A Structural Model of Asymmetric Lumpy Investment*

Francesco Lippi
Luiss University and EIEF
francescolippi@gmail.com

Aleksei Oskolkov
University of Chicago
Department of Economics
aoskolkov@uchicago.edu

November 13, 2023

Abstract

Firms' capital investments are lumpy and asymmetric, with disinvestments being less frequent and smaller than positive investments. We present a tractable model that reproduces the observed frequency and size distribution of investments by assuming that capital investment is subject to random fixed costs. We provide an inverse map to recover the model's primitives: in particular, we use the size distribution of investment to identify the underlying distribution of adjustment costs, as well as other fundamental parameters such as the drift and volatility of the firm's productivity. We apply the method to panel data on investments by Italian firms. We use the identified model to quantify the capital adjustment costs and to document their asymmetric nature: disinvestments are more costly than investments. We explore the extent to which these asymmetries imply a non-linear propagation of aggregate productivity shocks.

PRELIMINARY DRAFT

Key Words: Lumpy investment, Inaction, Generalized hazard function, fixed cost, impulse response

JEL Classification Numbers: D22, D25, E20, E22

*We benefited from discussions with Fernando Alvarez and Isaac Bailey. Artyom Lipin provided excellent research assistance. Lippi acknowledges financial support from the ERC grant: 101054421-DCS.

1 Introduction

Firms' investments in capital goods occur infrequently and in sizeable amounts, with positive investments being more common than disinvestments, see Doms and Dunne (1998). Common explanations for such "lumpy investment" behavior involve irreversibilities and fixed (convex) costs of adjustment at the firm level, as in e.g. Bertola and Caballero (1994). The lumpy nature of investment raises questions concerning the causes of the phenomenon and the implications of the micro-lumpiness for the propagation of aggregate shocks, see Khan and Thomas (2008); Winberry (2021); Baley and Blanco (2021).

A seminal contribution to this literature is the paper by Caballero and Engel (1999), which assumes that the firm's adjustment cost is drawn each period from a non-degenerate distribution. Unlike simple fixed-cost problems giving rise to an Ss rule, where all investments have the same size, the setup by Caballero and Engel (CE) gives rise to a smooth version of the Ss rule: at every moment some firms will invest if the adjustment cost drawn is sufficiently small. The setup by CE maps a set of primitives, involving the distribution of the fixed costs and other deep parameters, into an observable (non-degenerate) distribution of investment-sizes. Our paper adopts the CE setup and studies the inverse problem: starting from the observed distribution of the size of investments, we aim to recover the primitives that generate the data, in particular the distribution of the fixed costs associated to positive and negative capital adjustments.

Our goal is to provide a tractable method to identify a high dimensional object, namely the distribution of the adjustment costs, using data on the frequency and size of the firms' investments. We illustrate the approach exploiting a large dataset of capital investment by Italian firms, from several industries, over a 25 year period. We use the identified model to quantify the capital adjustment costs and to document its asymmetric nature: disinvestment is more costly than investment. This finding is not obvious: an asymmetric distribution of investments might simply reflect a large negative drift in the law of motion of capital. The results show that a large negative drift is indeed present in the data, but that accounting for the observed investment patterns also requires large asymmetries in the adjustment cost of capital. We also show that these asymmetries imply non-linearities for the propagation of an aggregate productivity shock, making the impulse response dependent on the size and sign of the shock.

The setup. We develop the analysis using Caballero and Engel (1999) setup, which assumes that firms are subject to idiosyncratic productivity shocks. Firms use capital to produce, and the optimal capital stock depends on the current productivity. Adjusting the capital stock is assumed to be costly: with probability κ per unit of time the firm draws a fixed cost ψ from a distribution $G(\psi)$ in each period and decides whether to pay the cost and adjust capital or continue to operate with the old capital stock. The setup generates lumpy investment behavior. In particular we allow both the rate κ_i and the distribution G_i to depend on the sign of the adjustment: $i = u$ for investments and $i = d$ for disinvestments. These assumptions can rationalize that negative investments are much less frequent than positive ones, and that their size distribution is markedly different in the data.

The setup is tractable in that the firm's state is described by a scalar variable: the ratio of capital to productivity, relative to the profit maximising ratio in the absence of frictions. We call this variable the capital gap, and denote it by x. The gap is unobservable since productivity is not recorded, and for this reason the primitives of the firm's problem cannot be simply read off the data. To recover such primitives, we analyze the optimal choices of firms in conjunction with investment data. The problem of the firm yields a generalized hazard function $\Lambda(x)$, specifying the adjustment intensity as a function of the underlying state x. The cross-sectional distribution of capital gaps, described by the density function $f(x)$, solves a differential equation that depends on Λ. The distribution f is not observable but is tightly related to the observed size-distribution of investments, $q(\Delta x)$. We illustrate how to recover both Λ and f using data from the distribution q.

This recovery procedure can in principle be done non-parametrically, using the probability density of investments as an input. However this approach requires solving a second-order differential equation complicated by the presence of drift in productivity and capital stock. Another second-order differential equation then has to be solved to recover the underlying distribution of adjustment costs. We suggest a less flexible procedure that simplifies the calculations tremendously. Specifically, we break the observed distribution of investments into bins and assume that the generalized hazard function Λ is constant within bins.[1] This gives us closed-form solutions for the differential equations and turns the problem of recovering Λ and f into solving a linear system. The fundamental distribution of fixed costs, $G(\psi)$, is also easy to recover: it reduces to a discrete distribution with a finite number of mass points. Identifying them requires to solve another linear system.

We illustrate the main results of the paper by applying the model to a panel of investment data for Italian firms obtained from the Company Accounts Data Service (*Centrale dei Bilanci*). The data cover a period of 25 years, from 1982 to 2006, and contain the information about industry and region in which the firm operates, its assets, investment in tangible and non-tangible assets as well as disinvestments.[2] We use the data from 9 broad sectors to compute the size distribution of investment q and to recover the generalized hazard function Λ and the underlying distributions of fixed costs G, for each sector.

Main results. The results reveal that a substantial degree of asymmetry in the primitives is needed to fit the data. The opportunity to downsize the capital stock arises less frequently, a result that confirms the findings by Baley and Blanco (2021). This finding is driven by the prevalence of positive investment in the data.[3] Negative investment is also more expensive

[1] In practice, since the data come in the form of histograms, the distribution of investment within each bin has to be ignored. Assuming that Λ is constant within bins is attractive from the standpoint of computational efficiency. The resulting discrete distributions of fixed costs are easy to use in quantitative evaluations of the model.

[2] This data have been used in Guiso and Schivardi (2007) and Guiso, Lai, and Nirei (2017).

[3] Table 3 in Baley and Blanco (2021) uses data on structures, which do not feature negative investments. For this reason these authors end up calibrating an infinite cost of downward adjustments.

on average, being associated to an average higher fixed cost and that arises less frequently ($\kappa_u > \kappa_d$). Another feature of the recovered primitives is that the generalized hazard function Λ increases as the firm gets further away from the optimal capital stock. The recovered magnitude of the adjustment costs is comparable to the ones discussed in the literature, e.g. by Caballero and Engel (1999).

We use the recovered primitives to study the impulse response of the economy to an aggregate productivity shock. Specifically, we hit all firms with a multiplicative shock to productivity. This changes everyone's capital gap and triggers a wave of investment. We then trace out the path of the average capital gap, which to the first order approximates the aggregate capital-to-productivity ratio. We find that the cumulative impulse response (CIR) of capital gaps is asymmetric around zero and non-linear in the size of the shock. Asymmetry is natural to expect given that we recover different fixed cost distributions for positive and negative investments as well as different arrival frequencies of investment opportunities. The non-linearity in the size of the shock is a direct consequence of a non-constant generalized hazard function.[4]

The model performance can be compared to two well-known benchmarks that are nested by discrete distributions. The simple Calvo model is nested as a special case of a discrete distribution of fixed cost with one mass point on zero. When an adjustment opportunity arises, the firm always draws a zero fixed cost. The generalized hazard function is constant and independent of x, meaning that firms are equally likely to adjust regardless of their desired adjustment size.

Another benchmark is what we call a two-sided distribution of fixed costs. In this case, adjustment opportunities for positive and negative investments may arise with different

[4]Suppose the generalized hazard function increases in the distance from the optimal gap. Consider an aggregate shock to productivity that shifts the whole distribution of capital gaps away from the optimal one. Then the average adjustment hazard faced by the bulk of the firm population is higher. It then takes the economy less time to reset most of the gaps to the optimal value and come back to the steady-state trajectory. See Cavallo, Lippi, and Miyahara (2023) for a similar idea applied to the context of sticky prices.

frequencies, but the fixed cost drawn by the firms is always zero. This specification allows for different propensity to invest depending on whether the firm has too much or too little capital. However, conditional on the sign of desired adjustment, the incidence is still independent of the desired adjustment size.

We estimate the Calvo model and the two-sided specification on the same data as the full model. A first advantage of our approach is that it provides a characterization of the costs of adjustment, whereas both Calvo and the two-sided specification are mute about the size of the adjustment costs.

A second set of results concerns the implications for the economy's response to an aggregate shock. The CIR of the Calvo model is linear in the size of the shock. We find that this model overestimates CIR relative to the baseline. This is intuitive since the generalized hazard functions in the two models are linked. Specifically, the harmonic averages of the two generalized hazard functions Λ have to be the same. In the full model, aggregate shocks move the distribution of firms into regions with a higher-than-average Λ, triggering faster adjustment. In the Calvo model, this is impossible, since Λ is constant.

The two-sided distribution overestimates CIR as well. Perhaps surprisingly, we find that it fares worse than the simple Calvo model in many cases, especially for negative shocks to productivity. The reason is that the two-sided model can capture asymmetry in Λ with respect to size, so it promptly assigns a low value of Λ to negative investments. When the economy gets a negative productivity shock, firms need negative investment, and in the two-sided model, such opportunities are estimated to arrive very rarely. The economy then takes longer to adjust than the Calvo economy.

For the same intuition, the two-sided model should generate faster adjustment for positive shocks than the Calvo model, at least right after the shock hits. We establish this as a theoretical result, showing that the slope of the impulse response function of capital gaps right after the shock is higher in the two-sided model than in Calvo for positive productivity shocks and lower for negative ones. In the direct aftermath of the shock, the two-sided distribution will always estimate a more sluggish response than Calvo for negative shocks and a less sluggish one for positive shocks. Of course, this does not generalize to the comparison of CIR between the two models, and we find that for large shocks the two-sided model outperforms the Calvo model.

Related literature. A few papers have studied the inverse inference problem in a fixed-cost model. Alvarez, Lippi, and Oskolkov (2022) study infrequent price adjustments through the lenses of a menu-cost model that has several elements in common with the problem studied here. A main difference with the present paper is that pricing behavior is quite symmetric, i.e. the shape of the adjustments with a positive size 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.[5] A methodological novelty of this paper is to solve for the inverse mapping in a problem where the distribution of adjustments is not symmetric. The solution we propose is tractable and can be applied to other problems with asymmetries, such as price setting with high inflation, portfolio management problems, as well as inventory problems.

Our analysis also relates to the recent paper by Baley and Blanco (2021) who study a lumpy investment model where the fixed cost is either zero, with some probability, or else equal to a constant (possibly different for positive and negative investments). 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 the 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

[5]Symmetry occurs if the state has a small drift (low inflation) and the firm's return function (profit) is locally symmetric.

main empirical application restricts the distribution of adjustment costs to have two values, our model allows for a general shape of this distribution.

The paper is organized as follows: the next section describes the economic setup for the investment problem that we consider and defines the optimal policy by the firms, given by the generalized hazard function. Section 3 presents the main theoretical result of the paper, an inverse mapping that allows us to recover the distribution of the fixed costs starting from a given distribution of investments. Section 4 applies this result using a large panel of data on investment by Italian firms. Section 5 uses the estimated investment hazards and attempts a preliminary analysis of the effects of different distributions for the propagation of aggregate shocks.

2 Setup

Firms use capital to produce a homogeneous good with a production function $F(K_t, z_t) = z_t^{1-\alpha} K_t^{\alpha}$. Productivity z_t evolves according to

$$d\ln(z_t) = \mu dt + \sigma dW_t \tag{1}$$

When uncontrolled, capital evolves according to

$$d\ln(K_t) = -\delta dt \tag{2}$$

Firms occasionally get an opportunity to invest. For that, they use the investment good, which is the same as what they produce. The price of this good is normalized to one.

When an adjustment opportunity arrives, firms can take it at a fixed cost of investment ψz_t. With a Poisson intensity κ_d, they 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 κ_u. Upward adjustment costs are distributed according to $G_u(\cdot)$ on interval $[0, \psi_u]$.

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

We guess and verify that only the ratio K_t/z_t matters for the adjustment decision. Specifically, the downward adjustment cutoff function $\boldsymbol{\psi}_d$ maps $[k^*, \infty)$ to $[0, \psi_d]$. The upward adjustment cutoff function $\boldsymbol{\psi}_u$ maps $(-\infty, k^*]$ to $[0, \psi_u]$. Appendix A provides details.

Let $i \in \{u, d\}$ be a binary function 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 will disinvest should the opportunity arrive. The value of the firm depends on the capital-to-productivity ratio $k \equiv K/z$. Denoting $\rho = r - \mu - \sigma^2/2$ and $\nu = r + \delta$, the Bellman equation in a stationary environment is

$$\begin{aligned}\rho v(k) = k^\alpha - \nu k + (\rho - \nu) k v'(k) + \frac{\sigma^2}{2} k^2 v''(k) \\ + \sum_{i=u,d} \mathbb{1}_i \kappa_i \int \max\{v(k^*) - v(k) - \psi, 0\} dG_i(\psi)\end{aligned} \tag{3}$$

Here $\mathbb{1}_i$ indicates that the firm is in the region where it wants to adjust up or down. The optimality condition for the choice of k^* is $v'(k^*) = 0$.

The cutoff functions $\boldsymbol{\psi}_i(\cdot)$ for $i \in \{u, d\}$ are given by $\boldsymbol{\psi}_i(k) = v(k^*) - v(k)$. The generalized hazard function, which is the intensity of adjustment given the state k, is

$$\lambda(k) = \sum_{i=u,d} \mathbb{1}_i \kappa_i G(\boldsymbol{\psi}_i(k)) \tag{4}$$

Equation (3) and equation (4) map the primitives $(\delta, \mu, \sigma^2, r, \alpha)$ and $(\kappa_i, G_i(\cdot))_{i=u,d}$ into generalized hazard $\lambda(\cdot)$. To obtain the inverse mapping, take the derivative of equation (3) with respect to k:

$$(\nu + \lambda(k))u(k) = \alpha k^{\alpha-1} - \nu + (\rho - \nu + \sigma^2)ku'(k) + \frac{\sigma^2}{2}k^2u''(k) \tag{5}$$

Here $u(k) = v'(k)$. Having recovered $U(\cdot)$, we can solve for $(\kappa_i, G_i(\cdot))_{i=u,d}$ using

$$\boldsymbol{\psi}_i(k) = v(k) - v(k^*) = \int_{k^*}^{k} U(t)dt \tag{6}$$

The random menu cost primitives satisfy $\kappa_d = \lim_{k\to\infty}\lambda(k)$, $\kappa_u = \lim_{k\to-\infty}\lambda(k)$, and $\kappa_i G_i(\psi) = \lambda((\boldsymbol{\psi}_i)^{-1}[\psi])$. Hence, having $\lambda(\cdot)$ and the set of parameters $(\delta, \mu, \sigma^2, r, \alpha)$ is enough to recover the primitivies of adjustment costs $(\kappa_i, G_i(\cdot))_{i=u,d}$.

It is convenient to work with $x \equiv \ln(k)$ and functions of this argument $\Lambda(x) = \lambda(k)$ and $U(x) = u(k)$. Denoting $x^* \equiv \ln(k^*)$, we have $U(x^*) = 0$. Equation (5) transforms into

$$(\nu + \Lambda(x))U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu + \delta)U'(x) + \frac{\sigma^2}{2}U''(x) \tag{7}$$

The equation for $G_i(\cdot)$ and κ_i is

$$\Lambda(x) = \sum_{i=u,d} \kappa_i G_i(v(e^x) - v(e^{x^*})) = \sum_{i=u,d} \kappa_i G_i\left(\int_{x^*}^{x} U(t)e^t dt\right) \tag{8}$$

In Appendix A, we provide details on the HJB equation (3) and include a more general version of the model, where firms can always pay a fixed cost to adjust, and the occasionally arriving opportunity decreases its value. In Appendix B, we show that this setup is equivalent to one where firms rent capital instead of owning it. In Appendix C, we show the solution for equation (7) under a specific assumption on the functional form of $\Lambda(\cdot)$ that makes it piece-wise constant. This is the case we use in our empirical application in Section 4.

3 Estimating the hazard function

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

$$f(x)\Lambda(x) = (\mu + \delta)f'(x) + \frac{\sigma^2}{2} f''(x) \tag{9}$$

Productivity z is not observable, so $f(\cdot)$ and $\Lambda(\cdot)$ cannot be simply read from the data. One needs to use the observed distribution and frequency of investments to recover them. Upon observing an adjustment, we can record its size in logarithms, $\Delta x = \ln(K_+) - \ln(K_-)$, where K_- and K_+ are the values of capital stock right before and right after the change. This equals the jump in capital gap: $\ln(K_+) - \ln(K_-) = (\ln(K_+) - \ln(z_+)) - (\ln(K_-) - \ln(z_-))$. Productivity $\ln(z_-)$ and $\ln(z_+)$ drops out, as z has continuous sample paths.

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

$$f(x)\Lambda(x) = Nq(x^* - x) \tag{10}$$

Here N is the frequency of adjustments. The interpretation of equation (10) is that the number of adjustments of the size $x^* - x$ per unit of time is equal to the number of firms who have a capital gap 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.

It is convenient to re-center f and Λ around x^*. Define $\tilde{f}(\cdot)$ and $\tilde{\Lambda}(\cdot)$ by $\tilde{f}(x) = f(x - x^*)$

and $\tilde{\Lambda}(x) = \Lambda(x - x^*)$. For these functions, we have

$$\tilde{f}(x)\tilde{\Lambda}(x) = (\mu + \delta)\tilde{f}'(x) + \frac{\sigma^2}{2}\tilde{f}''(x) \tag{11}$$

$$\tilde{f}(x)\tilde{\Lambda}(x) = Nq(-x) \tag{12}$$

In practice, estimation procedures use histograms, which pool observation within bins and provide limited information on the tails. We propose a method to recover $(\tilde{f}, \tilde{\Lambda})$ under a functional form assumption that maximizes computational convenience taking into account these data limitations. Specifically, we assume that $\tilde{\Lambda}$ is constant within each bin of the observed adjustment histogram. Since the information on the functional form of Q 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 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 U bins X_j with boundaries $\{x_j\}_{-U \leq j \leq 0}$, where $x_0 = 0$ and $x_{-U} = -\infty$. Accordingly, let the positive gaps corresponding to negative investment be pooled into D bins X_j with boundaries $\{x_j\}_{0 \leq j \leq D}$, where $x_0 = 0$ and $x_D = \infty$.

Assumption 1. The re-centered generalized hazard function $\tilde{\Lambda}$ is given by $\tilde{\Lambda}(x) = \lambda_j$ for $x \in X_j$, where $X_j = (x_j, x_{j+1}]$ for $-U \leq j \leq -1$ and $X_j = [x_{j-1}, x_j)$ for $1 \leq j \leq D$.

