u}+1,2j+2\mathbb{u}+3} = -e^{\xi_{1,j+1}\hat{x}_{j+1}} \tag{A.182}$$
$$\mathbb{A}_{2j+2\mathbb{u}+1,2j+2\mathbb{u}+4} = -e^{\xi_{2,j+1}\hat{x}_{j+1}} \tag{A.183}$$
$$\mathbb{A}_{2j+2\mathbb{u}+2,2j+2\mathbb{u}+1} = \xi_{1,j}e^{\xi_{1,j}\hat{x}_{j+1}} \tag{A.184}$$
$$\mathbb{A}_{2j+2\mathbb{u}+2,2j+2\mathbb{u}+2} = \xi_{2,j}e^{\xi_{2,j}\hat{x}_{j+1}} \tag{A.185}$$
$$\mathbb{A}_{2j+2\mathbb{u}+2,2j+2\mathbb{u}+3} = -\xi_{1,j+1}e^{\xi_{1,j+1}\hat{x}_{j+1}} \tag{A.186}$$
$$\mathbb{A}_{2j+2\mathbb{u}+2,2j+2\mathbb{u}+4} = -\xi_{2,j+1}e^{\xi_{2,j+1}\hat{x}_{j+1}} \tag{A.187}$$

The vector $\mathbf{b}$ for $-\mathbb{u} \leq j \leq -1$ is filled as follows:

$$\mathbf{b}_{2j+2\mathbb{u}+1} = (\theta_{1,j+1} - \theta_{1,j})e^{(\alpha-1)\hat{x}_{j+1}} + \theta_{2,j+1} - \theta_{2,j} \tag{A.188}$$
$$\mathbf{b}_{2j+2\mathbb{u}+2} = (\theta_{1,j+1} - \theta_{1,j})(1-\alpha)e^{(\alpha-1)\hat{x}_{j+1}} \tag{A.189}$$

The rows $2(\mathbb{d} + \mathbb{u}) - 1$ and $2(\mathbb{d} + \mathbb{u})$ of the matrix $\mathbb{A}$ are

$$\mathbb{A}_{2\mathbb{u}-1,2\mathbb{u}-1} = e^{\xi_{1,-1}\hat{x}_0} \tag{A.190}$$
$$\mathbb{A}_{2\mathbb{u}-1,2\mathbb{u}} = e^{\xi_{2,-1}\hat{x}_0} \tag{A.191}$$
$$\mathbb{A}_{2\mathbb{u}-1,2\mathbb{u}+1} = -e^{\xi_{1,1}\hat{x}_0} \tag{A.192}$$
$$\mathbb{A}_{2\mathbb{u}-1,2\mathbb{u}+2} = -e^{\xi_{2,1}\hat{x}_0} \tag{A.193}$$
$$\mathbb{A}_{2\mathbb{u},2\mathbb{u}-1} = \xi_{1,-1}e^{\xi_{1,-1}\hat{x}_0} \tag{A.194}$$
$$\mathbb{A}_{2\mathbb{u},2\mathbb{u}} = \xi_{2,-1}e^{\xi_{2,-1}\hat{x}_0} \tag{A.195}$$
$$\mathbb{A}_{2\mathbb{u},2\mathbb{u}+1} = -\xi_{1,1}e^{\xi_{1,1}\hat{x}_0} \tag{A.196}$$
$$\mathbb{A}_{2\mathbb{u},2\mathbb{u}+2} = -\xi_{2,1}e^{\xi_{2,1}\hat{x}_0} \tag{A.197}$$

The vector $\mathbf{b}$ at these positions is filled as follows:

$$\mathbf{b}_{2\mathbb{u}-1} = (\theta_{1,1} - \theta_{1,-1})e^{(\alpha-1)\hat{x}_0} + \theta_{2,1} - \theta_{2,-1} \tag{A.198}$$

$$\mathbf{b}_{2\mathbb{u}} = (\theta_{1,1} - \theta_{1,-1})(1-\alpha)e^{(\alpha-1)\hat{x}_0} \tag{A.199}$$

For $1 < j \leq \mathbb{d}$, rows of the matrix $\mathbb{A}$ are

$$\mathbb{A}_{2j+2\mathbb{u}-3,2j+2\mathbb{u}-3} = e^{\xi_{1,j-1}\hat{x}_{j-1}} \tag{A.200}$$

$$\mathbb{A}_{2j+2\mathbb{u}-3,2j+2\mathbb{u}-2} = e^{\xi_{2,j-1}\hat{x}_{j-1}} \tag{A.201}$$

$$\mathbb{A}_{2j+2\mathbb{u}-3,2j+2\mathbb{u}-1} = -e^{\xi_{1,j}\hat{x}_{j-1}} \tag{A.202}$$

$$\mathbb{A}_{2j+2\mathbb{u}-3,2j+2\mathbb{u}} = -e^{\xi_{2,j}\hat{x}_{j-1}} \tag{A.203}$$

$$\mathbb{A}_{2j+2\mathbb{u}-2,2j+2\mathbb{u}-3} = \xi_{1,j-1}e^{\xi_{1,j-1}\hat{x}_{j-1}} \tag{A.204}$$

$$\mathbb{A}_{2j+2\mathbb{u}-2,2j+2\mathbb{u}-2} = \xi_{2,j-1}e^{\xi_{2,j-1}\hat{x}_{j-1}} \tag{A.205}$$

$$\mathbb{A}_{2j+2\mathbb{u}-2,2j+2\mathbb{u}-1} = -\xi_{1,j}e^{\xi_{1,j}\hat{x}_{j-1}} \tag{A.206}$$

$$\mathbb{A}_{2j+2\mathbb{u}-2,2j+2\mathbb{u}} = -\xi_{2,j}e^{\xi_{2,j}\hat{x}_{j-1}} \tag{A.207}$$