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

The data provide a frequency of investments N and a histogram $\boldsymbol{Q} = \{Q_j\}_{-D \leq j \leq U, j \neq 0}$ 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 $\boldsymbol{H} = \{H_j\}_{-D \leq j \leq U, j \neq 0}$ is potentially different from $\boldsymbol{Q}$.

We next characterize the mapping $\mathcal{P} \mapsto \mathcal{D}$.

3.1 Mapping parameters to the data

Under Assumption 1, the solution to equation (9) is a linear combination of two exponentials:

$$\tilde{f}_j(x) = \eta_{1,j} e^{\xi_{1,j} x} + \eta_{2,j} e^{\xi_{2,j} x} \tag{13}$$

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

$$\{\xi_{1,j}, \xi_{2,j}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2 \lambda_j}}{\sigma^2} \tag{14}$$

Denote the vectors of these coefficients by $\boldsymbol{\xi}_1 = \{\xi_{1,j}\}_{-U \leq j \leq D, j \neq 0}$ and $\boldsymbol{\xi}_2 = \{\xi_{2,j}\}_{-U \leq j \leq D, j \neq 0}$. The sets of coefficients $\eta_{1,j}$ and $\eta_{2,j}$ satisfy the following continuity conditions:

$$\tilde{f}_{j-1}(x_j) = \tilde{f}_j(x_j) \text{ for } j \in \{-U+1, ... -1\} \tag{15}$$

$$\tilde{f}_{j+1}(x_j) = \tilde{f}_j(x_j) \text{ for } j \in \{1, ...D-1\} \tag{16}$$

$$\tilde{f}_{-1}(0) = \tilde{f}_1(0) \tag{17}$$

$$\tilde{f}'_{j-1}(x_j) = \tilde{f}'_j(x_j) \text{ for } j \in \{-U+1, ... -1\} \tag{18}$$

$$\tilde{f}'_{j-1}(x_j) = \tilde{f}'_j(x_j) \text{ for } j \in \{1, ...D-1\} \tag{19}$$

The probability density must be continuous, including at zero. It must also be differentiable between all segments except for the junction at $x = 0$. In addition, $f(\cdot)$ must not explode at infinity, and it must integrate to one:

$$\lim_{x \to \infty} \tilde{f}_D(x) = \lim_{x \to -\infty} \tilde{f}_{-U}(x) = 0 \tag{20}$$

$$\int_{-\infty}^{\infty} \tilde{f}(x) dx = 1 \tag{21}$$

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

$$\lim_{x \to -0} \tilde{f}'_{-1}(x) - \lim_{x \to +0} \tilde{f}'_1(x) = \frac{2N}{\sigma^2} \tag{22}$$

This can be shown by integrating equation (9) over the real line and using the statistical fact that $q(-x)N = \tilde{\Lambda}(x)\tilde{f}(x)$.

The conditions in equation (15), equation (16), equation (17), equation (18), equation (19), equation (20), and equation (21) provide $2(U+D)$ linear equations to solve for $2(U+D)$ unknowns $\boldsymbol{\eta}_1 = \{\eta_{1,j}\}_{-J \le j \le D, j \ne 0}$ and $\boldsymbol{\eta}_2 = \{\eta_{2,j}\}_{-J \le j \le D, j \ne 0}$. The next proposition states this fact formally and uses it to establish the mapping from parameters to the data:

Proposition 1. Fix $\mathcal{P}$: a $(U+D)$-dimensional vector $\boldsymbol{\lambda}$ with non-negative entries and a pair $(\mu + \delta, \sigma)$. The density of gaps is given by equation (13), where the coefficients $\boldsymbol{\xi}_1(\mathcal{P})$ and $\boldsymbol{\xi}_2(\mathcal{P})$ are given by equation (14), and the coefficients $\boldsymbol{\eta}_1(\mathcal{P})$ and $\boldsymbol{\eta}_2(\mathcal{P})$ solve a $2(U+D)$-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 $-U \le j \le D$ with $j \ne 0$:

$$n(\mathcal{P}) := \frac{\sigma^2}{2}(\eta_{1,-1}\xi_{1,-1} + \eta_{2,-1}\xi_{2,-1} - \eta_{1,1}\xi_{1,1} - \eta_{2,1}\xi_{2,1}) \tag{23}$$

$$h_j(\mathcal{P}) := \frac{1}{N}\left[\frac{\lambda_j \eta_{1,j}}{\xi_{1,j}}(e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}) + \frac{\lambda_j \eta_{2,j}}{\xi_{2,j}}(e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j})\right] \tag{24}$$

We show how to construct the linear system for $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ in the proof. In practice, the coefficients in this linear system only depend on parameters through $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, which need to be computed first. As a corollary, we note a homogeneity property:

Corollary 1. 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 $\mathcal{P} = (\mu + \delta, \sigma^2, \boldsymbol{\lambda})$. 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 $\mu + \delta$, optimize over all other parameters, and then scale the estimates by the ratio of the observed frequency to that in equation (23).

3.2 Estimation

The generalized hazard function under Assumption 1 is fully encoded in a finite set of numbers. It can be estimated by minimizing the error in equation (24) under one restriction: the menu cost model implies that $\tilde{\Lambda}(x)$ 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 $\hat{\lambda}_{j-1} \leq \hat{\lambda}_j$ 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 the following minimization problem:[6]

$$\hat{\mathcal{P}} = \arg \min_{\mathcal{P}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{25}$$

$$\text{s.t. } \lambda_{j+1} \geq \lambda_j \text{ for } j > 0 \tag{26}$$

$$\lambda_{j-1} \geq \lambda_j \text{ for } j < 0 \tag{27}$$

Here the true data $\boldsymbol{H}(\mathcal{P})$ are given by equation (24). The estimates $\hat{\mathcal{P}}$ correspond to a frequency $\hat{N} = n(\hat{\mathcal{P}})$ given by equation (23). They have to be divided by $\hat{N}/N$ to be consistent with the frequency N from the data.

Two main properties of the model are the asymmetry of the generalized hazard function

[6]Depreciation rate δ cannot be identified separately from μ, so we assume it is zero throughout the exercise and then use external data sources to estimate it.

and the fact that it increases in distance from zero. To gauge the role of these properties, we additionally estimate simple parametric specifications that shut them down one at a time. First, we fit a symmetric generalized hazard function that potentially increases in distance from zero. We choose a simple power form to minimize the number of parameters. This amounts to solving a version of the problem in equation (25):

$$\hat{\mathcal{P}}_{\text{symmetric}} = \arg \min_{\{\mathcal{P},\pi_1,\pi_2\}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{28}$$

$$\text{s.t. } \lambda_j = \pi_1 |\theta x_j + (1-\theta)x_{j-1}|^{\pi_2} \text{ for } j > 0 \tag{29}$$

$$\lambda_j = \pi_1 |\theta x_j + (1-\theta)x_{j+1}|^{\pi_2} \text{ for } j < 0 \tag{30}$$

The parameter $\theta \in (0,1)$ determines at which point in (x_{j-1}, x_j) we evaluate λ_j for every j. Optimization is over four numbers $(\pi_1, \pi_2, \mu + \delta, \sigma^2)$. This specification ignores the asymmetries in adjustment costs, making disinvestment as easy as a positive investment. However, the invariant distribution of capital gaps is still asymmetric due to drift $\mu + \delta$. The generalized hazard function is increasing in distance from zero, which we ensure by imposing a restriction $\pi_2 \geq 0$.

Second, 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. The corresponding version of the problem in equation (25) is

$$\hat{\mathcal{P}}_{\text{two-sided}} = \arg \min_{\{\mathcal{P},\lambda_u,\lambda_d\}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{31}$$

$$\text{s.t. } \lambda_j = \lambda_d \text{ for } j > 0 \tag{32}$$

$$\lambda_j = \lambda_u \text{ for } j < 0 \tag{33}$$

Optimization here is over four numbers $(\lambda_u, \lambda_d, \mu + \delta, \sigma^2)$. We call this model a two-sided distribution because the distribution of random menu cost has three points in its support:

a zero cost that is available upon arrival of an investment opportunity and two different (perhaps, infinite) costs for positive and negative adjustment that can always be paid. This case is considered, for example, by Baley and Blanco (2021).

This model has a particularly simple closed-form solution for the density of capital gaps. It is a single exponential function on either side of $x = 0$, as opposed to a sum of two exponential with potentially different exponents and coefficients on each segment. This 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. This follows from continuity and differentiability at all boundaries between segments.

Finally, we fit the original Calvo model with one value for all λ_j:

$$\hat{\mathcal{P}}_{\text{single-hazard}} = \arg\min_{\{\mathcal{P},\lambda\}} \mathbf{dist}(\boldsymbol{H}(\mathcal{P}), \boldsymbol{Q}) \tag{34}$$

$$\text{s.t. } \lambda_j = \lambda \text{ for all } i, j \tag{35}$$

We call this model a single-hazard one. The single hazard rate λ can be directly read off the data since it is equal to the adjustment frequency. The remaining coefficients are then easily computed given $\mu + \delta$ and σ^2. This model is nested by the symmetric benchmark if $\pi_2 = 0$ and by the two-sided benchmark if $\lambda_u = \lambda_d$.

We next summarize these properties of the two-sided model and a regular Calvo model.

Proposition 2. Consider a two-sided distribution model and suppose it is parameterized by $\mathcal{P} = (\mu + \delta, \sigma^2, \lambda_u, \lambda_d)$, with 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

$$\eta_{1,i} = \eta_{2,j} = \frac{\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} - \mu - \delta\right)\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d} + \mu + \delta\right)}{\sigma^2\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} + \sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d}\right)} \tag{36}$$

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

$$n(\mathcal{P}) = \frac{\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} - \mu - \delta\right)\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d} + \mu + \delta\right)}{2\sigma^2} \tag{37}$$

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

$$\eta_{1,i} = \eta_{2,j} = \frac{\lambda}{\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda}} \text{ for } i < 0, j > 0 \tag{38}$$

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

These restricted models cannot produce a better fit than the full one, but it is interesting to see how far from each other they are quantitative. One metric we will use to compare them is the impulse response of the average investment gap to a one-time aggregate productivity shock, which we compute in Section G after estimating the model on firm-level data.

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 3. Consider two models that generate the same frequency of adjustments. Let $\boldsymbol{\lambda}_1 = \{\lambda_{j,1}\}$ and $\boldsymbol{\lambda}_2 = \{\lambda_{k,2}\}$ define 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

$$\sum_j \frac{H_{j,1}}{\lambda_{j,1}} = \sum_k \frac{H_{k,2}}{\lambda_{k,2}} \tag{39}$$

In words, the Harmonic average of the generalized hazard function, weighted by the histogram of adjustments, is the same across models. In particular, for a Calvo model with generalized hazard λ and a two-sided model with hazards (λ_u, λ_d) the proposition implies that

$$\frac{1}{\lambda} = \frac{\mathbb{P}\{\text{positive investment}\}}{\lambda_u} + \frac{1 - \mathbb{P}\{\text{positive investment}\}}{\lambda_d} \tag{40}$$

Here λ_u is the upward adjustment hazard in the two-sided model, and λ_d is the downward one. A direct implication is that, if $\lambda_u > \lambda$, then $\lambda_d < \lambda$, so the Calvo model hazard is always in between the hazards of the two-sided model. This has implications for the asymmetric reaction of the economy to aggregate shocks, as discussed in Appendix G.

4 Panel Data on Investment in Italy.

For our analysis, we utilize panel data on Italian firms obtained from the Company Accounts Data Service, *Centrale dei Bilanci* (CB). The data cover a period of 25 years, from 1982 to 2006, and provides information about the industry and region in which each firm operates, its assets, investment in tangible and non-tangible assets, as well as disinvestments. These data have been used by Guiso and Schivardi (2007) and Guiso et al. (2017).

On average, there are about 45,000 firms included in the data set in a particular year. We calculate the net investment, I, by subtracting the disinvestment from the investment in tangible assets. In the baseline results, we normalize the investment by the stock of the illiquid assets that is given by 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 investments have been made. Capital stock before investment is then $A - I$.

Our variable of interest is the log change in assets

$$\Delta x \equiv \ln\left(\frac{A}{A-I}\right) = \ln\left(\frac{1}{1 - I/A}\right) \tag{41}$$

where I is recorded net investment and $A - I$ is capital before adjustment, as described above. Table 1 reports summary information sector-by-sector for the sample, using a 9-sector classification.

It is apparent that the data is characterized by a significant presence of inaction. In a typical year about one-fourth of the firms is inactive. We follow Cooper and Haltiwanger (2006), as well as Baley and Blanco (2021), and consider as "zero investment" all investments with an absolute size smaller than 1% of the firm's capital.

The average size of the investment (relative to assets) is about 15%, and the majority of the adjustments have a positive sign, i.e., the fraction of disinvestments is small in all sectors, pointing to the presence of drift and the possibility of asymmetries in the adjustment costs. The table also shows that there is significant variation across industries, both in the typical size of investments and in the prevalence of inaction.

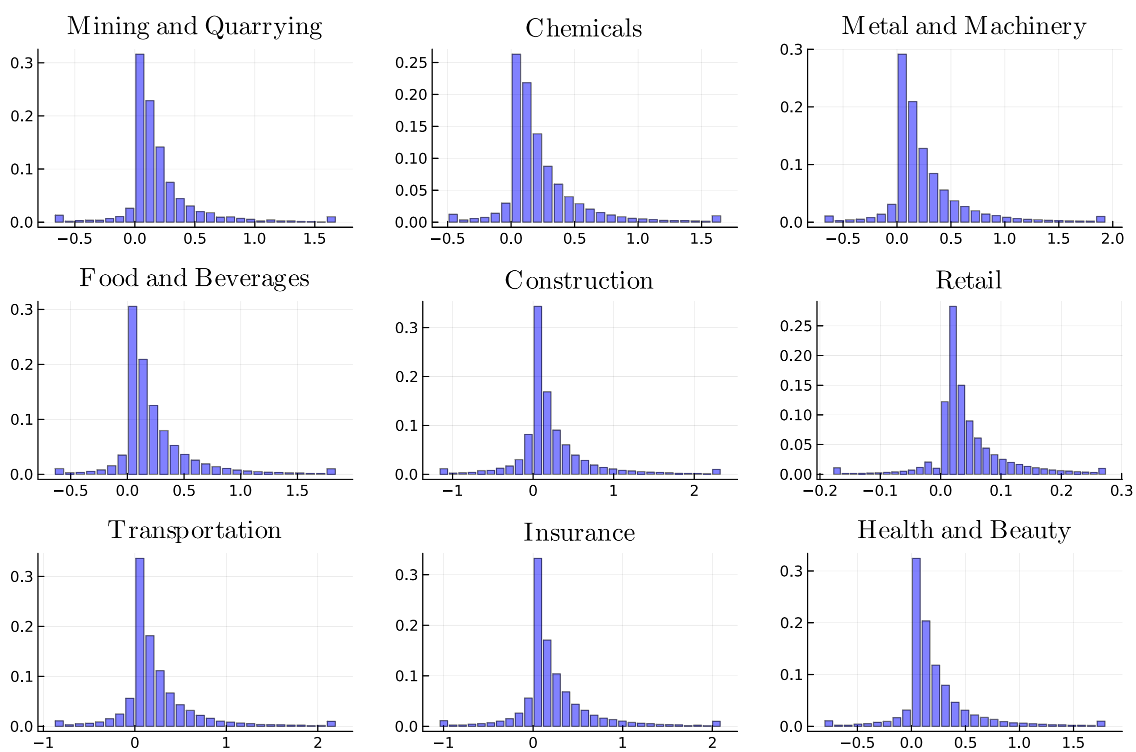
The histograms in Figure 1 describe the distribution of the size of the (non-zero) investments in each of the 9 industries considered. This distribution corresponds to the theoretical measure $Q(\cdot)$ described in equation (10). The leftmost and rightmost bins in each graph contain the entire tails of the observed distribution, which is why they are slightly larger than the neighboring bins.

Table 1: Summary statistics by industry

Industry	# Firms	Inactive	Mean Δx	SD Δx	share $\Delta x < 0$
Mining & Quarrying	814	0.239	0.133	0.347	0.053
Chemicals	5414	0.144	0.163	0.293	0.062
Metal & Machinery	13412	0.144	0.174	0.339	0.064
Food & Beverages	16346	0.156	0.161	0.328	0.067
Construction	7418	0.271	0.107	0.491	0.128
Retail	24589	0.200	0.143	0.407	0.090
Transportation	3529	0.212	0.145	0.396	0.102
Insurance	4659	0.383	0.124	0.464	0.087
Health & Beauty	2245	0.201	0.140	0.351	0.072
Total	78664	0.187	0.152	0.374	0.080

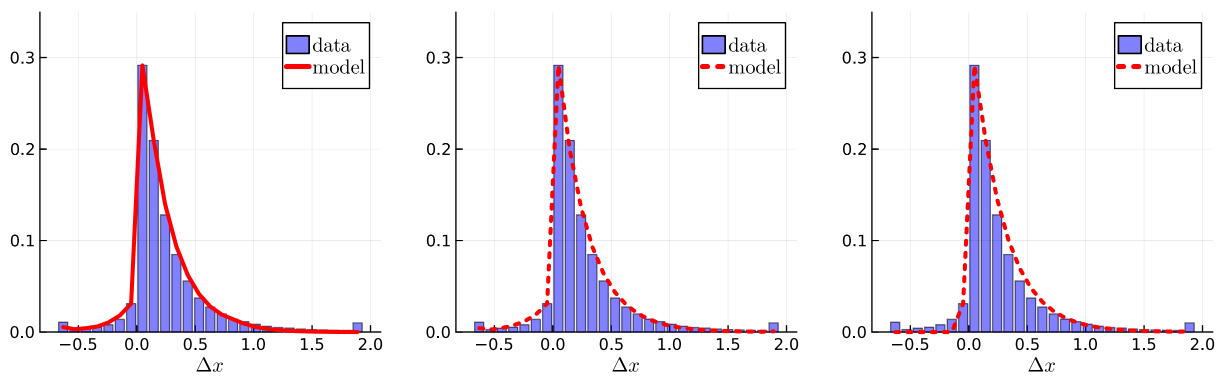
Notes: Investment is considered zero if the net investment to capital ratio is less than 1% in absolute value. *Inactive* is the share of observations with zero investment. The mean and standard deviation are computed for non-zero investments.

Figure 1: Distribution of non-zero investments.



4.1 Calibrating the model to the data

Steady-state objects. Having estimated model parameters, we can reconstruct the steady-state distribution of price gaps, the generalized hazard function, and the model-implied distribution of investments. Figure 2 shows the empirical histogram against the one generated by the model for one particular sector from our data, "Metal & Machinery". Panel (a) refers to the unrestricted specification. For comparison, panel (b) shows the same for a two-sided model, where Λ is restricted to two values: one for positive investments and one for disinvestments. Panel (c) shows the symmetric benchmark.



(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure 2: Data on investments and the histograms implied by the model. Full model on left panel, restricted models on center and right panels.

One observation to make about Figure 2 is that all three models are able to capture the form of the histogram quite closely. There is no qualitative difference between the steady-state distributions of investments in the three specifications. The differences will be more pronounced in the context of aggregate shocks, where the shape of the generalized hazard function is key for dynamics.

Figure 3 illustrates the estimated generalized hazard function $\tilde{\Lambda}$ and the recovered steady-state distribution $\tilde{f}$. Panel (a) shows the unrestricted benchmark and the two-sided model for comparison. Panel (b) does the same for the symmetric model. The corresponding figures for the rest of the sectors can be found in Appendix E.