The vector $\mathbf{b}$ for $1 < j \leq \mathbb{d}$ is

$$\mathbf{b}_{2j+2\mathbb{u}-3} = (\theta_{1,j-1} - \theta_{1,j})e^{(1-\alpha)\hat{x}_{j-1}} + \theta_{2,j-1} - \theta_{2,j} \tag{A.208}$$

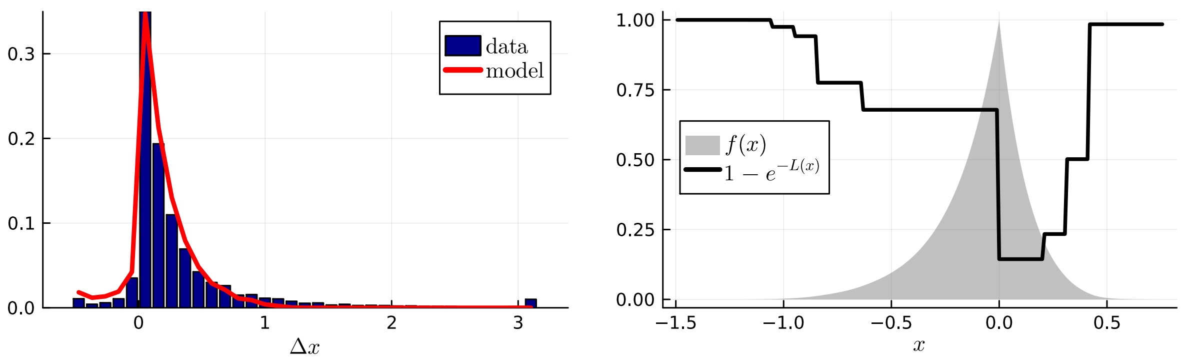
$$\mathbf{b}_{2j+2\mathbb{u}-2} = (\theta_{1,j-1} - \theta_{1,j})(1-\alpha)e^{(\alpha-1)\hat{x}_{j-1}} \tag{A.209}$$

This fills the first $2\mathbb{u} + 2\mathbb{d} - 2$ rows of A and $\mathbf{b}$. The remaining two take care of $\eta_{1,-\mathbb{u}} = 0$ and $\eta_{2,\mathbb{d}} = 0$: $\mathbb{A}_{2\mathbb{u}+2\mathbb{d}-1,1} = \mathbb{A}_{2\mathbb{u}+2\mathbb{d},2\mathbb{d}+2\mathbb{u}} = 1$ and $\mathbf{b}_{2\mathbb{u}+2\mathbb{d}} = \mathbf{b}_{2\mathbb{u}+2\mathbb{d}} = 0$.

Outer part. Given the coefficients $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, construct the function $U(\cdot; x^*_{(n)})$ and find $x^*_{(n+1)}$ such that $U(x^*_{(n+1)}; x^*_{(n)}) = 0$.

E Histograms and recovered hazards

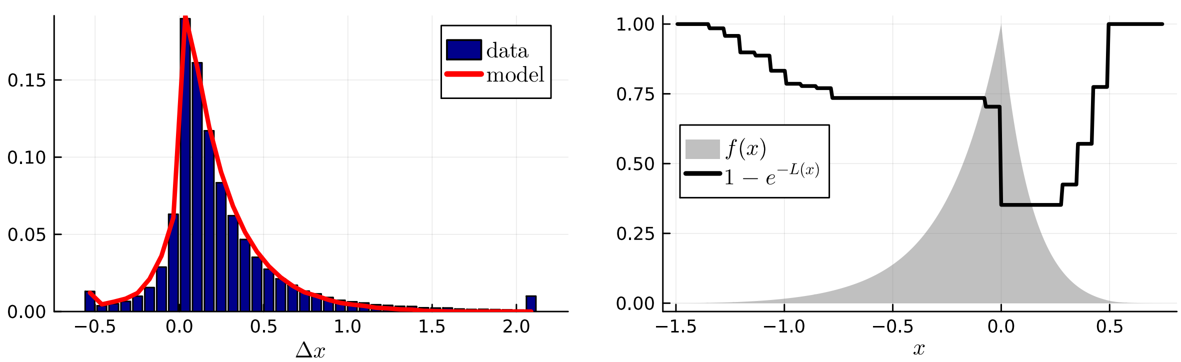
This section presents fitted histograms and recovered generalized hazard functions along with the underlying steady-state distributions of capital gaps for all sectors other than "Metal & Machinery". The graphs are organized in the same way as those for "Metal & Machinery" on Figure 2. The left panel shows data on investments and the histograms implied by the full model. The center and right panels show the same for two restricted benchmarks: the two-sided and symmetric models. The right panel shows the generalized hazard function and the implied steady-state distribution of capital gaps.



(a) Data and the full model.

(b) The two-sided benchmark.

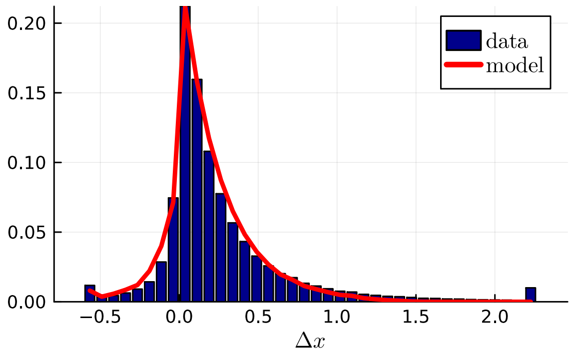
Figure A.1: Mining & Quarrying



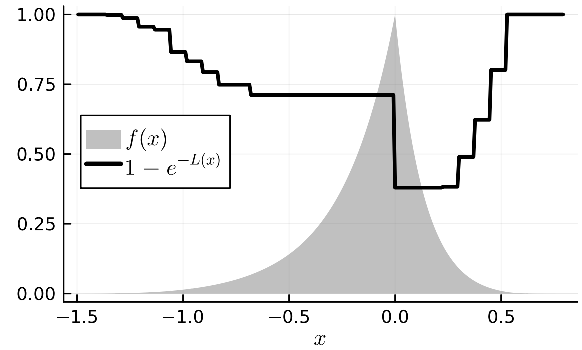
(a) Data and the full model.

(b) The two-sided benchmark.

Figure A.2: Chemicals

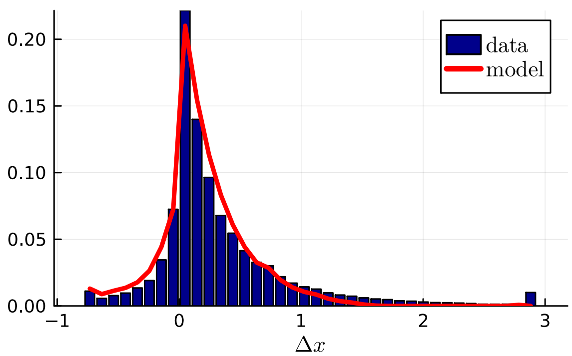


(a) Data and the full model.

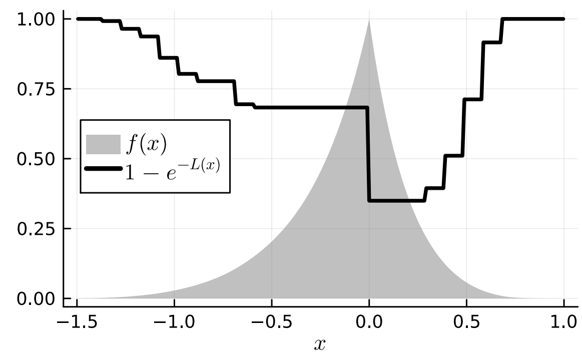


(b) The two-sided benchmark.

Figure A.3: Food & Beverages

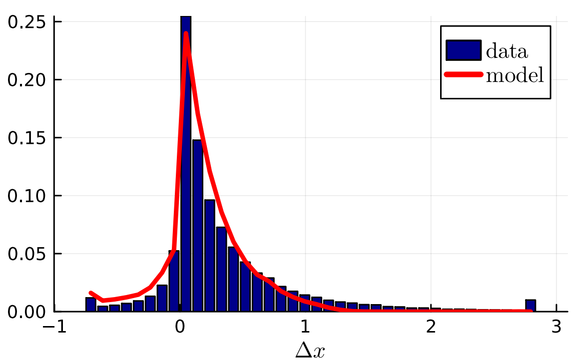


(a) Data and the full model.

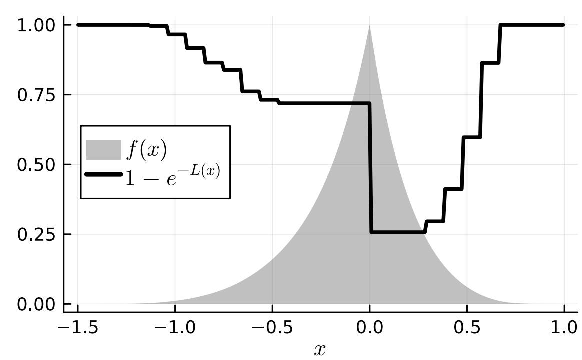


(b) The two-sided benchmark.

Figure A.4: Construction

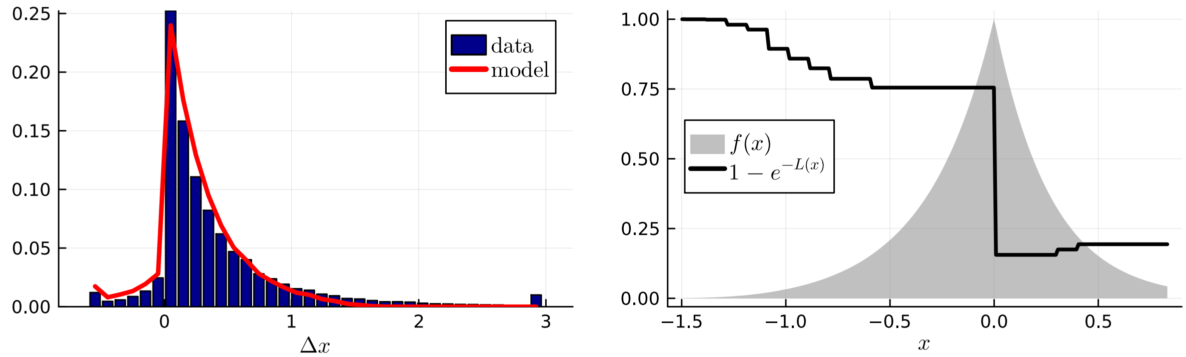


(a) Data and the full model.



(b) The two-sided benchmark.

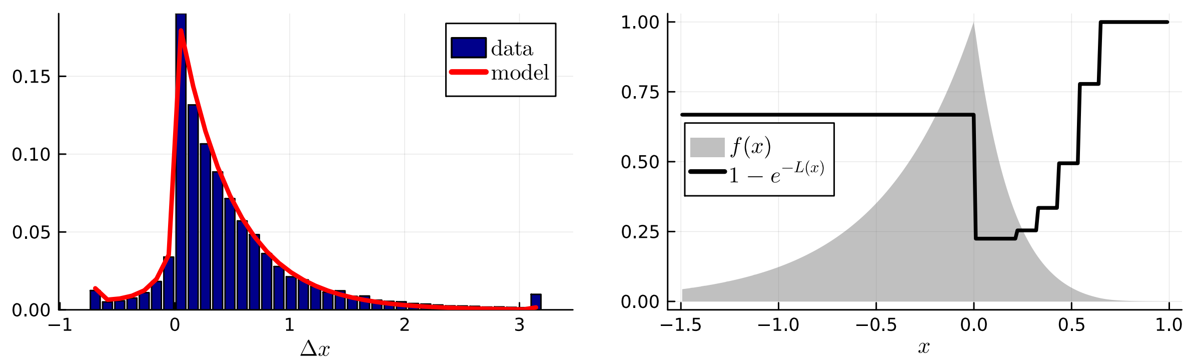
Figure A.5: Retail



(a) Data and the full model.