The symmetric model is particularly far from the unrestricted benchmark in terms of the underlying distribution of capital gaps $\tilde{f}$. The reason is that imposing symmetry on $\tilde{\Lambda}$ forces the model to overestimate $\tilde{\Lambda}$ for positive gaps, that is, for negative adjustments. This leads to an underestimated $\tilde{f}$ for positive gaps since the model matches the observed density $q(-x)$ that is proportional to the product $\tilde{\Lambda}(x)\tilde{f}(x)$. As a result, the symmetric model implies that very few firms have a positive capital gap in the steady state.

The way in which the symmetric model generates this is through a relatively large negative drift in capital gaps and a relatively small volatility of idiosyncratic shocks to productivity. Table 2 shows the estimated drift and volatility. Recall that the capital gap is the log of capital-to-productivity ratio, $x = \log(K/z)$. The drift in capital gaps is equal to $-(\mu + \delta)$ since $-\delta$ is the drift in log capital and μ is that of log productivity. It is larger in absolute value in the symmetric benchmark, and the volatility is much smaller. This pattern is consistent across sectors.

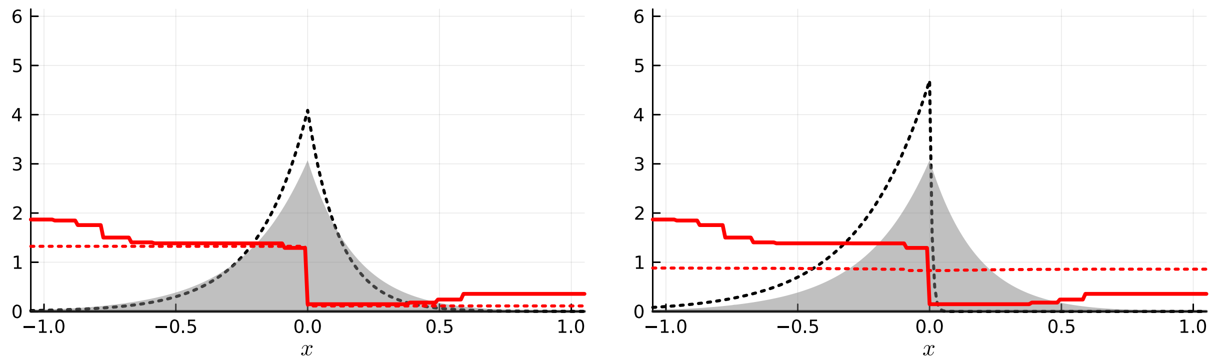
Table 2: Estimated drift and volatility for the three models

	full model	two-sided model	symmetric model
$\mu + \delta$	0.178	0.186	0.218
σ	0.272	0.294	0.062

Imposing symmetry is thus consequential for two aspects of the model. First, it strongly affects the estimated parameters of the underlying productivity process. Second, it makes the steady-state distribution much more asymmetric than the other two benchmarks.

4.2 Recovering adjustment costs

To recover the primitive distributions of adjustment costs G_i and arrival intensities κ_i, we need the true hazard Λ instead of the re-centered version $\tilde{\Lambda}(x) = \Lambda(x - x^*)$. The challenge is



(a) Full model and two-sided benchmark.

(b) Full model and symmetric benchmark.

Figure 3: Estimated generalized hazard $\tilde{\Lambda}$ and recovered steady-state distributions $\tilde{f}$. The solid line and the shaded distribution in the background correspond to the full model. Dashed lines illustrate restricted models: two-sided on the left, symmetric on the right.

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 U satisfies $U(x^*) = 0$. We can guess x^* to obtain marginal value U from equation (7) and then update the guess based on this optimality condition until convergence.

Specifically, having $\tilde{\Lambda}$ and a guess of x^*, we solve

$$(\nu + \tilde{\Lambda}(x + x^*))U(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu + \delta)U'(x) + \frac{\sigma^2}{2}U''(x) \tag{42}$$

using the procedure described in Appendix C. This procedure takes advantage of the fact that $\tilde{\Lambda}$ 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 and hence discrete sets of adjustment cost ψ. Take upward adjustments first. They happen when the capital gap is negative, $x < 0$. It is straightforward to compute the arrival intensity $\kappa_u = \lim_{x \to -\infty} \Lambda(x)$. To find the values of ψ with positive mass, use

$$\Lambda(x) = \kappa_u G_u \left(\int_x^{x^*} U(t) e^t dt \right) \tag{43}$$

The values of x at which Λ jumps map into values of ψ at which G_u jumps: for any such x, the corresponding cost is $\psi = \int_x^{x^*} U(t)e^t dt$. The size of jumps in G_u corresponds to the size of jumps in Λ scaled by κ_u. The case of downward adjustments is treated in the same way.

Figure 4 shows the recovered distributions for "Metal & Machinery". The left panel plots cumulative distribution functions G_u and G_d. We express adjustment costs ψ in percent of $(x^*)^\alpha - \nu x^*$, the instantaneous profits at the optimal level of capital gap x^*. This is the maximum attainable level of profits conditional on the environment. Firms would earn this if their capital gap was always set to x^*.

The average adjustment costs for positive investments is equal to 5.5% of this profit, and these opportunities arrive with an annual intensity of $\kappa_u = 3.07$. The opportunity for a negative adjustment arrives with an annual intensity of $\kappa_d = 0.36$, and the average cost, again, is 5.5% of maximal profits. This is an example of an industry where the average cost, by chance, is similar for positive and negative investments, although the distribution G_u and G_d are different, as Figure 4 shows.

The probability to draw a zero cost for positive adjustment is about 41%, and about the same for negative adjustments. In this particular sector, the asymmetry between positive and negative investments comes from the frequency of opportunities. Table 3 shows the same quantities for other sectors. In most of them, the average size of the cost drawn for positive and negative investments is different.

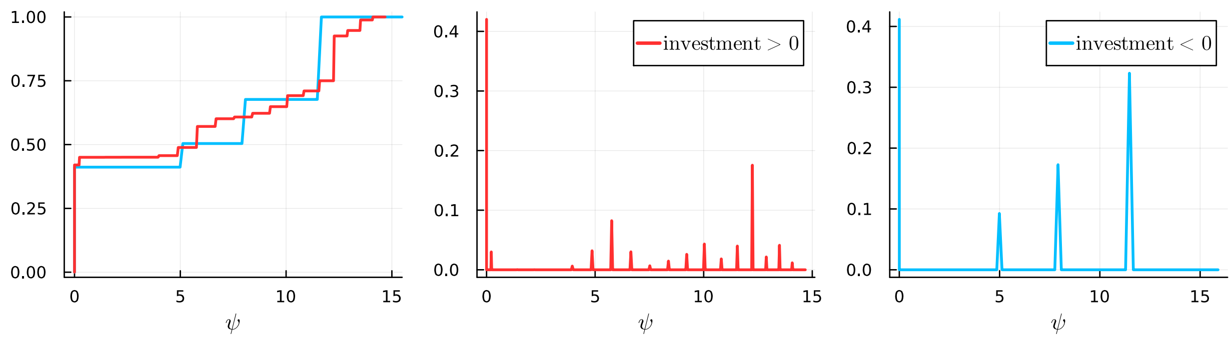


Figure 4: Recovered distributions of adjustment cost ψ. Left panel: cumulative distribution functions G_u (red) and G_d (blue) for costs of positive and negative adjustment. Center and right panels: distributions of costs of positive and negative adjustments. Costs expressed in percent of instantaneous profits at optimal capital gap $(x^*)^\alpha - \nu x^*$.

Table 3: Estimation results

Industry	κ_u	$\mathbb{E}[\psi_u]$	$\mathbb{P}\{\psi_u = 0\}$	κ_d	$\mathbb{E}[\psi_d]$	$\mathbb{P}\{\psi_d = 0\}$
Mining & Quarrying	1.874	2.292	0.623	0.277	4.009	0.371
Chemicals	2.277	3.057	0.507	0.558	3.832	0.254
Metal & Machinery	3.073	5.514	0.42	0.357	5.536	0.412
Food & Beverages	1.481	1.829	0.823	0.357	2.378	0.442
Construction	1.775	4.163	0.632	0.54	10.721	0.504
Retail	1.733	2.78	0.743	0.448	6.224	0.399
Transportation	1.827	3.709	0.671	0.53	5.914	0.406
Insurance	1.761	4.93	0.523	0.839	8.66	0.197
Health & Beauty	1.877	2.229	0.768	0.411	7.303	0.339

Panels (b) and (c) on Figure 4 show the arrival frequencies of different adjustment cost values ψ. These figures are histograms corresponding to G_u and G_d. For comparison, the two-sided model would simply put all the mass on $\psi = 0$ on both panels. The symmetric model would generate non-degenerate G_u and G_d but make them exactly the same. The corresponding figures for all sectors are in Appendix F.

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.

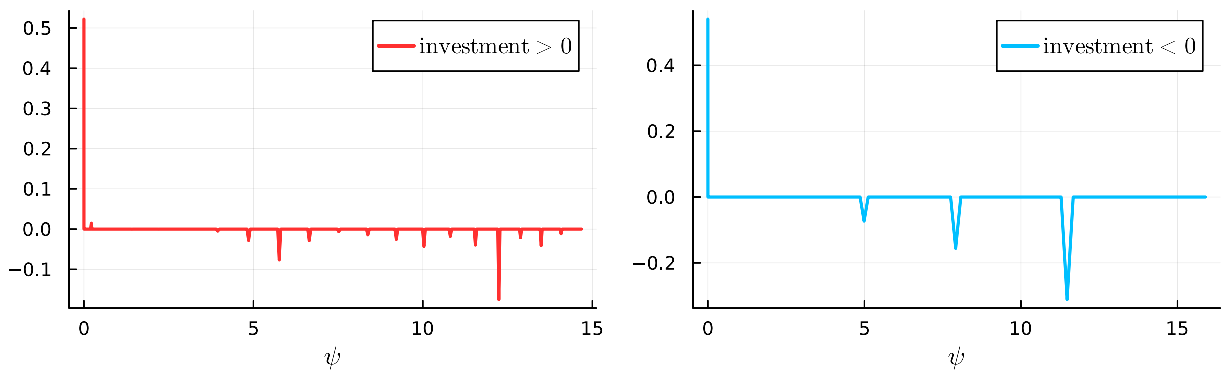


Figure 5: Difference $\hat{g}_i(\psi) - g_i(\psi)$ between the distributions of costs paid and drawn. Positive investments on the left panel, negative ones on the right. Costs expressed in percent of instantaneous profits at optimal capital gap $(x^*)^\alpha - \nu x^*$.

Figure 5 plots the difference $\hat{g}_i(\psi) - g_i(\psi)$ between the distributions of costs that firms pay and draw. In both sectors, the distribution of paid costs puts more weight on small ψ, especially on zero. The average cost paid for a positive investment in "Metal & Machinery" is just 0.08% of the maximum profit, which is 70 times lower than the average cost drawn. For negative investments, the average paid is 0.36% of the maximum profit, a 15-fold decrease relative to the average arriving cost.

5 Impulse Response Functions

Consider the economy outside of the steady state. Keeping the decision rules of the firms constant, the law of motion for the distribution of log capital gaps is

$$\partial_t f(x,t) = (\mu + \delta)\partial_x f(x,t) + \frac{\sigma^2}{2}\partial^2_{xx} f(x,t) - \Lambda(x) f(x,t) \tag{44}$$

The fact that the hazard function $\Lambda(x)$ only depends on x reflects that the firms use the steady-state decision rules. This equation is satisfied at all $x \neq x^*$.

We consider aggregate productivity shocks that shift the productivity of all firms by the same percentage. Concretely, every firm's z_t is replaced with $e^{\varepsilon} z_t$. In terms of capital gaps, a firm with x instantly moves to $x - \varepsilon$. A positive ε shifts capital gaps into negative territory.

This shock is permanent, and the productivity processes z_t continue evolving as before. This means that the initial condition for equation (44) is given by

$$f(x,0) = f(x + \varepsilon) \tag{45}$$

Here $f(x)$ with a single argument is the steady-state distribution, in a slight abuse of notation. The object of interest is the impulse response of the average capital gap given by

$$X(t;\mathcal{P}) = \int_{-\infty}^{\infty} (f(x,t) - f(x)) x dx \tag{46}$$

The cumulative impulse response $C(\mathcal{P})$ is the accumulated deviation

$$C(\mathcal{P}) = \int_0^{\infty} X(t;\mathcal{P}) dt \tag{47}$$

The argument $\mathcal{P}$ makes explicit the dependence of these objects on model parameters.

The impulse response $X(t;\mathcal{P})$ captures excess capital in the economy. A positive $X(t;\mathcal{P})$

means that firms are on average above their optimal scale in terms of capital stock. Conversely, a negative $X(t;\mathcal{P})$ means that there is too little capital. Cumulative impulse response $C(\mathcal{P})$ measures the economy's speed of adjustment. A high absolute value of $C(\mathcal{P})$ indicates that firms are slow to reach their new optimal scale, and the economy overall has too much or too little capital for long periods of time after the shock. In the economy without frictions, both $X(t;\mathcal{P})$ and $C(\mathcal{P})$ would be zero.

The structure of equation (44) suggests a useful homogeneity property.

PROPOSITION 4. Fix a parameterization $\mathcal{P} = (\mu+\delta, \sigma^2, \boldsymbol{\lambda})$. For any $\alpha > 0$, the impulse responses satisfy

$$X(t;\mathcal{P}) = X(\alpha^{-1}t;\alpha\mathcal{P}) \tag{48}$$

$$C(\mathcal{P}) = \alpha C(\alpha\mathcal{P}) \tag{49}$$

Here $\alpha\mathcal{P} = (\alpha(\mu+\delta), \alpha\sigma^2, \alpha\,\boldsymbol{\lambda})$ is a rescaling of $\mathcal{P}$.

Scaling all parameters by a positive constant is exactly like changing the time units in the model. Doubling the drift, the variance, and the adjustment rate makes convergence back to the steady-state distribution twice as fast. Cumulative impulse response halves.

Corollary 1 states that doubling the parameters also doubles the steady-state adjustment frequency. Taken together, these facts imply that the model can be first estimated under a normalization $N = 1$, and the cumulative impulse response just needs to be scaled by the true frequency in the data. Alternatively, the model can be estimated for a fixed σ^2 or $\mu+\delta$. In this case, C has to be scaled by the ratio of the true frequency to that in equation (23).

In the single-hazard Calvo model, this frequency is equal to the hazard rate λ, and the impulse response function has a simple exponential form:

PROPOSITION 5. Consider a single-hazard parameterization $\mathcal{P}_1 = (\mu+\delta, \sigma^2, \lambda)$. The impulse response to a one-time permanent aggregate productivity shock of the size ε is

$X(t;\mathcal{P}_1) = -\varepsilon e^{-\lambda t}$, and the cumulative impulse response is $C(\mathcal{P}_1) = -\varepsilon/\lambda$.

This impulse response does not depend on $\mu+\delta$ and σ^2. It is also symmetric and, in fact, linear in ε. Both of these properties are lost when we generalize the model to a two-sided distribution case and beyond.

In the general case of the piece-wise generalized hazard function, it is possible to characterize the slope of the impulse response function at $t = 0$. This slope is not a complete summary of the impulse response but it reflects how fast the economy adjusts initially. We use it to illustrate asymmetries with respect to the sign of the shock. The proposition below provides a general characterization, after which we demonstrate a tractable special case of the two-sided model.

Proposition 6. Consider a model $\mathcal{P} = (\mu+\delta, \sigma^2, \boldsymbol{\lambda})$ that generates an adjustment frequency N. The slope of the impulse response to a productivity shock of size ε at $t = 0$ is

$$\begin{aligned}\partial_t X(0;\mathcal{P}) = \varepsilon N &+ (\lambda_1 - \lambda_{-1}) \int_{x^*}^{x^*+\varepsilon} f(x)(x-\varepsilon-x^*)dx \qquad (50)\\ &+ \sum_{j<0}(\lambda_j - \lambda_{j-1}) \int_{x_j}^{x_j+\varepsilon} f(x)(x-\varepsilon-x^*)dx \qquad (51)\\ &+ \sum_{j>0}(\lambda_{j+1} - \lambda_j) \int_{x_j}^{x_j+\varepsilon} f(x)(x-\varepsilon-x^*)dx\end{aligned}$$

This object is non-linear in ε and asymmetric around $\varepsilon = 0$, meaning that the economy's adjustment has fundamentally different speeds for positive and negative shocks. Alvarez et al. (2022) defined the flexibility index $\mathcal{F}$ as the derivative of $X'(0)$ with respect to ε. Proposition 6 implies the following form for this measure:

Corollary 2. The flexibility index is

$$\mathcal{F} = N + \sum_{j<0}(\lambda_j - \lambda_{j-1}) f(x_j)(x_j - x^*) + \sum_{j>0}(\lambda_{j+1} - \lambda_j) f(x_j)(x_j - x^*) \qquad (52)$$

In the Calvo case, only N is left. This is intuitive because firms adjust for exogenous reasons, and the economy's speed of adjustment only depends on the arrival frequency of these opportunities. In the two-sided case, this is also the only term that remains: small aggregate shocks change the hazard of adjustments $\Lambda(x)$ only for firms in close proximity of x^*, where adjustments are small anyway.

We therefore have to go beyond small ε to illustrate the differences between the Calvo and two-sided cases. Proposition 6 still leads to a tractable expression for the two-sided model. Consider a Calvo model and a two-sided model with the same adjustment frequency in the steady state. The differences in slopes depend

Corollary 3. Consider a single-hazard model $\mathcal{P}_1 = (\mu + \delta, \sigma^2, \lambda)$ and a two-sided model $\mathcal{P}_2 = (\tilde{\mu} + \tilde{\delta}, \tilde{\sigma}^2, \{\lambda_u, \lambda_d\})$ in which $\Lambda(x) = \lambda_u$ for $x < x^*$ and $\Lambda(x) = \lambda_d$ for $x > x^*$. If the steady-state adjustment frequency is the same in these models,

$$X'(0; \mathcal{P}_2) = X'(0; \mathcal{P}_1) + (\lambda_u - \lambda_d) \int_{x^*+\varepsilon}^{x^*} (x - \varepsilon - x^*) f(x) dx \qquad \text{if } \varepsilon < 0 \tag{53}$$

$$X'(0; \mathcal{P}_2) = X'(0; \mathcal{P}_1) - (\lambda_u - \lambda_d) \int_{x^*}^{x^*+\varepsilon} (x - \varepsilon - x^*) f(x) dx \qquad \text{if } \varepsilon > 0 \tag{54}$$

Here $f(\cdot)$ is the steady-state distribution of investment gaps in the two-sided model.

Note that the result does not depend on the drift and volatility of the idiosyncratic shocks. They may even be different in the two models.

The integral term is similar in equation (54) and equation (53). The sign in front of it changes because the sign of the slope $X'(0)$ is different for positive and negative shocks. The difference between $X'(0; \mathcal{P}_2)$ and $X'(0; \mathcal{P}_1)$ arises because, in the two-sided case, it matters whether the shock shifts the bulk of the distribution toward a higher or lower adjustment hazard. Suppose $\lambda_d < \lambda_u$, meaning that the negative investment hazard is lower. The two models have the same frequency in steady state, so it must be that $\lambda_d < \lambda < \lambda_u$. If the shock to productivity is negative ($\varepsilon < 0$), firms are shifted to the positive gap territory. Negative investment opportunities are taken less frequently in the two-sided model, $\lambda_d < \lambda$, so investment gaps will stay high for longer than with a single hazard. The slopes of the impulse responses are negative, and they are less negative in the two-sided model.