(b) The two-sided benchmark.

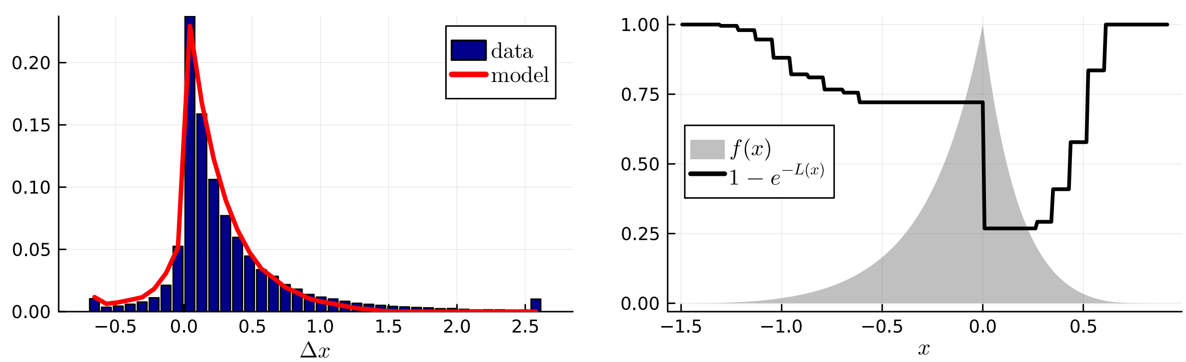
Figure A.6: Transportation



(a) Data and the full model.

(b) The two-sided benchmark.

Figure A.7: Insurance



(a) Data and the full model.

(b) The two-sided benchmark.

Figure A.8: Health & Beauty

F Recovered distributions of adjustment costs

This section presents our estimates for the underlying random menu costs. We plot the recovered distributions of adjustment costs for all sectors other than “Metal & Machinery”. The graphs are organized in the same way as those for “Metal & Machinery” on Figure 3. The left panel shows cumulative distribution functions G_u and G_d for costs of positive and negative adjustment. The center panel shows the arrival intensity of costs of positive adjustments. The right panel shows the arrival intensity of costs of negative adjustments. Costs are expressed in percent of instantaneous profits at optimal capital $e^{\alpha x^*} - (r+\delta)e^{x^*}$.