If, on the other hand, the productivity shock is positive ($\varepsilon > 0$), the two-sided economy will adjust faster. Both slopes are positive since gaps are shifted primarily to the negative territory and adjust upwards on aggregate. The slope in the two-sided model is higher because positive investment opportunities in this model are taken more frequently, $\lambda_u > \lambda$.

The differences in equation (54) and equation (53) are second-order in ε, as follows from Corollary 2 that shows first-order equivalence. The second-order terms are readily computed by differentiation:

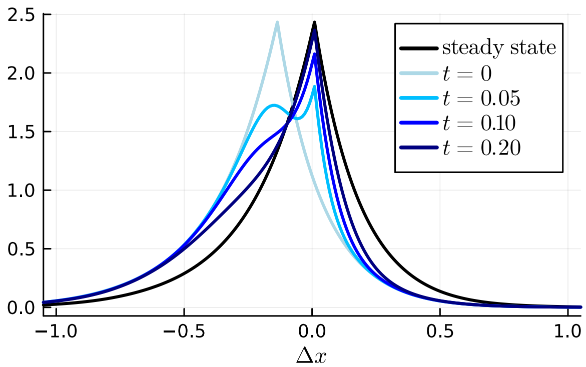
$$X'(0; \mathcal{P}_2) = X'(0; \mathcal{P}_1) + (\lambda_u - \lambda_d) f(x^*) \varepsilon^2 + o(\varepsilon^2) \tag{55}$$

In the following orders, the derivatives at $\varepsilon = 0$ do not exist in general, since $f(x)$ does not have derivatives at $x = x^*$.

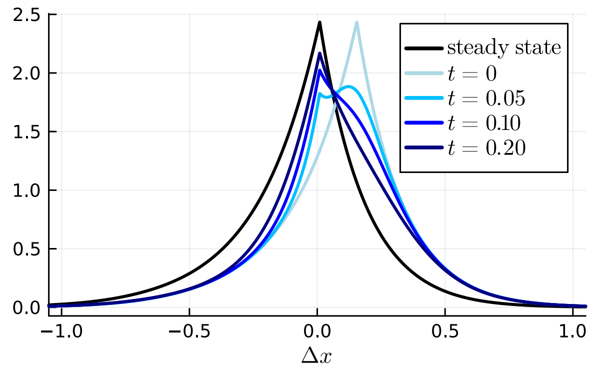
5.1 Comparing impulse responses

We next trace out the impulse response functions for the nine sectors in our data. Figure 6 shows the evolution of the capital gap distribution after a 15% dislocation in "Metal & Machinery". We choose the size of the shock close to the average adjustment in the data.

Immediately after the shock, the whole density function moves to the left or to the right, depending on the sign of ε. Right after $t = 0$, the density loses differentiability at x^*, which is normalized to zero in the figure. This is because firms immediately begin reinjecting at x^*. The density slowly reverts to the steady-state position as time progresses.



(a) Positive shock to productivity.

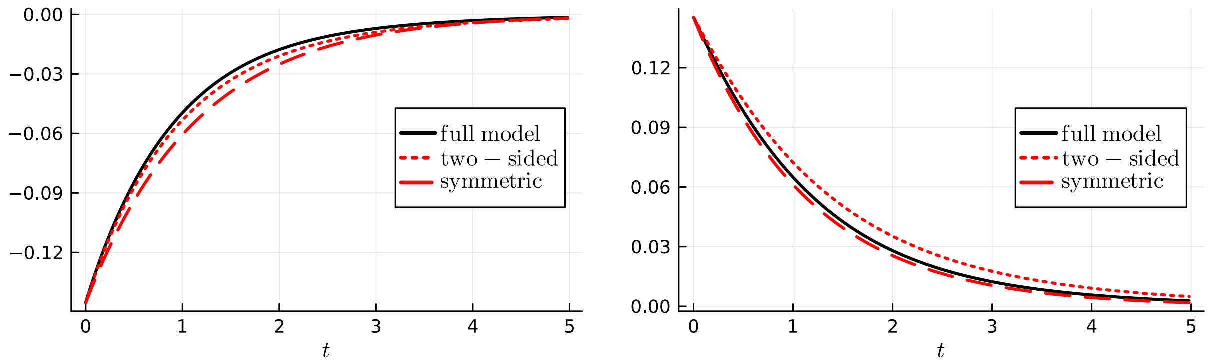


(b) Negative shock to productivity.

Figure 6: Evolution of the capital gap distribution after a positive and a negative shock to productivity. The absolute size of the shock is 15% in both cases.

It is evident from Figure 6 that reversion is slower when the productivity shock is negative. Opportunities to adjust capital stock downwards arrive less frequently, so positive capital gaps persist. Positive aggregate shocks to productivity generate a fast build-up of capital, and negative shocks unfold slowly.

We compare impulse responses across the three models, replacing our symmetric benchmark with a simple Calvo model. We do it for two reasons. First, in the Calvo model, there is a closed-form solution for the impulse response functions, and there is full symmetry with respect to the size of the shock. This illustrative special case is useful for highlighting asymmetries in the other two specifications. Second, our estimates in the symmetric case are quantitatively very close to Calvo model.



(a) Positive shock to productivity.

(b) Negative shock to productivity.

Figure 7: Impulse responses of the average capital gap x across models. Positive shock to productivity on the left panel, negative on the right. The absolute size of the shock is 15%.

Figure 7 shows the impulse response functions in the full, two-sided, and symmetric models. Unsurprisingly, responses are symmetric with respect to the sign of the shock in the Calvo model. The generalized adjustment hazard is symmetric around zero, so firms adjust with the same speed regardless of the initial dislocation. In the asymmetric models, the response to a negative shock is more sluggish. This is because the recovered generalized hazard is much lower for positive capital gaps.

The two-sided model shows a faster initial adjustment (around $t = 0$) than the symmetric model for a positive shock. For a negative shock, the relationship reverses, and the symmetric model economy adjusts faster. The symmetric model is quantitatively close to a simple Calvo model, so the intuition is the same as in Proposition 6.

Table 4 shows cumulative impulse responses to positive and negative shocks of the same absolute size. The symmetric model is very close to the Calvo specification, and the cumulative impulse response (CIR) in it is almost symmetric. Whether it overestimates or underestimates the CIR depends on the size of the shock. The two-sided model performs better than the symmetric one for a positive shock and much worse for a negative one.

Table 4: Cumulative impulse responses in the three models.

	full model	two-sided model	symmetric model
15%	0.149	0.157	0.177
-15%	-0.186	-0.224	-0.177

We next focus on cumulative impulse responses and explore the asymmetry and non-linearity with respect to the size of the shock. Figure 8 plots cumulative impulse responses for the baseline model, the two-sided model, and the Calvo model. We use the latter instead of our symmetric benchmark because they are quantitatively very similar, and the plain Calvo specification produces an impulse response with a simple formula.

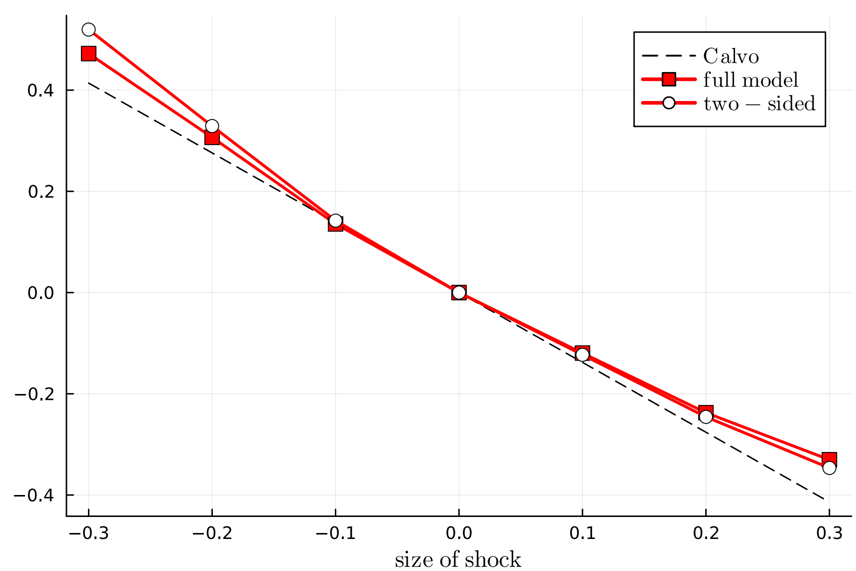


Figure 8: CIR of capital gaps as a function of the shock to productivity.

The Calvo model, expectedly, produces a linear cumulative impulse response. The intuition for this is simple: however far the distribution shifts, firms will not start adjusting more frequently, and shocks of different sizes only differ along the intensive margin. Capital gaps will spend the same time away from the steady-state distribution, but the dislocations will be larger for larger shocks. The CIR for any model other than Calvo is generically non-linear in the size of the shock and asymmetric. Asymmetry is clearly seen on Figure 8.

For positive shocks, the Calvo model overestimates the CIR relative to the baseline. The reason is that the generalized hazard functions are tightly connected between the two models. They have to fit the same adjustment frequency, so the harmonic average of Λ weighted with the observed distribution of adjustments in them has to coincide, according to Proposition 3. In the baseline model, this means that Λ is lower than that in Calvo for $x < 0$ and higher for $x > 0$. The shock thus shifts the distribution of firms into the territory with a higher adjustment probability than in Calvo. This makes the economy adjust faster. For negative shocks, the Calvo model underestimates the CIR since it is hardwired to impute a larger arrival intensity for disinvestment opportunities than the less restricted models.

The two-sided distribution model also slightly overestimates the response. It does it to a different extent depending on the sign of the shock. For negative shocks, overestimation is more pronounced, since the two-sided model assigns a low adjustment hazard to negative adjustments that does not grow with distance from $x = x^*$. Allowing for asymmetry in the model without allowing for an increasing hazard leads to overshooting in this case.

References

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.

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

G. Bertola and R. J. Caballero. Irreversibility and aggregate investment. The Review of Economic Studies, 61(2):223–246, 1994.

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.

A. Cavallo, F. Lippi, and K. Miyahara. Large shocks travel fast. Working Paper 31659, National Bureau of Economic Research, September 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.

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

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

L. Guiso, L. Pistaferri, and F. Schivardi. Insurance within the firm. Journal of Political Economy, 113(5):1054–1087, 2005.

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.

A. Khan and J. K. Thomas. Adjustment costs. The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke, 2008.

T. Winberry. Lumpy investment, business cycles, and stimulus policy. American Economic Review, 111(1):364–396, 2021.

A 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

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

The optimality condition is

$$\partial_k V(k^*z, z) = 1 \qquad \text{(A.2)}$$

The value-matching conditions at all cutoffs are

$$V(k^*z,z) - k^*z - V(K,z) + K = \boldsymbol{\psi}^i(K/z), \text{ for } i \in \{u,d\} \qquad \text{(A.3)}$$

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 k^d and k^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.

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

$$\partial_z V(k^*z,z) - \partial_z V(k^i z, z) = \boldsymbol{\psi}_i, \text{ for } i \in \{u,d\} \qquad \text{(A.4)}$$

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

$$V(K,z) = zv(K/z) + K \qquad \text{(A.5)}$$

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:

$$\partial_k V(K,z) = v'\left(\frac{K}{z}\right) + 1 \tag{A.6}$$

$$\partial_z V(K,z) = v\left(\frac{K}{z}\right) - \frac{K}{z}v'\left(\frac{K}{z}\right) \tag{A.7}$$

$$\partial_{zz} V(K,z) = \frac{K^2}{z^3}v''\left(\frac{K}{z}\right) \tag{A.8}$$

Plugging the relations for derivatives of $V(\cdot)$ into equation (A.1) and denoting $k = K/z$,

$$\begin{aligned} rzv(k) + rzk = zk^\alpha - \delta zkv'(k) - \delta zk + \left(\mu + \frac{\sigma^2}{2}\right)(zv(k) - zkv'(k)) + \frac{\sigma^2}{2}k^2 v''(k) \\ + z\sum\nolimits_{i=u,d} \mathbb{1}_i \kappa_i \int \max\{v(k^*) - v(k) - \psi, 0\} dG_i(\psi) \end{aligned} \tag{A.9}$$

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

$$\begin{aligned} \rho v(k) = k^\alpha - \nu k + (\rho - \nu)kv'(k) + \frac{\sigma^2}{2}k^2 v''(k) \\ + \sum\nolimits_{i=u,d} \mathbb{1}_i \kappa_i \int \max\{v(k^*) - v(k) - \psi, 0\} dG_i(\psi) \end{aligned} \tag{A.10}$$

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

$$v(k^*) - v(k) = \boldsymbol{\psi}^i(k)a(k) \tag{A.11}$$

$$v(k^*) - k^* v'(k^*) - v(k^i) + k^i v'(k^i) = \boldsymbol{\psi}^i \tag{A.12}$$

for $i \in \{u, d\}$. Since $v'(k^*) = 0$, these two equations together imply $v'(k^i) = -\boldsymbol{\psi}^i$ for both $i \in \{u, d\}$. Taking the derivative of equation (A.10) with respect to k and denoting $u(k) = v'(k)$,

$$\begin{aligned} \rho u(k) = \alpha k^{\alpha-1} - \nu + (\rho - \nu)u(k) + (\rho - \nu + \sigma^2)ku'(k) + \frac{\sigma^2}{2}k^2 u''(k) \\ = \sum_{i=u,d} \mathbb{1}_i \kappa_i \int \mathbb{1}\{v(k^*) - v(k) \geq \psi\} u(k) dG_i(\psi) \end{aligned} \tag{A.13}$$

Rearranging,

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

The generalized hazard function here is

$$\lambda(k) = \sum_{i=u,d} \mathbb{1}_i \kappa_i G_i \left(v(k^*) - v(k)\right) \tag{A.14}$$

Plugging,

$$(\nu + \lambda(k))u(k) = \alpha k^{\alpha-1} - \nu + (\rho - \nu + \sigma^2)ku'(k) + \frac{\sigma^2}{2}k^2 u''(k) \tag{A.15}$$

Boundary conditions for this equation are $u(k^u) = -\boldsymbol{\psi}^u$, $u(k^d) = -\boldsymbol{\psi}^d$, and $u(k^*) = 0$. Again, the smooth-pasting conditions $u(k^u) = -\boldsymbol{\psi}^u$ and $u(k^d) = -\boldsymbol{\psi}^d$ need not hold if the normal fixed cost is infinite. In this case, $k^u = 0$ and $k^d = \infty$, and the firm does not take a decision that would call for smooth pasting.

Recovering $v(\cdot)$ from $u(\cdot)$ can start at k^*:

$$\rho v(k^*) = (k^*)^\alpha - \nu k^* + \frac{\sigma^2}{2}(k^*)^2 u'(k^*) \tag{A.16}$$

The rest can be obtained by integration:

$$v(k) - v(k^*) = \int_{k^*}^{k} u(x)dx \tag{A.17}$$

Proof. (of Proposition 1). Conditions (15)-(21) contain exactly $2(U + D)$ equations that are linear in $\{\eta_{1,j}, \eta_{2,j}\}$. Equation (21) can be rewritten as

$$\begin{aligned} 1 = & \sum_{j=-U}^{-1} \left(\frac{\eta_{1,j}}{\xi_{1,j}}(e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}) + \frac{\eta_{2,j}}{\xi_{2,j}}(e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j}) \right) \\ & + \sum_{j=1}^{D} \left(\frac{\eta_{1,j}}{\xi_{1,j}}(e^{\xi_{1,j}x_j} - e^{\xi_{1,j}x_{j-1}}) + \frac{\eta_{2,j}}{\xi_{2,j}}(e^{\xi_{2,j}x_j} - e^{\xi_{2,j}x_{j-1}}) \right) \end{aligned} \tag{A.18}$$

Using this and equations (15)-(20), 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 $-U+1 \le j \le -1$, set $k = j+U+1$ and

$$\mathbb{A}_{2k-1,2k-3} = -e^{\xi_{1,j-1}x_j} \tag{A.19}$$

$$\mathbb{A}_{2k-1,2k-2} = -e^{\xi_{2,j-1}x_j} \tag{A.20}$$

$$\mathbb{A}_{2k-1,2k-1} = e^{\xi_{1,j}x_j} \tag{A.21}$$

$$\mathbb{A}_{2k-1,2k} = e^{\xi_{2,j}x_j} \tag{A.22}$$

$$\mathbb{A}_{2k,2k-3} = -\xi_{1,j-1}e^{\xi_{1,j-1}x_j} \tag{A.23}$$

$$\mathbb{A}_{2k,2k-2} = -\xi_{2,j-1}e^{\xi_{2,j-1}x_j} \tag{A.24}$$

$$\mathbb{A}_{2k,2k-1} = \xi_{1,j}e^{\xi_{1,j}x_j} \tag{A.25}$$

$$\mathbb{A}_{2k,2k} = \xi_{2,j}e^{\xi_{2,j}x_j} \tag{A.26}$$

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

$$\mathbb{A}_{2k-1,2k-1} = e^{\xi_{1,j}x_j} \tag{A.27}$$

$$\mathbb{A}_{2k-1,2k} = e^{\xi_{2,j}x_j} \tag{A.28}$$

$$\mathbb{A}_{2k-1,2k+1} = -e^{\xi_{1,j+1}x_j} \tag{A.29}$$

$$\mathbb{A}_{2k-1,2k+2} = -e^{\xi_{2,j+1}x_j} \tag{A.30}$$

$$\mathbb{A}_{2k,2k-1} = \xi_{1,j}e^{\xi_{1,j}x_j} \tag{A.31}$$

$$\mathbb{A}_{2k,2k} = \xi_{2,j}e^{\xi_{2,j}x_j} \tag{A.32}$$

$$\mathbb{A}_{2k,2k+1} = -\xi_{1,j+1}e^{\xi_{1,j+1}x_j} \tag{A.33}$$

$$\mathbb{A}_{2k,2k+2} = -\xi_{2,j+1}e^{\xi_{2,j+1}x_j} \tag{A.34}$$

The remaining rows $\{1, 2, 2(D+U)-1, 2(D+U)\}$. They encode (20), (17), and (21). This is achieved by setting $\mathbb{A}_{1,1} = 1$, $\mathbb{A}_{2(U+D),2(U+D)} = 1$, $(\mathbb{A}_{2,2U-1}, \mathbb{A}_{2,2U}, \mathbb{A}_{2,2U+1}, \mathbb{A}_{2,2U+2}) = (1, 1, -1, -1)$, and

$$\mathbb{A}_{2(U+D)-1,2k-1} = \frac{1}{\xi_{1,j}}(e^{\xi_{1,j}x_{j+1}} - e^{\xi_{1,j}x_j}\mathbb{1}\{j > -U\}),\ k = j+U+1,\ -U \le j \le -1 \tag{A.35}$$

$$\mathbb{A}_{2(U+D)-1,2k} = \frac{1}{\xi_{2,j}}(e^{\xi_{2,j}x_{j+1}} - e^{\xi_{2,j}x_j}),\ k = j+U+1,\ -U \le j \le -1 \tag{A.36}$$

$$\mathbb{A}_{2(U+D)-1,2k-1} = \frac{1}{\xi_{1,j}}(e^{\xi_{1,j}x_j} - e^{\xi_{1,j}x_{j-1}}),\ k = j+U,\ 1 \le j \le D \tag{A.37}$$

$$\mathbb{A}_{2(U+D)-1,2k} = \frac{1}{\xi_{2,j}}(e^{\xi_{2,j}x_j}\mathbb{1}\{j < D\} - e^{\xi_{2,j}x_{j-1}}),\ k = j+U,\ 1 \le j \le D \tag{A.38}$$

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 $-U \le j \le -1$ and $(\eta_{1,j}, \eta_{2,j}) = (\boldsymbol{\eta}_{2(j+U)-1}, \boldsymbol{\eta}_{2(j+U)})$ for $1 \le j \le D$. □

Proof. (of Corollary 1). It follows from inspecting equation (23) and equation (24).

B Alternative formulation

In this section, 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 κ_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 κ_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 $K>k^*z_t$, firms only adjust down, and do this if and only if the corresponding cost reduction arrives and the new value drawn ψ satisfies $\psi\leq\boldsymbol{\psi}^d(K/z)$. Here $\boldsymbol{\psi}^d(\cdot)$ is a cutoff function. When $K>k^*z_t$, firms only adjust up, and do this if and only if the corresponding cost reduction arrives and the new value drawn ψ satisfies $\psi\leq\boldsymbol{\psi}^u(K/z)$. The function $\boldsymbol{\psi}^d(\cdot)$ maps $[k^*,k^d]$ to $[0,\psi^d]$, and $\boldsymbol{\psi}^u(\cdot)$ maps $[k^u,k^*]$ to $[0,\psi^u]$. The thresholds k^d and k^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

$$r\overline{V}(K,z)=z^{1-\alpha}K^\alpha-(r+\delta)K-\delta K\partial_k\overline{V}(K,z)+\left(\mu+\frac{\sigma^2}{2}\right)z\partial_z\overline{V}(K,z)+\frac{\sigma^2}{2}z^2\partial_{zz}\overline{V}(K,z)$$
$$+\sum\nolimits_{i=u,d}\mathbb{1}_i\kappa_i\int\max\{\overline{V}(k^*z,z)-\overline{V}(K,z)-\psi z,0\}dG_i(\psi)\tag{A.39}$$

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

$$\overline{V}(k^*z,z)-\overline{V}(K,z)=\boldsymbol{\psi}^i(K/z)z,\text{ for } i\in\{u,d\}\tag{A.40}$$

The smooth-pasting conditions are

$$\partial_z\overline{V}(k^*z,z)-\partial_z\overline{V}(k^iz,z)=\psi^i,\text{ for } i\in\{u,d\}\tag{A.41}$$

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

$$\overline{V}(K,z)=zv(K/z)\tag{A.42}$$

Note the following relations for derivatives:

$$\partial_k \overline{V}(K, z) = v' \left(\frac{K}{z} \right) \tag{A.43}$$

$$\partial_z \overline{V}(K, z) = v \left(\frac{K}{z} \right) - \frac{K}{z} v' \left(\frac{K}{z} \right) \tag{A.44}$$

$$\partial_{zz} \overline{V}(K, z) = \frac{K^2}{z^3} v'' \left(\frac{K}{z} \right) \tag{A.45}$$

Plugging this into equation (A.39) and denoting $k = K/z$,

$$\begin{aligned} rzv(k) = zk^\alpha - (r+\delta)zk - \delta zkv'(k) + \left(\mu + \frac{\sigma^2}{2} \right) (v(k) - kv'(k)) + \frac{\sigma^2}{2} zk^2 v''(k) \\ + z \sum\nolimits_{i=u,d} \mathbb{1}_i \kappa_i \int \max\{v(k^*) - v(k) - \psi, 0\} dG_i(\psi) \end{aligned} \tag{A.46}$$

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

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

This equation coincides with equation (3). Notice that $v'(k^*) = v'(k^u) = v'(k^d) = 0$, so all decisions are exactly the same as in the baseline.

C Recovering marginal value $u(\cdot)$

In this section, we describe the algorithm to recover the marginal value function $u(\cdot)$. It is convenient first to transform the capital-to-productivity ratio into its logarithm. Denoting $x = \ln(k)$, transform $u(\cdot) : (0, \infty) \mapsto \mathbb{R}$ into a function $\Upsilon(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ such that $\Upsilon(x) = u(k)$. Similarly, transform $\lambda(\cdot) : (0, \infty) \mapsto \mathbb{R}$ into a function $\Lambda(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ such that $\Lambda(x) = \lambda(k)$.

Rewriting equation (5),

$$\nu(x)\Upsilon(x) = \alpha e^{(\alpha-1)x} - \nu - (\mu + \delta)\Upsilon'(x) + \frac{\sigma^2}{2}\Upsilon''(x) \tag{A.47}$$

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

Now suppose $\Lambda(\cdot)$ is piece-wise constant. Specifically, let there be $U + D + 1$ nodes given by $\{y_j\} = \{x_j + x^*\}$ such that $\Lambda(\cdot)$ is constant on any $(x_{j-1} + x^*, x_j + x^*)$ for $1 \le j \le D$, any $(x_j + x^*, x_{j+1} + x^*)$ for $-U \le j \le -1$:

- $\Lambda(x) = \lambda_j$ on $(x_{j-1} + x^*, x_j + x^*)$ for $1 \le j \le D$
- $\Lambda(x) = \lambda_j$ on $(x_j + x^*, x_{j+1} + x^*)$ for $-U \le j \le -1$

The leftmost and rightmost nodes are infinite: $x_{-U} = -\infty$ and $x_D = \infty$. Denote $\nu_j = \nu + \lambda_j$ and let $\Upsilon_j(\cdot)$ be the part of $\Upsilon(\cdot)$ defined on the segment j. The solution to equation (A.47) on any segment j is

$$\Upsilon_j(x) = \eta_{1,j} e^{\xi_{1,j} x} + \eta_{2,j} e^{\xi_{2,j} x} + \theta_{1,j} e^{(\alpha-1)x} + \theta_{2,j} \tag{A.48}$$

The non-homogeneous part can be recovered immediately:

$$\theta_{1,j} = \frac{\alpha}{\nu_j - (1-\alpha)(\mu+\delta) - (1-\alpha)^2\sigma^2/2} \tag{A.49}$$

$$\theta_{2,j} = -\frac{\nu}{\nu_j} \tag{A.50}$$

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

$$\{\xi_{1,j}, \xi_{2,j}\} = \frac{\mu + \delta \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2\nu_j}}{\sigma^2} \tag{A.51}$$

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

$$\Upsilon_{j-1}(y_j) = \Upsilon_j(y_j) \tag{A.52}$$

$$\Upsilon'_{j-1}(y_j) = \Upsilon'_j(y_j) \tag{A.53}$$

This yields $2(U+D-1)$ conditions. Two other equations are $\eta_{1,-U} = 0$ and $\eta_{2,D} = 0$ ensuring that the homogeneous part of $\Upsilon(\cdot)$ 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 $\Upsilon(\cdot)$, it is straightforward to recover $u(k) = \Upsilon(\ln(k))$ and integrate $v(k) - v(k^*)$:

$$v(k) - v(k^*) = \int_{k^*}^{k} u(t)dt = \int_{x^*}^{x} \Upsilon(x)e^x dx \tag{A.54}$$

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

Algorithm. The conditions on $\eta_{1,j}$ and $\eta_{2,j}$ combine into a linear system of dimensionality $2U + 2J$. Let the vector $\boldsymbol{\eta}$ combine the unknowns in the following way: $\boldsymbol{\eta}_{2U+2j+1} = \eta_{1,j}$ and $\boldsymbol{\eta}_{2U+2j+2} = \eta_{2,j}$ for all j such that $-U \leq j \leq -1$. For $1 \leq j \leq D$, set $\boldsymbol{\eta}_{2U+2j-1} = \eta_{1,j}$ and $\boldsymbol{\eta}_{2U+2j} = \eta_{2,j}$.

Let A be a square matrix of size $(2U + 2D) \times (2U + 2D)$ and b be a column vector of length $2U + 2D$. The $(2j + 2U - 1)$-th rows of $\mathbb{A}$ and $\mathbf{b}$ represent equation (A.52) for $-U \leq j < -1$:

$$\begin{aligned} \eta_{1,j}e^{\xi_{1,j}y_{j+1}} + \eta_{2,j}e^{\xi_{2,j}y_{j+1}} - \eta_{1,j+1}e^{\xi_{1,j+1}y_{j+1}} - \eta_{2,j+1}e^{\xi_{2,j+1}y_{j+1}} \\ = (\theta_{1,j+1} - \theta_{1,j})e^{(1-\alpha)y_{j+1}} + \theta_{2,j+1} - \theta_{2,j} \end{aligned} \tag{A.55}$$

The $(2j + 2U)$-th rows represent equation (A.53) for $-U \leq j < -1$:

$$\begin{aligned} \eta_{1,j}\xi_{1,j}e^{\xi_{1,j}y_{j+1}} + \eta_{2,j}\xi_{2,j}e^{\xi_{2,j}y_{j+1}} - \eta_{1,j+1}\xi_{1,j+1}e^{\xi_{1,j+1}y_{j+1}} - \eta_{2,j+1}\xi_{2,j+1}e^{\xi_{2,j+1}y_{j+1}} \\ = (\theta_{1,j+1} - \theta_{1,j})(1-\alpha)e^{(1-\alpha)y_{j+1}} \end{aligned} \tag{A.56}$$

Then, the rows $2U - 1$ takes care of continuity at $y_0 = x^*$:

$$\begin{aligned} \eta_{1,-1}e^{\xi_{1,-1}y_0} + \eta_{2,-1}e^{\xi_{2,-1}y_0} - \eta_{1,1}e^{\xi_{1,1}y_0} - \eta_{2,1}e^{\xi_{2,1}y_0} \\ = (\theta_{1,1} - \theta_{1,-1})e^{(1-\alpha)y_0} + \theta_{2,1} - \theta_{2,-1} \end{aligned} \tag{A.57}$$

The row $2U$ takes care of differentiability at $y_0 = x^*$:

$$\begin{aligned} \eta_{1,-1}\xi_{1,-1}e^{\xi_{1,-1}y_0} + \eta_{2,-1}\xi_{2,-1}e^{\xi_{2,-1}y_0} - \eta_{1,1}\xi_{1,1}e^{\xi_{1,1}y_0} - \eta_{2,1}\xi_{2,1}e^{\xi_{2,1}y_0} \\ = (\theta_{1,1} - \theta_{1,-1})(1-\alpha)e^{(1-\alpha)y_0} \end{aligned} \tag{A.58}$$

Then, the $(2j + 2U - 3)$-th rows of $\mathbb{A}$ and $\mathbf{b}$ represent equation (A.52) for $1 < j \leq D$:

$$\begin{aligned} \eta_{1,j}e^{\xi_{1,j}y_{j-1}} + \eta_{2,j}e^{\xi_{2,j}y_{j-1}} - \eta_{1,j-1}e^{\xi_{1,j-1}y_{j-1}} - \eta_{2,j-1}e^{\xi_{2,j-1}y_{j-1}} \\ = (\theta_{1,j-1} - \theta_{1,j})e^{(1-\alpha)y_{j-1}} + \theta_{2,j-1} - \theta_{2,j} \end{aligned} \tag{A.59}$$

The $(2j+2U-2)$-th rows represent equation (A.53) for $1 < j \le D$:

$$\eta_{1,j}\xi_{1,j}e^{\xi_{1,j}y_{j-1}} + \eta_{2,j}\xi_{2,j}e^{\xi_{2,j}y_{j-1}} - \eta_{1,j-1}\xi_{1,j-1}e^{\xi_{1,j-1}y_{j-1}} - \eta_{2,j-1}\xi_{2,j-1}e^{\xi_{2,j-1}y_{j-1}} \\ = (\theta_{1,j-1} - \theta_{1,j})(1-\alpha)e^{(1-\alpha)y_{j-1}} \tag{A.60}$$

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

$$\mathbb{A}_{2j+2U+1,2j+2U+1} = e^{\xi_{1,j}y_{j+1}} \tag{A.61}$$
$$\mathbb{A}_{2j+2U+1,2j+2U+2} = e^{\xi_{2,j}y_{j+1}} \tag{A.62}$$
$$\mathbb{A}_{2j+2U+1,2j+2U+3} = -e^{\xi_{1,j+1}y_{j+1}} \tag{A.63}$$
$$\mathbb{A}_{2j+2U+1,2j+2U+4} = -e^{\xi_{2,j+1}y_{j+1}} \tag{A.64}$$
$$\mathbb{A}_{2j+2U+2,2j+2U+1} = \xi_{1,j}e^{\xi_{1,j}y_{j+1}} \tag{A.65}$$
$$\mathbb{A}_{2j+2U+2,2j+2U+2} = \xi_{2,j}e^{\xi_{2,j}y_{j+1}} \tag{A.66}$$
$$\mathbb{A}_{2j+2U+2,2j+2U+3} = -\xi_{1,j+1}e^{\xi_{1,j+1}y_{j+1}} \tag{A.67}$$
$$\mathbb{A}_{2j+2U+2,2j+2U+4} = -\xi_{2,j+1}e^{\xi_{2,j+1}y_{j+1}} \tag{A.68}$$

The vector $\mathbf{b}$ for $-U \le j \le -1$ is filled as follows:

$$\mathbf{b}_{2j+2U+1} = (\theta_{1,j+1} - \theta_{1,j})e^{(1-\alpha)y_{j+1}} + \theta_{2,j+1} - \theta_{2,j} \tag{A.69}$$
$$\mathbf{b}_{2j+2U+2} = (\theta_{1,j+1} - \theta_{1,j})(1-\alpha)e^{(1-\alpha)y_{j+1}} \tag{A.70}$$

The rows $2J-1$ and $2J$ of the matrix $\mathbb{A}$ are

$$\mathbb{A}_{2U-1,2U-1} = e^{\xi_{1,-1}y_0} \tag{A.71}$$
$$\mathbb{A}_{2U-1,2U} = e^{\xi_{2,-1}y_0} \tag{A.72}$$
$$\mathbb{A}_{2U-1,2U+1} = -e^{\xi_{1,1}y_0} \tag{A.73}$$
$$\mathbb{A}_{2U-1,2U+2} = -e^{\xi_{2,1}y_0} \tag{A.74}$$
$$\mathbb{A}_{2U,2U-1} = \xi_{1,-1}e^{\xi_{1,-1}y_0} \tag{A.75}$$
$$\mathbb{A}_{2U,2U} = \xi_{2,-1}e^{\xi_{2,-1}y_0} \tag{A.76}$$
$$\mathbb{A}_{2U,2U+1} = -\xi_{1,1}e^{\xi_{1,1}y_0} \tag{A.77}$$
$$\mathbb{A}_{2U,2U+2} = -\xi_{2,1}e^{\xi_{2,1}y_0} \tag{A.78}$$

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

$$\mathbf{b}_{2j+2U+1} = (\theta_{1,1} - \theta_{1,-1})e^{(1-\alpha)y_0} + \theta_{2,1} - \theta_{2,-1} \tag{A.79}$$
$$\mathbf{b}_{2j+2U+2} = (\theta_{1,1} - \theta_{1,-1})(1-\alpha)e^{(1-\alpha)y_0} \tag{A.80}$$

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

$$\mathbb{A}_{2j+2U-3,2j+2U-3} = e^{\xi_{1,j} y_{j-1}} \tag{A.81}$$

$$\mathbb{A}_{2j+2U-3,2j+2U-2} = e^{\xi_{2,j} y_{j-1}} \tag{A.82}$$

$$\mathbb{A}_{2j+2U-3,2j+2U-1} = -e^{\xi_{1,j-1} y_{j-1}} \tag{A.83}$$

$$\mathbb{A}_{2j+2U-3,2j+2U} = -e^{\xi_{2,j-1} y_{j-1}} \tag{A.84}$$

$$\mathbb{A}_{2j+2U-2,2j+2U-3} = \xi_{1,j} e^{\xi_{1,j} y_{j-1}} \tag{A.85}$$

$$\mathbb{A}_{2j+2U-2,2j+2U-2} = \xi_{2,j} e^{\xi_{2,j} y_{j-1}} \tag{A.86}$$

$$\mathbb{A}_{2j+2U-2,2j+2U-1} = -\xi_{1,j-1} e^{\xi_{1,j-1} y_{j-1}} \tag{A.87}$$

$$\mathbb{A}_{2j+2U-2,2j+2U} = -\xi_{2,j-1} e^{\xi_{2,j-1} y_{j-1}} \tag{A.88}$$

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

$$\mathbf{b}_{2j+2U-3} = (\theta_{1,j-1} - \theta_{1,j}) e^{(1-\alpha) y_{j-1}} + \theta_{2,j-1} - \theta_{2,j} \tag{A.89}$$

$$\mathbf{b}_{2j+2U-2} = (\theta_{1,j-1} - \theta_{1,j})(1-\alpha) e^{(1-\alpha) y_{j-1}} \tag{A.90}$$

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

D Proofs

Proof. (of Proposition 2). Recall that the density of gaps on a segment j is

$$\tilde{f}_j(x) = \eta_{1,j} e^{\xi_{1,j} x} + \eta_{2,j} e^{\xi_{2,j} x} \tag{A.91}$$

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

$$\{\xi_{1,-1}, \xi_{2,-1}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2 \lambda_u}}{\sigma^2} \tag{A.92}$$

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

$$\{\xi_{1,1}, \xi_{2,1}\} = \frac{-(\mu+\delta) \pm \sqrt{(\mu+\delta)^2 + 2\sigma^2 \lambda_d}}{\sigma^2} \tag{A.93}$$

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:

$$\eta \left(\frac{1}{\xi_{1,-1}} - \frac{1}{\xi_{2,1}} \right) = 1 \tag{A.94}$$

Plugging,

$$\eta = \frac{\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d} + (\mu+\delta)\right)\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} - (\mu+\delta)\right)}{\sigma^2\left(\sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_u} + \sqrt{(\mu+\delta)^2 + 2\sigma^2\lambda_d}\right)} \tag{A.95}$$

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

$$N = \eta \left(\frac{\lambda_u}{\xi_{1,-1}} - \frac{\lambda_d}{\xi_{2,1}} \right) \tag{A.96}$$

Plugging η,

$$N = \frac{\lambda_u \xi_{2,1} - \lambda_d \xi_{1,-1}}{\xi_{2,1} - \xi_{1,-1}} \tag{A.97}$$

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

$$\frac{\sigma^2}{2}\xi_{2,1}^2 + (\mu+\delta)\xi_{2,1} = \lambda_d \tag{A.98}$$

$$\frac{\sigma^2}{2}\xi_{1,-1}^2 + (\mu+\delta)\xi_{1,-1} = \lambda_u \tag{A.99}$$

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. □

Proof. (of Proposition 3). Start with the accounting identity

$$Nq(-x) = \Lambda(x)f(x) \tag{A.100}$$

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

$$\frac{H_j}{\lambda_j} = \frac{1}{N}\int_j f(x)dx \tag{A.101}$$

Now summing over all these segments,

$$\sum_j \frac{H_j}{\lambda_j} = \frac{1}{N} \tag{A.102}$$

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. □

Proof. (of Proposition 4). The impulse response function is given by

$$X(t;\mathcal{P}) = \int_{-\infty}^{\infty}(f(x,t) - f(x))x dx \tag{A.103}$$

First, note that the steady-state distributions $f(x)$ are the same under $\mathcal{P} = (\mu+\delta, \sigma^2, \boldsymbol{\lambda})$ and $\alpha\mathcal{P} = (\alpha(\mu+\delta), \alpha\sigma^2, \alpha\,\boldsymbol{\lambda})$ since $f(x)$ that satisfies

$$\Lambda(x)f(x) = (\mu+\delta)f'(x) + \frac{\sigma^2}{2}f''(x) \tag{A.104}$$

also satisfies

$$\alpha\Lambda(x)f(x) = \alpha(\mu+\delta)f'(x) + \frac{\alpha\sigma^2}{2}f''(x) \tag{A.105}$$

It is enough to show that $f(x, \alpha^{-1}t; \alpha\mathcal{P}) = f(x, t; \mathcal{P})$ for all x and $t > 0$.