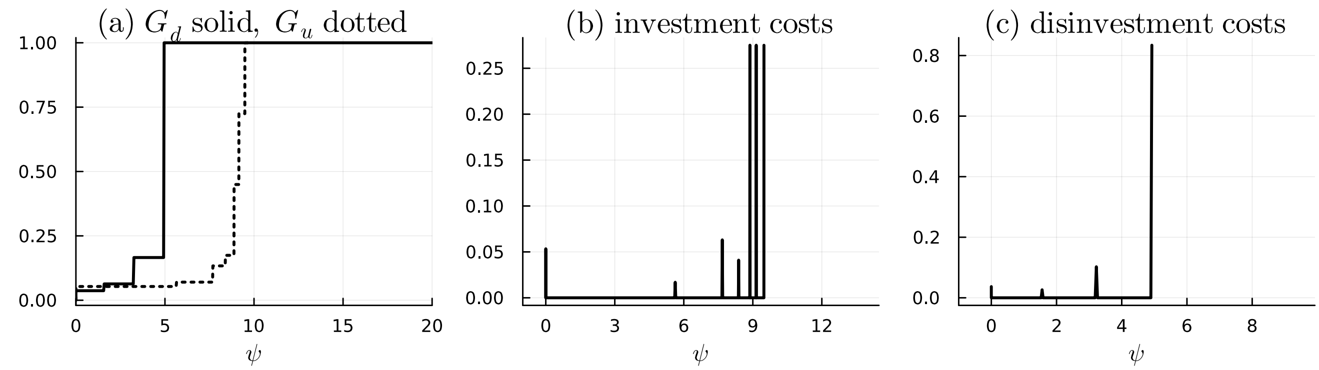


Figure A.9: Mining & Quarrying

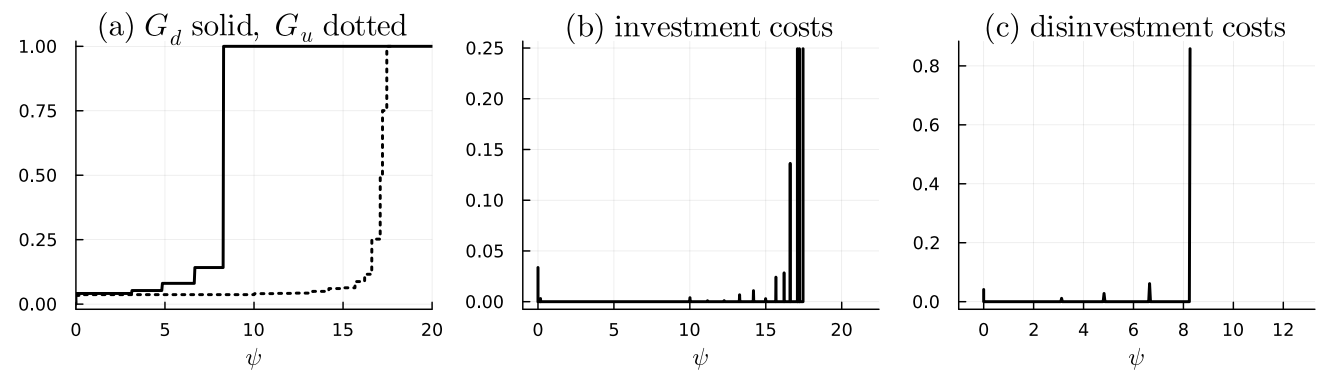


Figure A.10: Chemicals

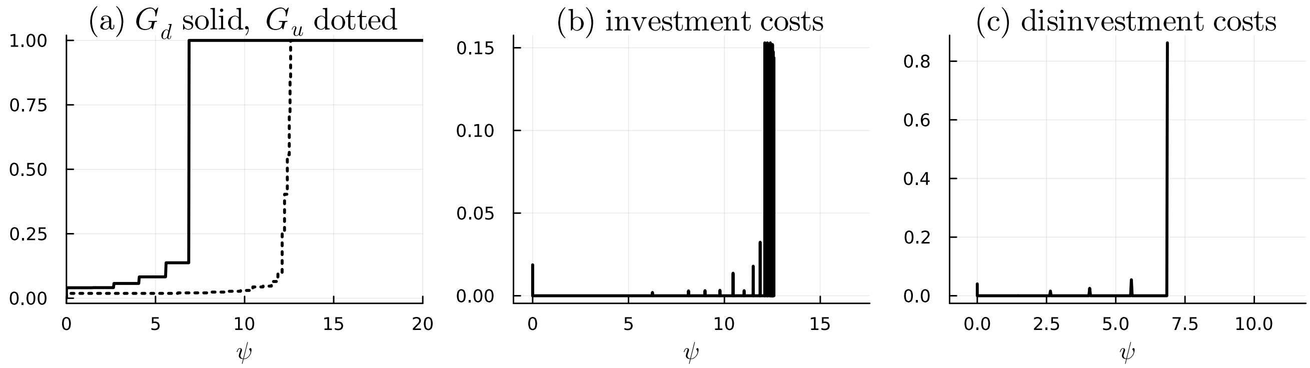


Figure A.11: Food & Beverages

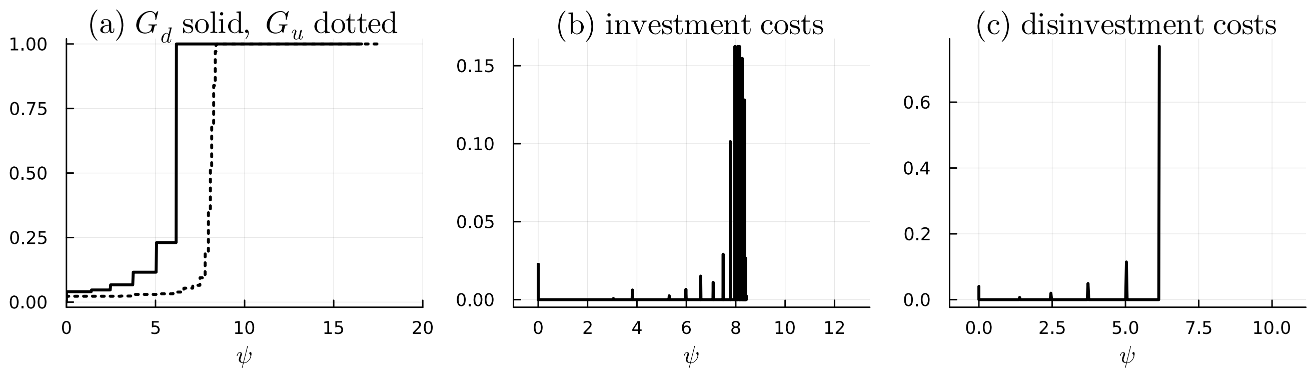


Figure A.12: Construction

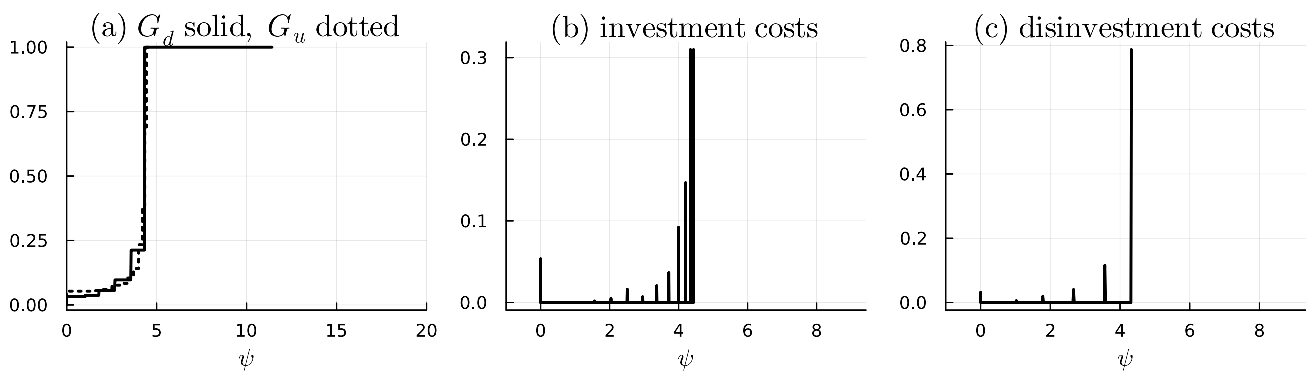


Figure A.13: Retail

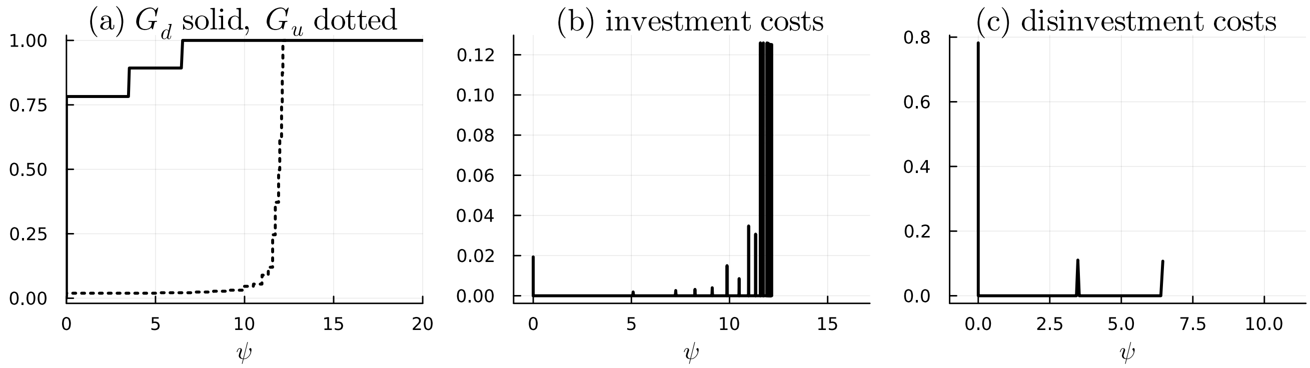


Figure A.14: Transportation