To that end, write down the Kolmogorov forward equation for $f(x,t;\alpha\mathcal{P})$:

$$\begin{aligned}\partial_t f(x,t;\alpha\mathcal{P}) &= \alpha(\mu+\delta)\partial_x f(x,t;\alpha\mathcal{P}) + \frac{\alpha\sigma^2}{2}\partial_{xx} f(x,t;\alpha\mathcal{P}) - \alpha\Lambda(x) f(x,t;\alpha\mathcal{P}) \\ &= \alpha\partial_t f(x,t;\mathcal{P}) \end{aligned} \tag{A.106}$$

Since the initial conditions coincide, $f(x,0;\mathcal{P}) = f(x,0;\alpha\mathcal{P})$, we can integrate this to get

$$f(x,\alpha^{-1}t;\alpha\mathcal{P}) = \int_0^{\alpha^{-1}t} \alpha\partial_s f(x,s;\mathcal{P})ds = \int_0^t \partial_\tau f(x,\tau;\mathcal{P})d\tau = f(x,t;\mathcal{P}) \tag{A.107}$$

Hence, it holds that $X(\alpha^{-1}t;\alpha\mathcal{P}) = X(t,\mathcal{P})$, and $C(\mathcal{P}) = \alpha C(\alpha\mathcal{P})$ trivially follows. □
Proof. (of Proposition 5). Recall the definition of the impulse response function $X(t)$:

$$X(t) = \int_{-\infty}^{\infty} (f(x,t) - f(x))xdx \tag{A.108}$$

The time derivative is

$$X'(t) = \int_{-\infty}^{\infty} \partial_t f(x,t)xdx \tag{A.109}$$

Take the Kolmogorov forward equation for $f(x,t)$:

$$\partial_t f(x,t) + \lambda f(x,t) = (\mu+\delta)\partial_x f(x,t) + \frac{\sigma^2}{2}\partial_{xx} f(x,t) \tag{A.110}$$

Plugging this into equation (A.109),

$$\begin{aligned} X'(t) &= (\mu+\delta)\int_{-\infty}^{\infty}\partial_x f(x,t)xdx + \frac{\sigma^2}{2}\int_{-\infty}^{\infty}\partial_{xx} f(x,t)xdx - \lambda\int_{-\infty}^{\infty} f(x,t)xdx \\ &= (\mu+\delta)\left[f(x,t)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x,t)dx\right] + \frac{\sigma^2}{2}\left[\partial_x f(x,t)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\partial_x f(x,t)dx\right] \\ &- \lambda\int_{-\infty}^{\infty} f(x,t)xdx = -(\mu+\delta) + \frac{\sigma^2 x^*}{2}(\partial_x^- f(x^*,0) - \partial_x^+ f(x^*,0)) - \lambda\int_{-\infty}^{\infty} f(x,t)xdx \end{aligned} \tag{A.111}$$

Here $\partial_x^- f(x^*,0)$ and $\partial_x^+ f(x^*,0)$ are the left and right derivatives of $f(x,t)$ with respect to x. Now take the Kolmogorov forward equation (A.110) and integrate it over the real line:

$$\int_{-\infty}^{\infty}\partial_t f(x,t)dx + \int_{-\infty}^{\infty}\lambda f(x,t)dx = \frac{\sigma^2}{2}(\partial_x^- f(x^*,0) - \partial_x^+ f(x^*,0)) \tag{A.112}$$

The left-hand side here is equal to N, since the first integral is zero, and the second one integrates to $\lambda = N$. Plugging this back into the expression for $\partial_t X(t)$,

$$X'(t) = -(\mu+\delta) + x^* N - \lambda \int_{-\infty}^{\infty} f(x,t) x dx \tag{A.113}$$

Now take the Kolmogorov forward equation for the steady-state distribution and integrate it over the real line:

$$\int_{-\infty}^{\infty} \lambda f(x) dx = \frac{\sigma^2}{2}(f'_-(x^*) - f'_+(x^*)) \tag{A.114}$$

Now multiply the same equation by x and integrate over the real line:

$$\begin{aligned}\lambda \int_{-\infty}^{\infty} f(x) x dx &= (\mu+\delta)\left[f(x)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)dx \right] + \frac{\sigma^2}{2}\left[f'(x)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)dx \right] \\ &= -(\mu+\delta) + \frac{\sigma^2 x^*}{2}(f'_-(x^*) - f'_+(x^*)) = -(\mu+\delta) + x^* \int_{-\infty}^{\infty} \lambda f(x) dx \\ &= -(\mu+\delta) + x^* N \end{aligned} \tag{A.115}$$

Integrating Plugging this into equation (A.120),

$$X'(t) = \lambda \int_{-\infty}^{\infty} (f(x) - f(x,t)) x dx = -\lambda X(t) \tag{A.116}$$

With the initial condition $X(0;\mathcal{P}) = -\varepsilon$, we get $X(t;\mathcal{P}) = -\varepsilon e^{-\lambda t}$ and $\mathcal{C}(\mathcal{P}) = -\varepsilon/\lambda$. □
Proof. (of Proposition 6). Recall the definition of the impulse response function $X(t)$:

$$X(t) = \int_{-\infty}^{\infty} (f(x,t) - f(x)) x dx \tag{A.117}$$

The time derivative is

$$X'(t) = \int_{-\infty}^{\infty} \partial_t f(x,t) x dx \tag{A.118}$$

Take the Kolmogorov forward equation for $f(x,t)$:

$$\partial_t f(x,t) + \lambda f(x,t) = (\mu+\delta)\partial_x f(x,t) + \frac{\sigma^2}{2}\partial_{xx} f(x,t) \tag{A.119}$$

Plugging it,

$$
\begin{aligned}
X'(t) &= (\mu+\delta)\int_{-\infty}^{\infty}\partial_x f(x,t)xdx + \frac{\sigma^2}{2}\int_{-\infty}^{\infty}\partial_{xx}f(x,t)xdx - \int_{-\infty}^{\infty}\Lambda(x)f(x,t)xdx \quad \text{(A.120)}\\
&= (\mu+\delta)\left[f(x,t)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x,t)dx\right] + \frac{\sigma^2}{2}\left[\partial_x f(x,t)x\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\partial_x f(x,t)dx\right]\\
&- \int_{-\infty}^{\infty}\Lambda(x)f(x,t)xdx\\
&= -(\mu+\delta) - \int_{-\infty}^{\infty}\Lambda(x)f(x,t)xdx + \frac{\sigma^2 x^*}{2}(\partial_x^- f(x^*,0) - \partial_x^+ f(x^*,0))
\end{aligned}
$$

Integrating the Kolmogorov forward equation itself over the real line,

$$\int_{-\infty}^{\infty}\Lambda(x)f(x,0)dx = \frac{\sigma^2}{2}(\partial_x^- f(x^*,0) - \partial_x^+ f(x^*,0)) \tag{A.121}$$

Hence,

$$X'(t) = -(\mu+\delta) - \int_{-\infty}^{\infty}\Lambda(x)f(x,t)(x-x^*)dx \tag{A.122}$$

Integrating the Kolmogorov forward equation for the steady-state distribution,

$$\int_{-\infty}^{\infty}\Lambda(x)f(x)dx = \frac{\sigma^2}{2}(f'_-(x^*) - f'_+(x^*)) \tag{A.123}$$

Integrating the same equation multiplied by x,

$$\int_{-\infty}^{\infty}\Lambda(x)f(x)xdx = -(\mu+\delta) + x^*\frac{\sigma^2}{2}(f'_-(x^*) - f'_+(x^*)) \tag{A.124}$$

Taken together, these two expressions imply

$$\int_{-\infty}^{\infty}\Lambda(x)f(x)(x-x^*)dx = -(\mu+\delta) \tag{A.125}$$

Plugging this into the expression for $X'(t)$,

$$X'(t) = \int_{-\infty}^{\infty}\Lambda(x)(f(x)-f(x,t))(x-x^*)dx \tag{A.126}$$

Now, evaluating this at $t = 0$ and plugging the piece-wise constant $\Lambda(\cdot)$,

$$
\begin{aligned}
X'(0) &= \sum_j \lambda_j \int_{\Delta_j} (f(x) - f(x+\varepsilon))(x - x^*)dx \qquad \text{(A.127)} \\
&= \sum_j \lambda_j \int_{\Delta_j} f(x)(x - x^*)dx - \sum_j \lambda_j \int_{\Delta_j} f(x+\varepsilon)(x - x^*)dx \\
&= \sum_j \lambda_j \int_{l_j}^{u_j} f(x)(x - x^*)dx - \sum_j \lambda_j \int_{l_j+\varepsilon}^{u_j+\varepsilon} f(y)(y - \varepsilon - x^*)dy \\
&= \sum_j \lambda_j \left(\int_{l_j}^{u_j} f(x)(x - \varepsilon - x^*)dx - \int_{l_j+\varepsilon}^{u_j+\varepsilon} f(x)(x - \varepsilon - x^*)dx \right) \\
&+ \varepsilon \sum_j \lambda_j \int_{l_j}^{u_j} f(x)dx \\
&= \sum_j (\lambda_{j+1} - \lambda_j) \int_{u_j}^{u_j+\varepsilon} f(x)(x - \varepsilon - x^*)dx + \varepsilon \int_{-\infty}^{\infty} \Lambda(x) f(x)dx
\end{aligned}
$$

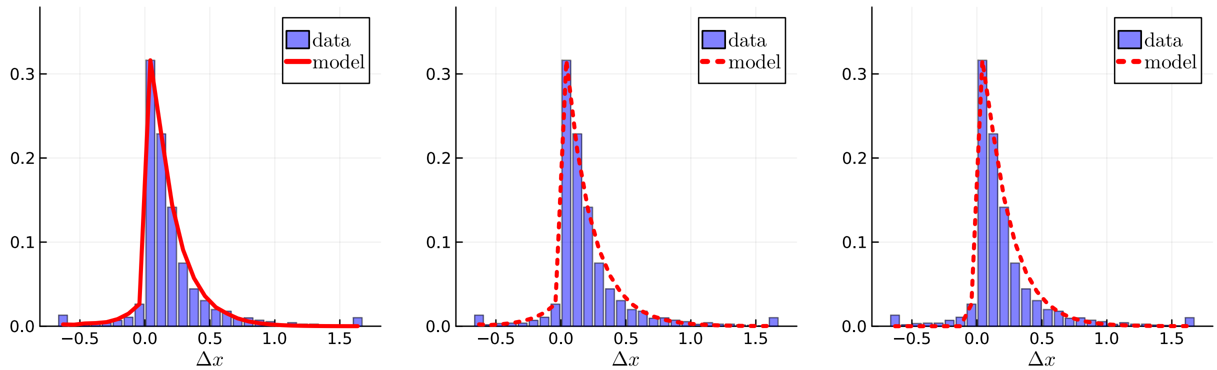
Here Δ_j is the segment on which $\Lambda(x) = \lambda_j$, and (l_j, u_j) are its left and right boundaries. The last term is equal to εN, since $\Lambda(x)f(x) = Nq(-x)$, and $q(\cdot)$ is a density that integrates to one. This proves the proposition. □

Proof. (of Corollary 2). The corollary follows from differentiating $X'(0; \mathcal{P})$ with respect to ε and evaluating the derivative at $\varepsilon = 0$.

Proof. (of Corollary 3). The corollary follows from setting all $\lambda_j - \lambda_{j+1} = 0$ for $j \geq 1$ and $\lambda_{j-1} - \lambda_j = 0$ for $j \leq -1$ to evaluate the slope for the two-sided distribution. In the Calvo model, all λ_j are equal to each other. The other summand is the same across the two models by assumption.

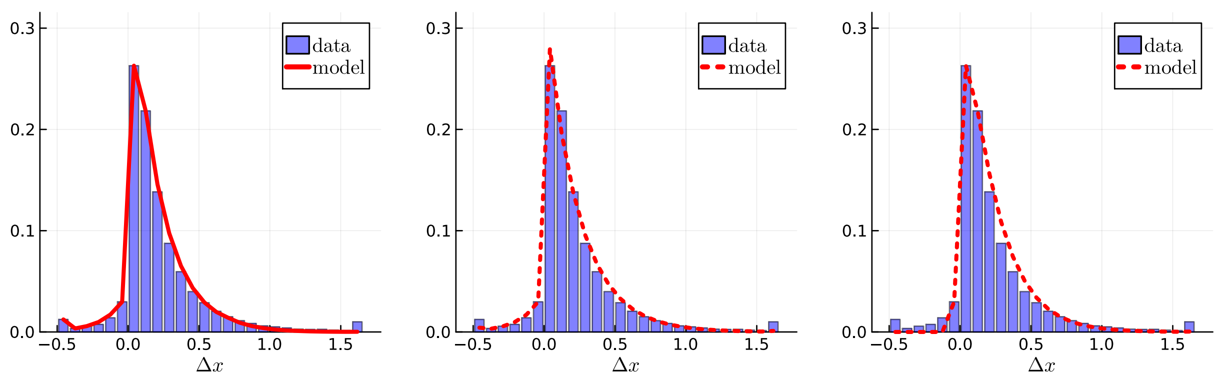
E Histograms and recovered hazards

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.



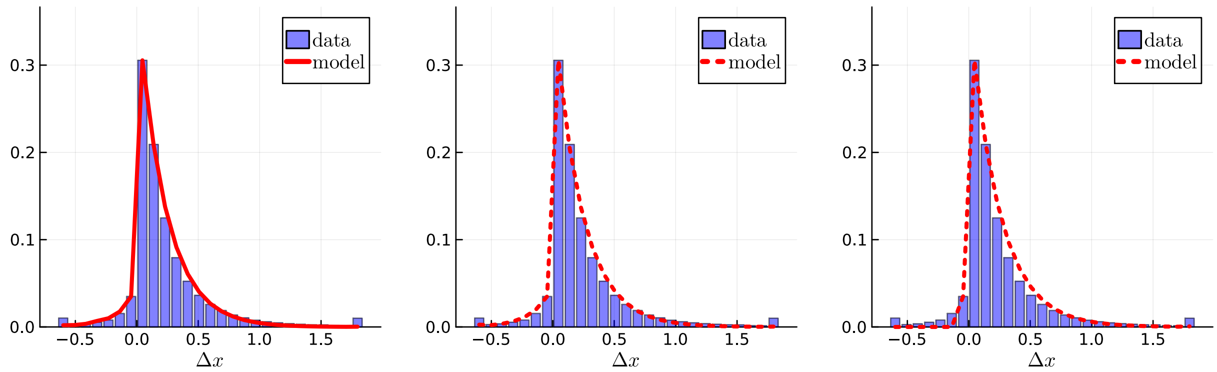
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.1: Mining & Quarrying



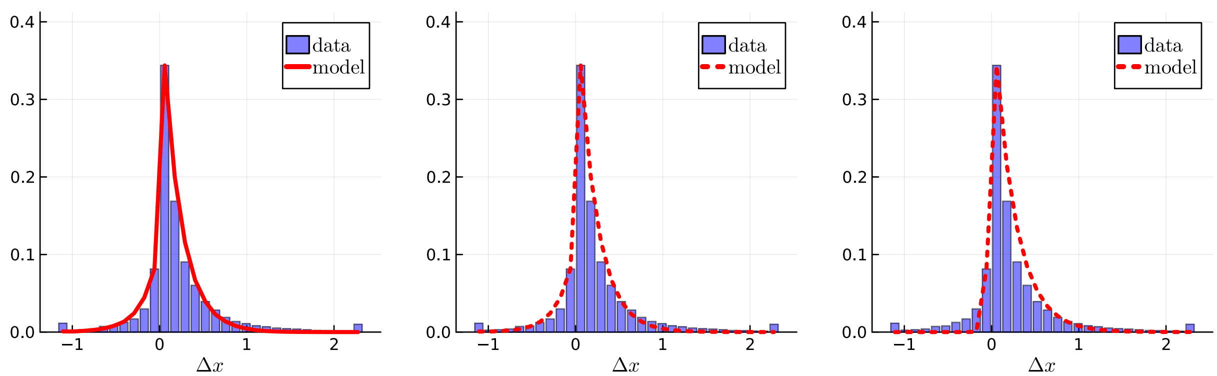
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.2: Chemicals



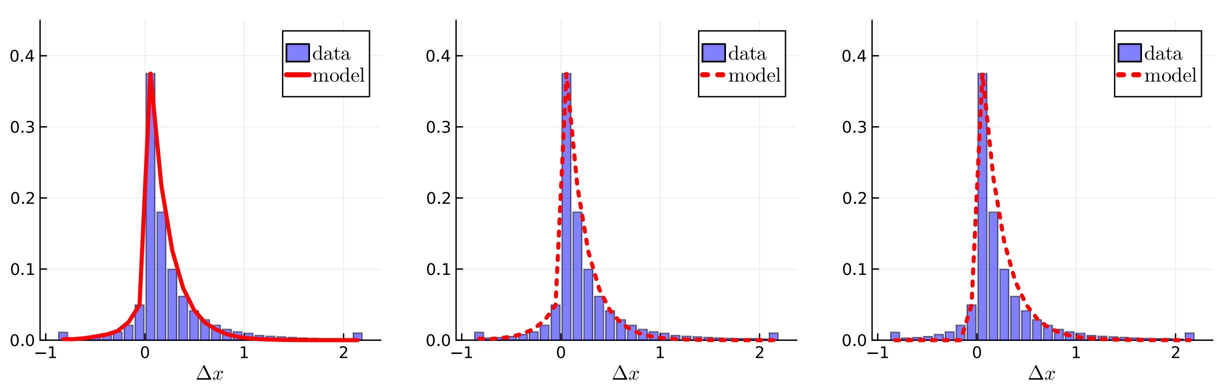
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.3: Food & Beverages



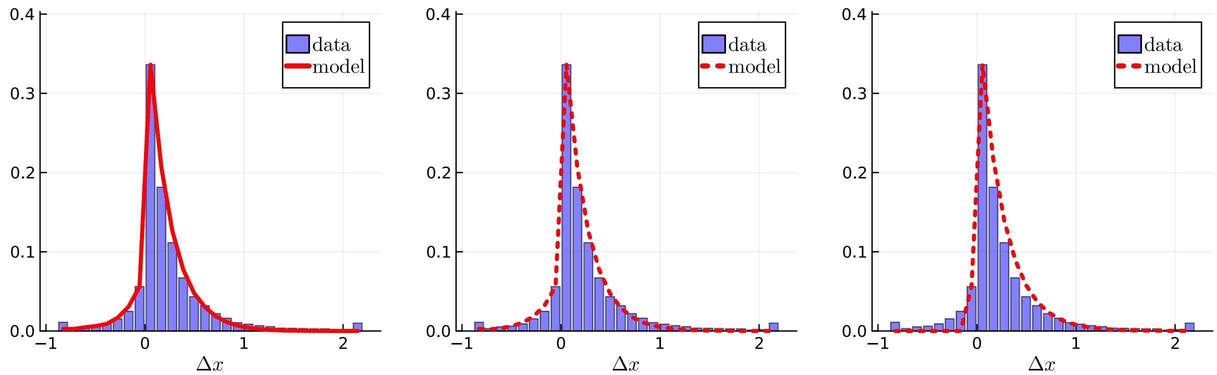
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.4: Construction



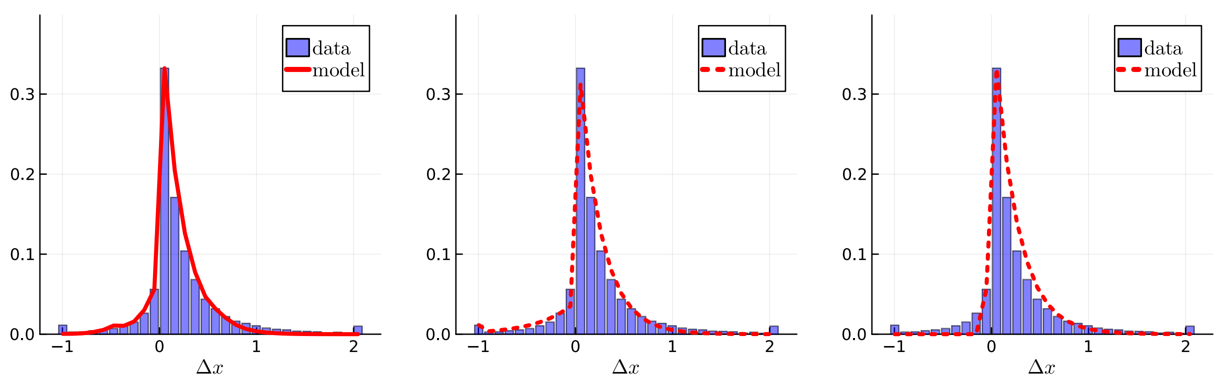
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.5: Retail



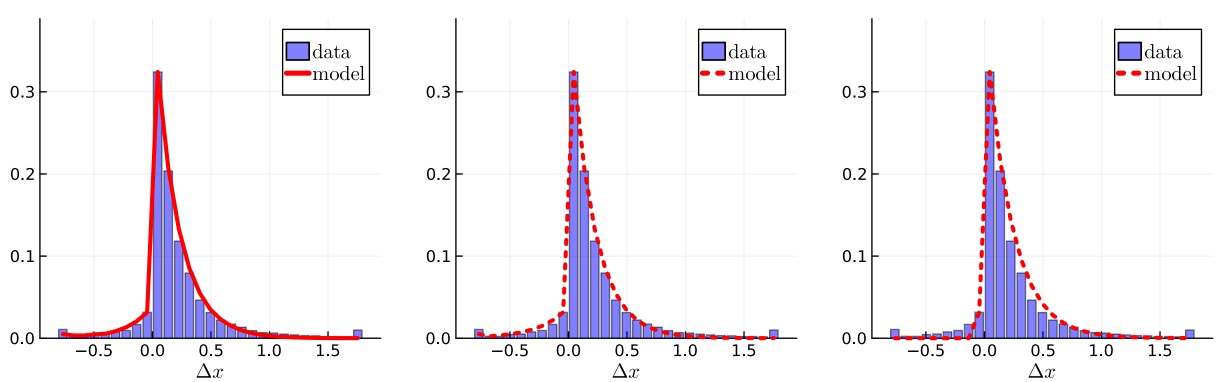
(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

Figure A.6: Transportation



(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

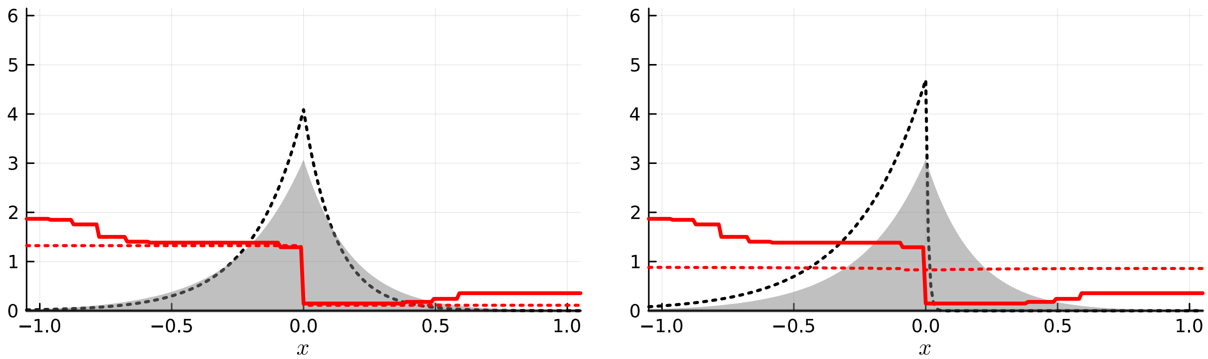
Figure A.7: Insurance



(a) Data and the full model. (b) The two-sided benchmark. (c) The symmetric benchmark.

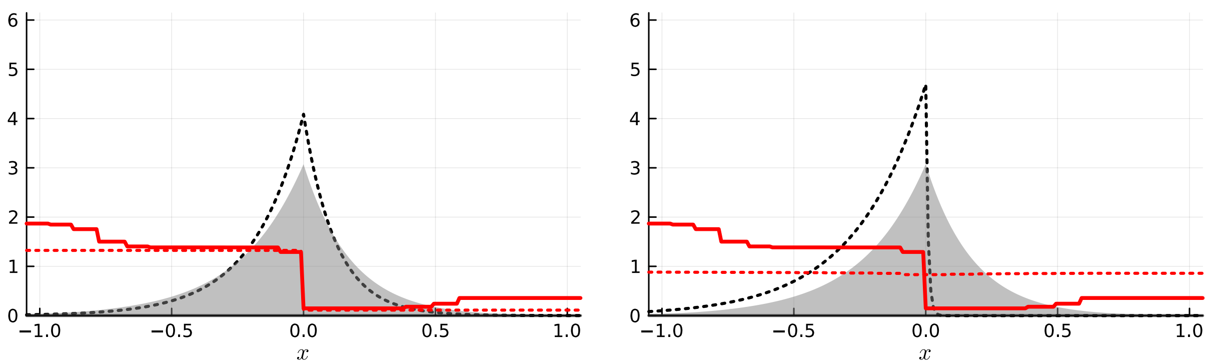
Figure A.8: Health & Beauty

The figures below show the analogs of Figure 3 for sectors other than "Metals & Machinery". They are organized in the same way: the left panel shows the generalized hazard function and the implied steady-state distribution of capital gaps for the full model in solid and for the two-sided benchmark in dash. The right panel does the same for the symmetric benchmark instead of the two-sided.



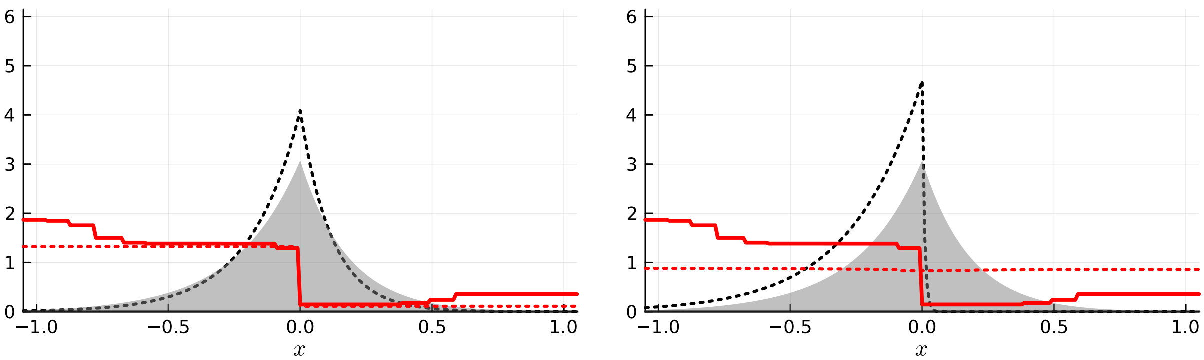
(a) Full model and two-sided benchmark. (b) Full model and symmetric benchmark.

Figure A.9: Mining & Quarrying



(a) Full model and two-sided benchmark. (b) Full model and symmetric benchmark.

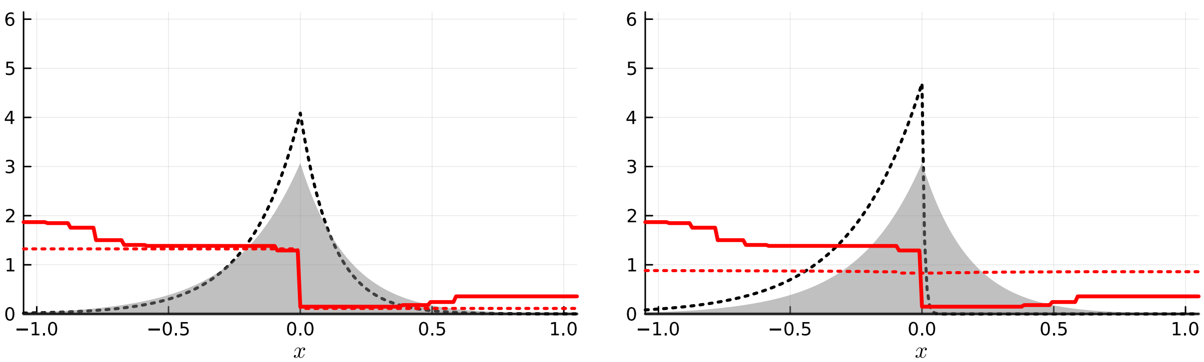
Figure A.10: Chemicals



(a) Full model and two-sided benchmark.

(b) Full model and symmetric benchmark.

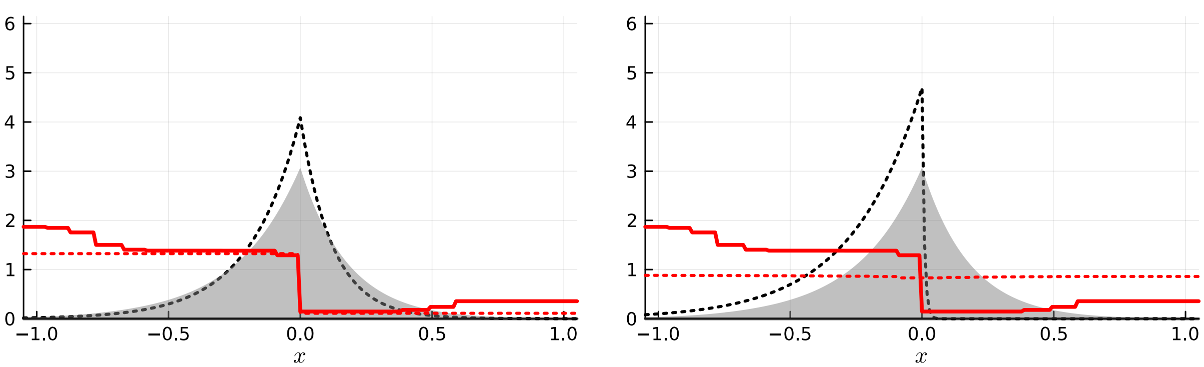
Figure A.11: Food & Beverages



(a) Full model and two-sided benchmark.

(b) Full model and symmetric benchmark.

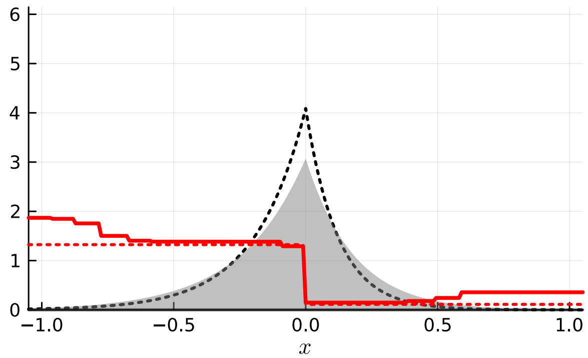
Figure A.12: Construction



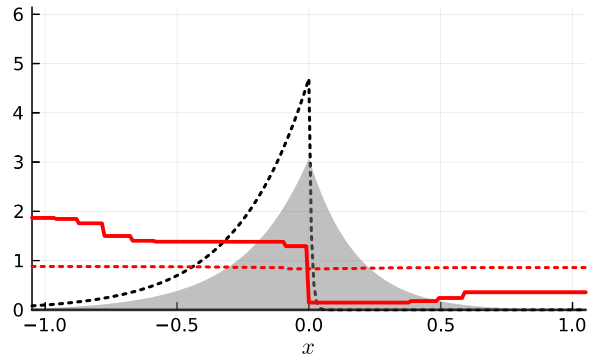
(a) Full model and two-sided benchmark.

(b) Full model and symmetric benchmark.

Figure A.13: Retail

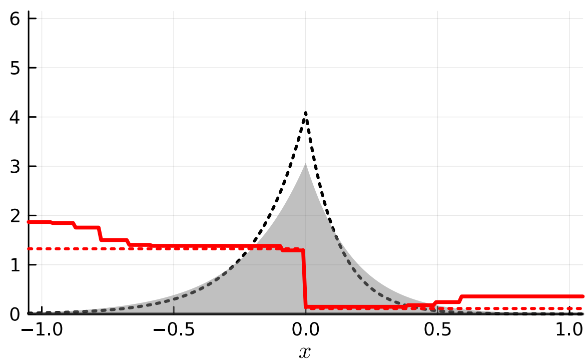


(a) Full model and two-sided benchmark.

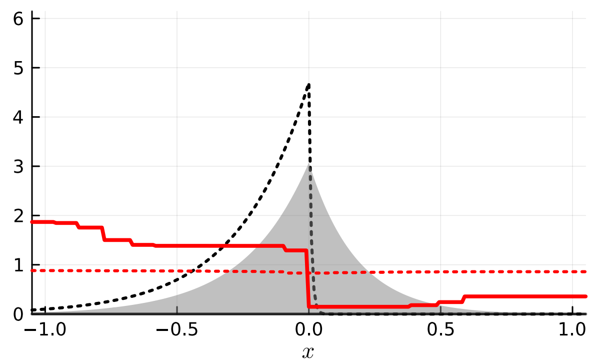


(b) Full model and symmetric benchmark.

Figure A.14: Transportation

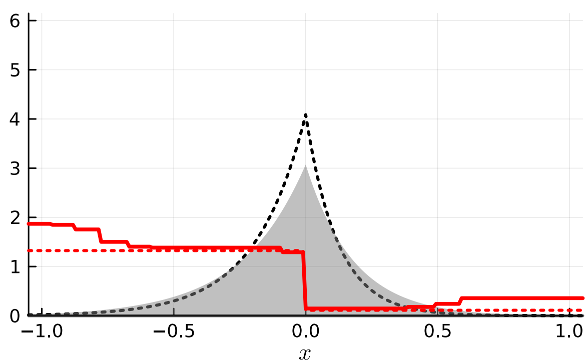


(a) Full model and two-sided benchmark.

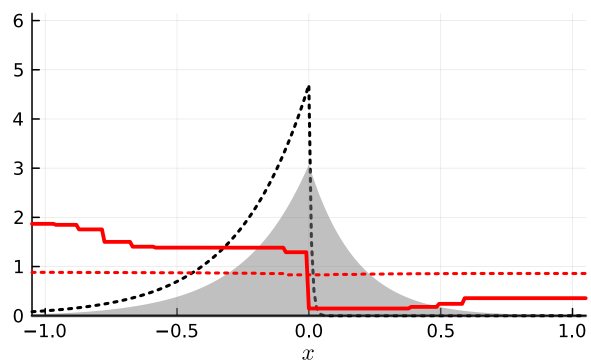


(b) Full model and symmetric benchmark.

Figure A.15: Insurance



(a) Full model and two-sided benchmark.



(b) Full model and symmetric benchmark.

Figure A.16: 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 4. The left panel shows cumulative distribution functions G_u (in purple) and G_d (in green) 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 $(k^*)^\alpha - \nu k^*$.