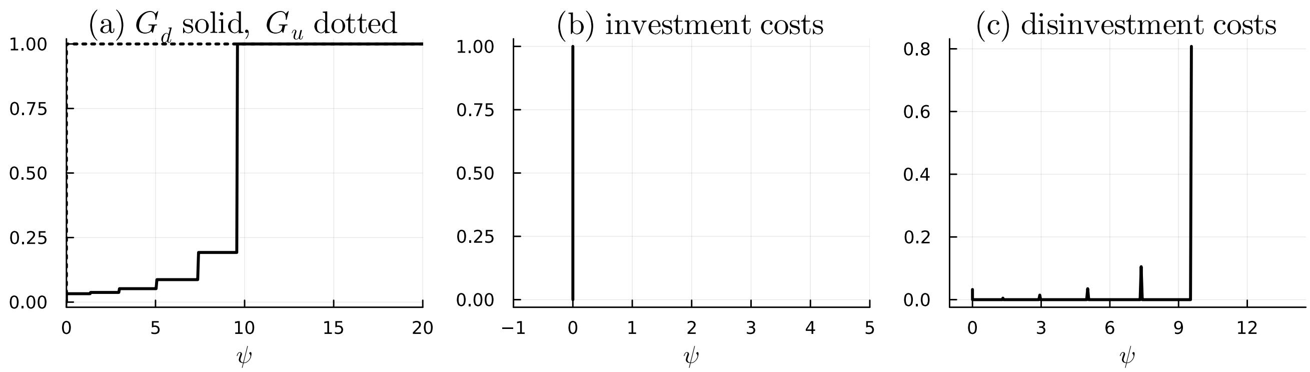


Figure A.15: Insurance

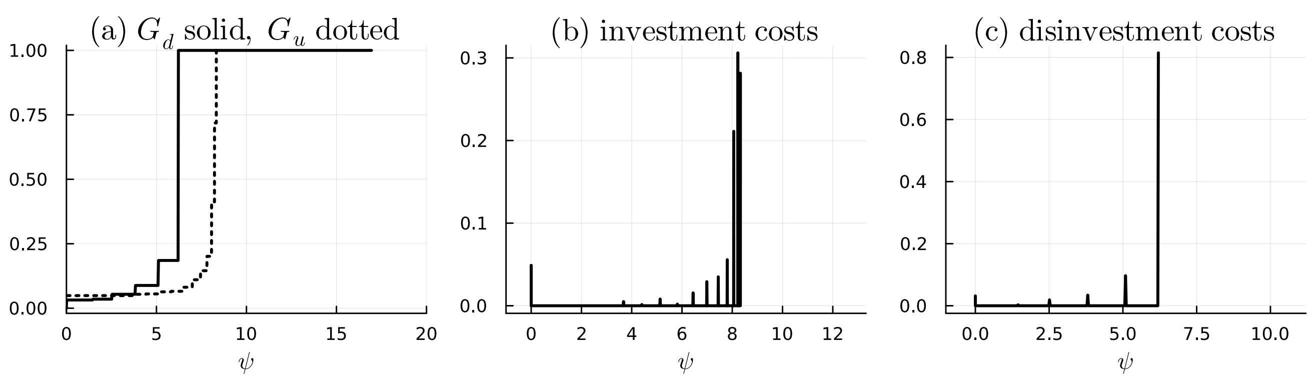


Figure A.16: Health & Beauty

G Estimated drift and volatility over time

This section presents our estimates of the drift of log capital stock between adjustments $\mu + \delta$ and the volatility of log productivity σ. We estimate these parameters in two subsamples: 1983-2003 and 2003-2023.

Table 8: Estimated parameters by subsample, 95% confidence intervals in brackets

	1983-2003		2003-2023	
Industry	$\mu + \delta$	σ	$\mu + \delta$	σ
Mining and Quarrying	0.257	0.365	0.112	0.216
	[0.225, 0.289]	[0.321, 0.409]	[0.096, 0.128]	[0.174, 0.259]
Chemicals	0.261	0.316	0.155	0.235
	[0.251, 0.271]	[0.306, 0.325]	[0.148, 0.162]	[0.229, 0.240]
Metals and Machinery	0.307	0.326	0.197	0.284
	[0.296, 0.318]	[0.302, 0.350]	[0.190, 0.205]	[0.280, 0.288]
Food and Beverages	0.256	0.260	0.147	0.240
	[0.248, 0.265]	[0.250, 0.270]	[0.143, 0.152]	[0.238, 0.243]
Construction	0.244	0.389	0.176	0.308
	[0.226, 0.261]	[0.366, 0.412]	[0.167, 0.185]	[0.303, 0.313]
Retail	0.208	0.370	0.155	0.306
	[0.199, 0.216]	[0.356, 0.384]	[0.150, 0.160]	[0.303, 0.309]
Transportation	0.296	0.475	0.207	0.356
	[0.265, 0.327]	[0.392, 0.559]	[0.191, 0.222]	[0.299, 0.413]
Insurance	0.459	0.544	0.348	0.389
	[0.409, 0.508]	[0.502, 0.585]	[0.329, 0.367]	[0.356, 0.423]
Health and Beauty	0.283	0.355	0.181	0.289
	[0.249, 0.317]	[0.311, 0.400]	[0.168, 0.194]	[0.282, 0.296]

H Two-sided model

Testing the two-sided model. The two-sided model is quantitatively far from the full one in terms of the estimated misallocation. This, however, is not in itself an indication that the two-sided model is prohibitively restrictive. We now conduct an econometric test to see if we can formally reject it against the full specification. Our approach is semi-parametric bootstrap. We adopt the estimated two-sided model as the null hypothesis and compare the goodness of fit under this null to the goodness of fit delivered by the full specification. To make this comparison, we need a measure of dispersion for goodness of fit under the null, which we obtain by drawing random subsamples of firms and recording the distance between the "true" data under the null and that produced by the subsample.

Formally, we do the following steps:

- estimate the flexible model $\hat{\mathcal{P}}_1$ on the full sample $\boldsymbol{Q}_{\text{full}}$, record $\mathcal{D}_{1,\text{full}} = \mathbf{dist}(\boldsymbol{H}(\hat{\mathcal{P}}_1), \boldsymbol{Q}_{\text{full}})$
- estimate the two-sided model $\hat{\mathcal{P}}_2$ on the full sample $\boldsymbol{Q}_{\text{full}}$
- draw B random subsamples, collect $\{\mathcal{D}_b\}_{b=1}^{b=B}$ measures of fit $\mathcal{D}_b = \mathbf{dist}(\boldsymbol{H}(\hat{\mathcal{P}}_2), \boldsymbol{Q}_b)$
- compare $\mathcal{D}_{1,\text{full}}$ to the distribution of $\{\mathcal{D}_b\}_{b=1}^{b=B}$
- define the p-value as $p = \hat{\mathbb{P}}\{\mathcal{D}_b < \mathcal{D}_{1,\text{full}}\}$

This procedure provides a measure of how extreme is the goodness of fit under the flexible specification compared to the norm under the null. We choose the rejection criterion for the p-value to be 0.05, although it could in principle be any other number.

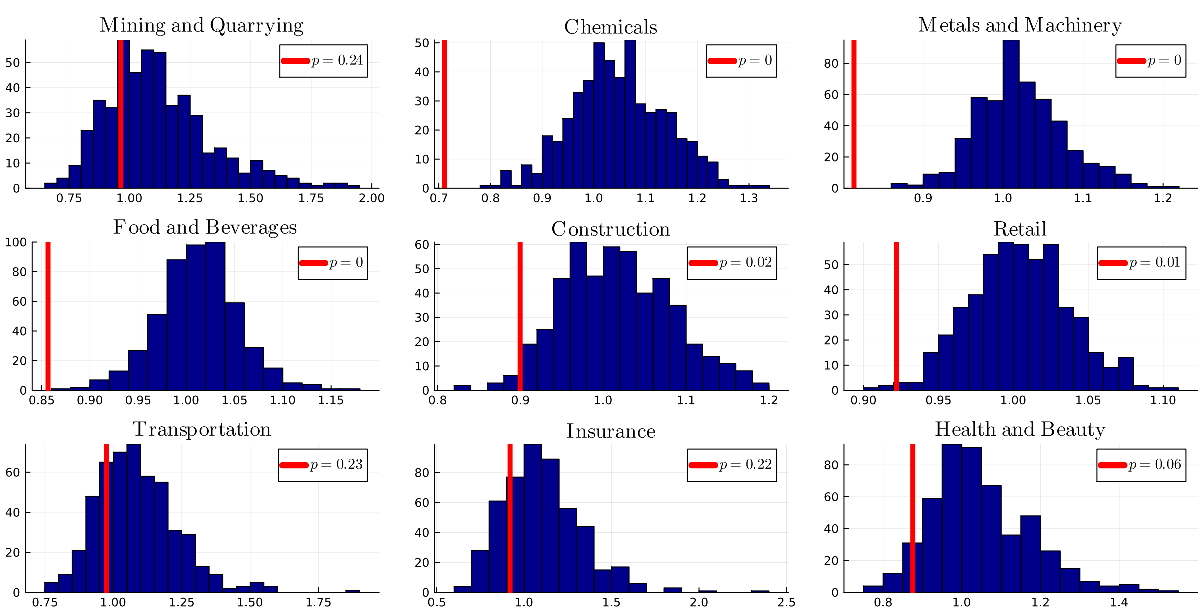


Figure A.17: distribution of $\mathcal{D}_b = \mathbf{dist}(\boldsymbol{H}(\hat{\mathcal{P}}_2), \boldsymbol{Q}_b)$ in the background, $\mathcal{D}_{1,\text{full}} = \mathbf{dist}(\boldsymbol{H}(\hat{\mathcal{P}}_1), \boldsymbol{Q}_{\text{full}})$ as the vertical line. All values normalized by $\mathcal{D}_{2,\text{full}} = \mathbf{dist}(\boldsymbol{H}(\hat{\mathcal{P}}_2), \boldsymbol{Q}_{\text{full}})$.

With the criterion $p > 0.05$, in five out of nine sectors we can reject the two-sided model. One of the sectors, "Healt & Beauty", is on the border with $p = 0.06$, while "Mining & Quarrying", "Insurance", and "Transportation" are rejected strongly. Table 9 collects p-values for all sectors.

Table 9: $p-$values for all sectors.

Industry	p
Mining and Quarrying	0.24
Chemicals	0.00
Metals and Machinery	0.00
Food and Beverages	0.00
Construction	0.02
Retail	0.01
Transportation	0.23
Insurance	0.22
Health and Beauty	0.06

I Data Appendix

In this appendix, we present additional correlation tables for firm-level activity measures and observable characteristics: revenue, employment, and time in the sample.

Table 10: correlations between firm-level observables and activity (Mining & Quarrying).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.14	-0.99	0.05	0.04	0.17
$down_i$	0.14		-0.18	0.12	0.09	0.1
$inactive_i$	-0.99	-0.18		-0.05	-0.04	-0.17
rev_i	0.05	0.12	-0.05		0.38	0.1
emp_i	0.04	0.09	-0.04	0.38		0.05
$time_i$	0.17	0.1	-0.17	0.1	0.05	
mean	0.71	0.04	0.28	150903	128.24	8.87
std	0.34	0.13	0.34	1074070	2116.04	6.72

Table 11: correlations between firm-level observables and activity (Chemicals).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.1	-0.94	0.08	0.06	0.17
$down_i$	0.1		-0.22	0.18	0.15	0.09
$inactive_i$	-0.94	-0.22		-0.08	-0.05	-0.16
rev_i	0.08	0.18	-0.08		0.67	0.09
emp_i	0.06	0.15	-0.05	0.67		0.03
$time_i$	0.17	0.09	-0.16	0.09	0.03	
mean	0.86	0.14	0.13	26006	73.66	10.43
std	0.23	0.21	0.22	115416	471.57	8.5

Table 12: correlations between firm-level observables and activity (Food & Beverages).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.13	-0.95	0.11	0.1	0.2
$down_i$	0.13		-0.25	0.26	0.26	0.16
$inactive_i$	-0.95	-0.25		-0.11	-0.1	-0.2
rev_i	0.11	0.26	-0.11		0.67	0.13
emp_i	0.1	0.26	-0.1	0.67		0.11
$time_i$	0.2	0.16	-0.2	0.13	0.11	
mean	0.82	0.13	0.16	13110	37.51	10.09
std	0.25	0.21	0.24	48912	163.05	8.39

Table 13: correlations between firm-level observables and activity (Construction).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.11	-0.93	0.08	0.1	0.21
$down_i$	0.11		-0.3	0.2	0.23	0.09
$inactive_i$	-0.93	-0.3		-0.07	-0.1	-0.21
rev_i	0.08	0.2	-0.07		0.66	0.06
emp_i	0.1	0.23	-0.1	0.66		0.08
$time_i$	0.21	0.09	-0.21	0.06	0.08	
mean	0.72	0.15	0.24	10250	27.73	7.76
std	0.32	0.24	0.31	39798	146.33	6.45

Table 14: correlations between firm-level observables and activity (Retail).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.11	-0.95	0.02	0.04	0.14
$down_i$	0.11		-0.24	0.09	0.09	0.09
$inactive_i$	-0.95	-0.24		-0.02	-0.04	-0.14
rev_i	0.02	0.09	-0.02		0.12	0.03
emp_i	0.04	0.09	-0.04	0.12		0.03
$time_i$	0.14	0.09	-0.14	0.03	0.03	
mean	0.79	0.1	0.19	22154	33.69	9.24
std	0.27	0.19	0.26	234877	458.7	7.38

Table 15: correlations between firm-level observables and activity (Transportation).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.07	-0.98	0.04	0.03	0.19
$down_i$	0.07		-0.15	0.12	0.1	0.11
$inactive_i$	-0.98	-0.15		-0.04	-0.03	-0.19
rev_i	0.04	0.12	-0.04		0.71	0.04
emp_i	0.03	0.1	-0.03	0.71		0.05
$time_i$	0.19	0.11	-0.19	0.04	0.05	
mean	0.77	0.05	0.22	32637	133.87	9.7
std	0.28	0.14	0.27	342662	2146.53	7.53

Table 16: correlations between firm-level observables and activity (Insurance).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.05	-0.96	0.06	0.11	0.26
$down_i$	0.05		-0.2	0.13	0.19	0.03
$inactive_i$	-0.96	-0.2		-0.06	-0.11	-0.25
rev_i	0.06	0.13	-0.06		0.31	0.06
emp_i	0.11	0.19	-0.11	0.31		0.13
$time_i$	0.26	0.03	-0.25	0.06	0.13	
mean	0.75	0.08	0.23	17987	69.14	8.03
std	0.32	0.19	0.31	114391	300.63	6.61

Table 17: correlations between firm-level observables and activity (Health & Beauty).

	up_i	$down_i$	$inactive_i$	rev_i	emp_i	$time_i$
up_i		0.09	-0.95	0.07	0.11	0.19
$down_i$	0.09		-0.22	0.17	0.18	0.05
$inactive_i$	-0.95	-0.22		-0.07	-0.1	-0.18
rev_i	0.07	0.17	-0.07		0.59	0.06
emp_i	0.11	0.18	-0.1	0.59		0.11
$time_i$	0.19	0.05	-0.18	0.06	0.11	
mean	0.82	0.11	0.16	15214	118.26	11.07
std	0.25	0.18	0.24	61418	397.89	7.84