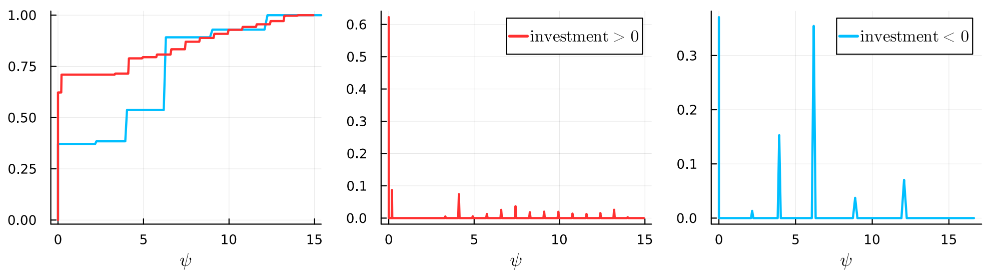


Figure A.17: Mining & Quarrying

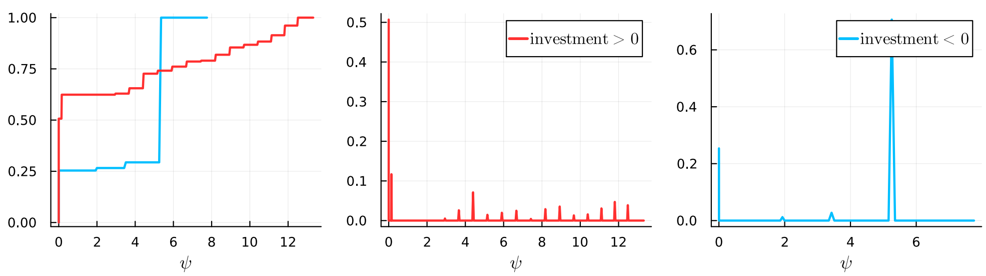


Figure A.18: Chemicals

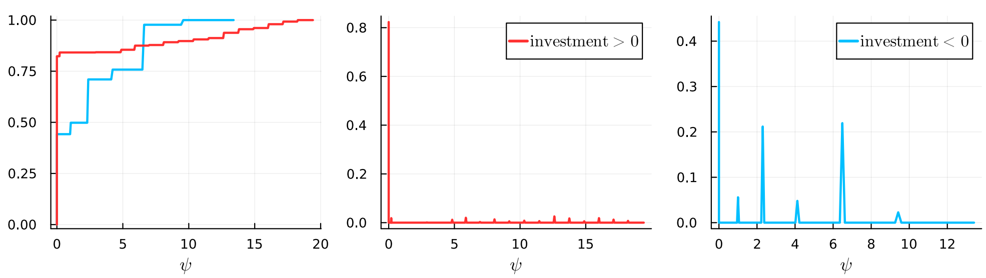


Figure A.19: Food & Beverages

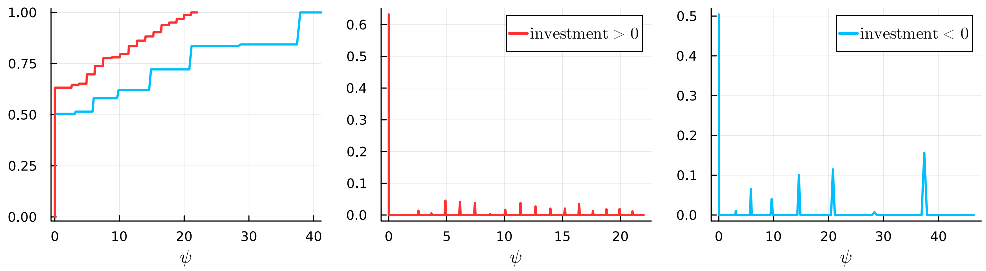


Figure A.20: Construction

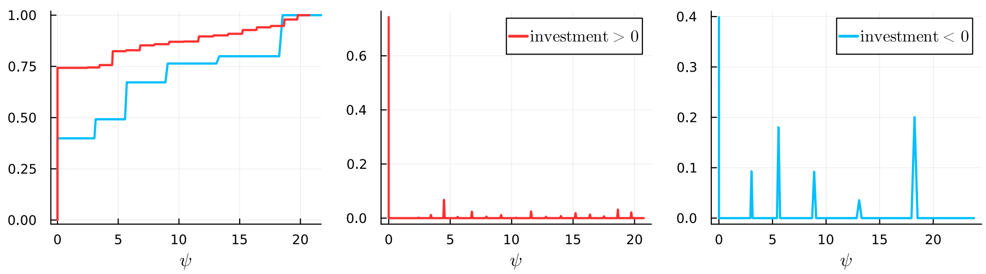


Figure A.21: Retail

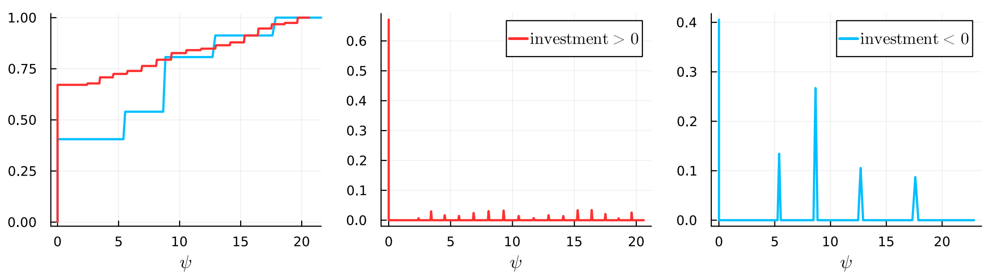


Figure A.22: Transportation

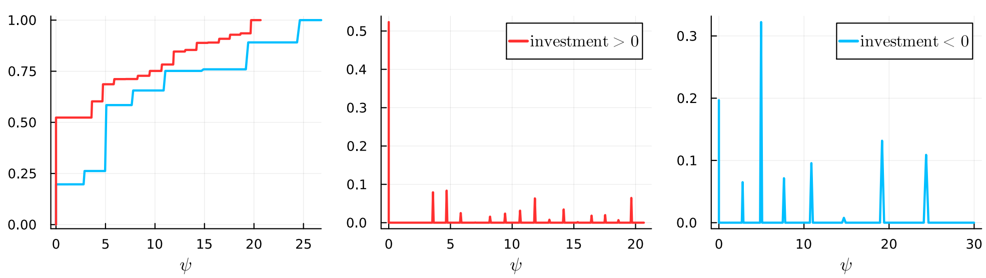


Figure A.23: Insurance

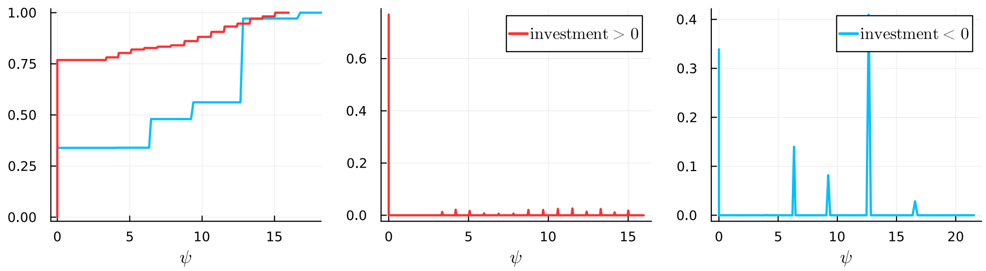
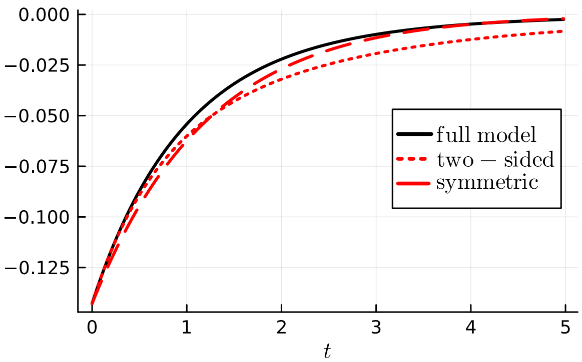


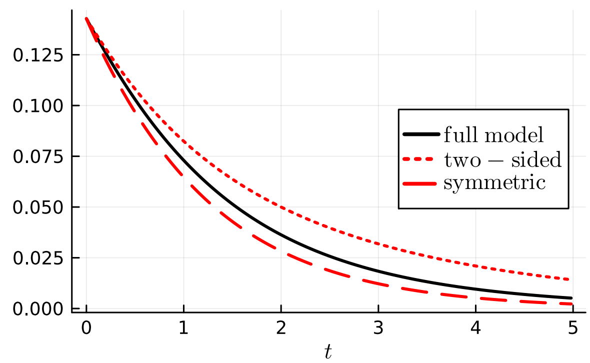
Figure A.24: Health & Beauty

G Impulse responses across models

This section presents impulse responses in the full model, the two-sided model, and the regular Calvo model for all sectors other than "Construction". The graphs are organized in the same way as those for "Construction" on Figure 7. The left panel shows impulse responses of the average capital gap x to a 5% productivity shock. The right panel shows the same for a -5% shock.

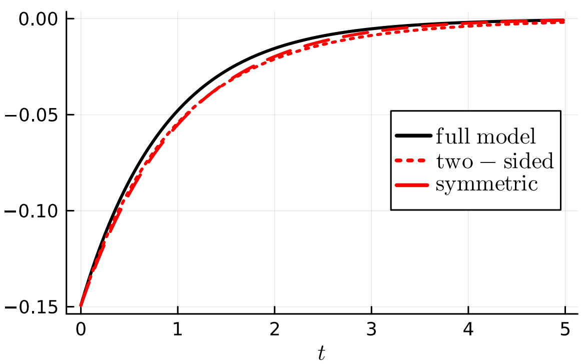


(a) Positive shock to productivity.

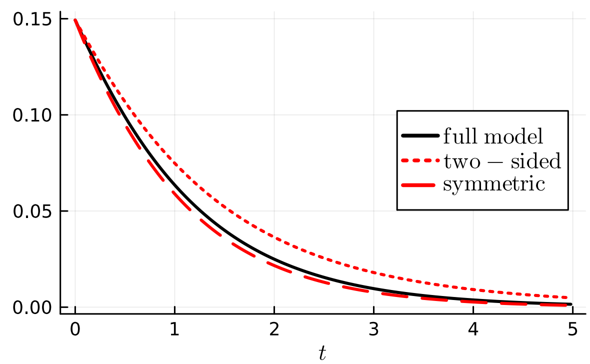


(b) Negative shock to productivity.

Figure A.25: Mining & Quarrying

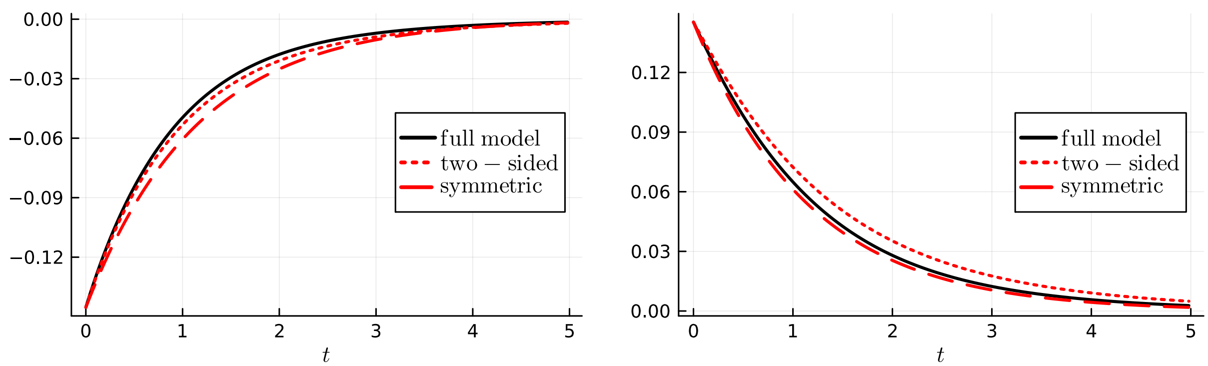


(a) Positive shock to productivity.



(b) Negative shock to productivity.

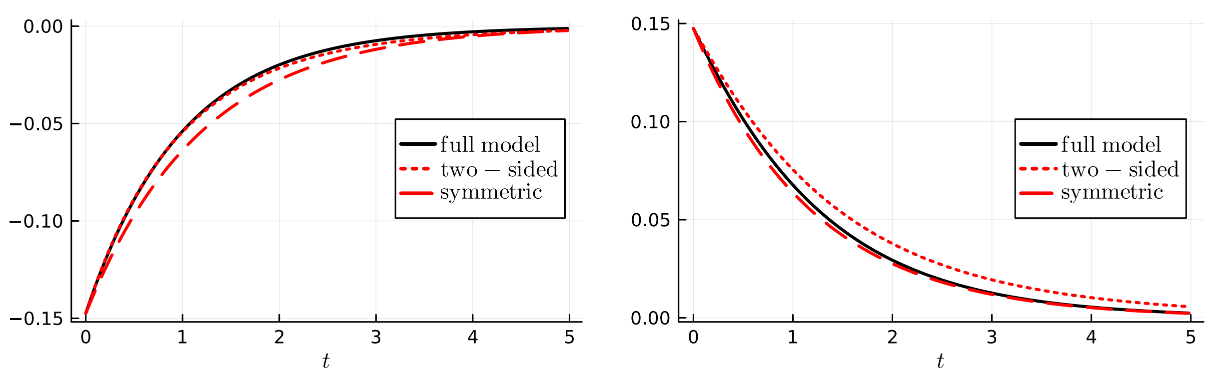
Figure A.26: Chemicals



(a) Positive shock to productivity.

(b) Negative shock to productivity.

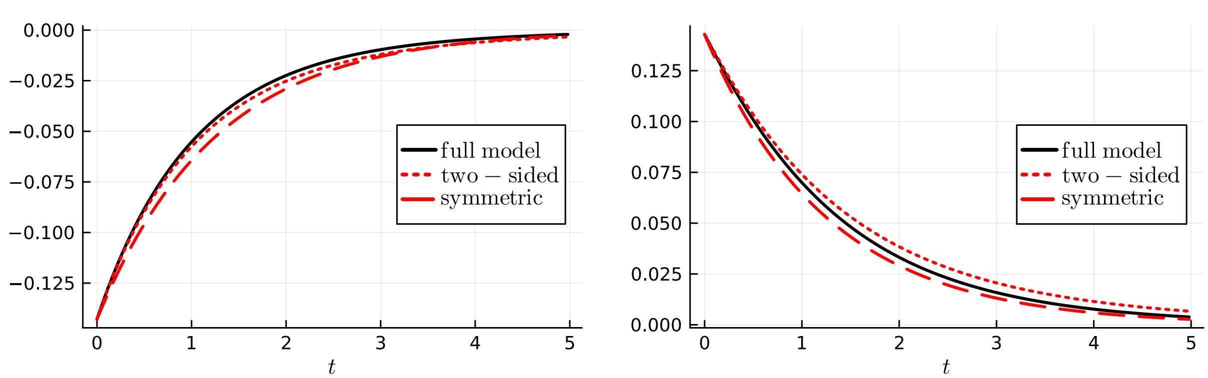
Figure A.27: Metal & Machinery



(a) Positive shock to productivity.

(b) Negative shock to productivity.

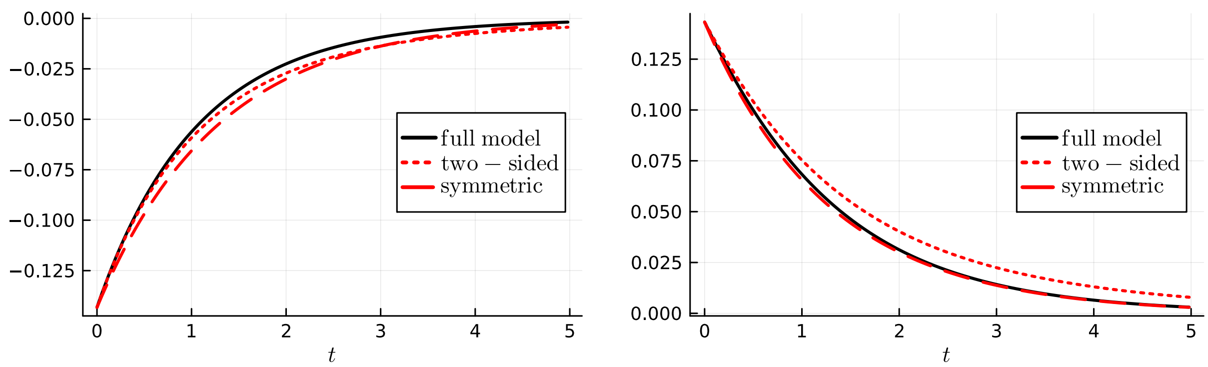
Figure A.28: Food & Beverages



(a) Positive shock to productivity.

(b) Negative shock to productivity.

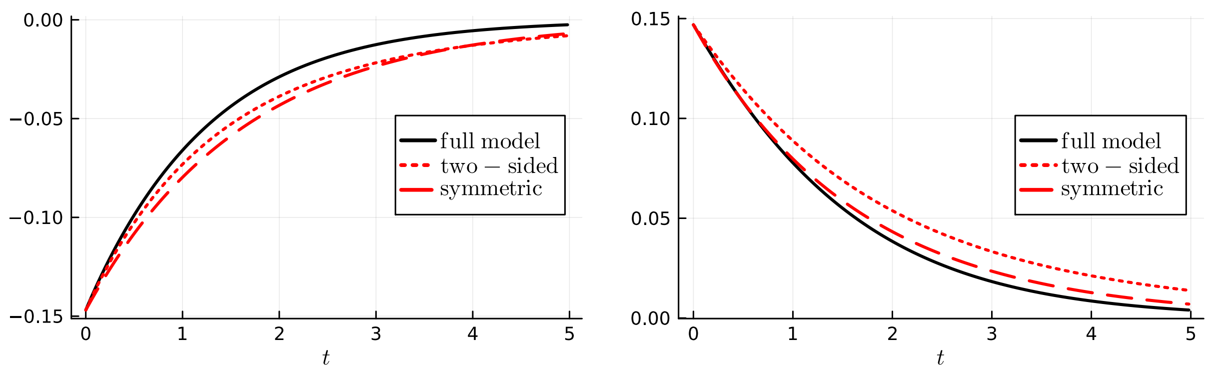
Figure A.29: Retail



(a) Positive shock to productivity.

(b) Negative shock to productivity.

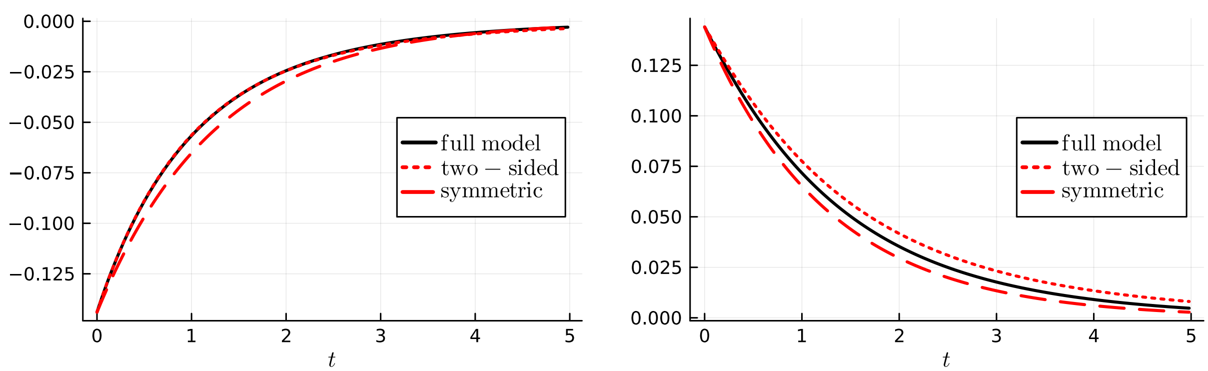
Figure A.30: Transportation



(a) Positive shock to productivity.

(b) Negative shock to productivity.

Figure A.31: Insurance



(a) Positive shock to productivity.

(b) Negative shock to productivity.

Figure A.32: Health & Beauty

H Data Appendix

The empirical analysis uses panel data on Italian firms obtained from the Company Accounts Data Service (*Centrale dei Bilanci*, CB). The dataset covers a period of 25 years, from 1982 to 2006, and contains information about industry and region in which the firm operates, its assets, depreciation, investment in tangible and non-tangible assets as well as the disinvestments. This data have been used in Guiso and Schivardi (2007) and Guiso et al. (2005).

On average there are about 45,000 firms included in the data set in a particular year. We calculate the net investment, I, by subtracting the disinvestment from the investment in tangible assets. In the baseline results, we normalize the investment by the stock of the illiquid assets, used as proxy for the firm's capital.[7] in the main text we will refer to the illiquid assets simply as "assets", A. We follow Cooper and Haltiwanger (2006), as well as Baley and Blanco (2021), and consider as "zero investment" all investments with an absolute size smaller than 1% of the assets.

We filter the data to remove outliers and observations with partial information.[8] We drop observations for firms with missing industry or region. We drop observations for agricultural firms.[9] Additionally, we drop firms for which we observe values of net investment to assets ratio in top or bottom 1% of the distribution and firms for which we have only one observation. This leaves us with 676,033 observations for 78,664 firms, i.e. on average 8.6 observations per firm.[10]

H.1 The size distribution of investment

We present the distribution of non-zero investments by industry in Figure 1. In accord to the theory the investment figures are normalized by our measure of capital, so that the size of investment is measured by $I/(A-I)$ (observations with the investment to assets ratio less than -1, about 0.85% of the total, are not shown in the graph). We observe an asymmetric distribution where positive investments more common than negative ones. Note that the modes of the distribution, for both positive and negative investment, are close to zero.

H.2 Summary Statistics

We present the baseline summary statistics of investment by industry in Table 1. We observe that on average about 18.7% of observations correspond to zero investment. There is a notable degree of heterogeneity in the degree of lumpiness, e.g. the construction almost doubles the share of zero investment observation in the chemical industry. On average, when

[7] In the raw balance-sheet data, investment in tangible assets corresponds to V168, disinvestment to V169, illiquid assets to V010, total assets to V023, depreciation to V121.

[8] Before filtering the data include 1,131,629 observations for 135,356 firms.

[9] For the duration analysis of the timing between non-zero investment of the firms, we also drop observations for firms with any gaps in the data.

[10] See the robustness section for an analysis of the importance of the trimming choices on the number of observations.

investing, the firm invests 15.2% of its assets. It is also notable that 8% of observations are characterized by the negative net investment.

We also present the results concerning robustness of the summary statistics. Table 1 ignores the fact that for some of the firms we cannot calculate the duration statistics, i.e. when the firm never invests or invests just once. In Table 6, we present the descriptive statistics that excludes such firms. The results are quantitatively similar to the ones obtained in the baseline Table 1. We also present the summary statistics that uses total assets instead ofilliquid assets both for the whole sample in Table 7 and for the restricted sample in Table 8.

H.3 Durations

We present the summary statistics by industry concentrating on the summary statistics that are calculated after we compute durations. The duration is measured in years and computed as a time between two consecutive non-zero investments. For example, if a firm invests every single year, the duration for that firm always equals 1. If a firm invested in 1990 and then did not invest until 2000, the duration for such firm is 10. Effectively, the number of durations for each firm equals a number of years in which the firm invests minus one. Thus, all firms that invest only once are not included in the summary statistics of duration.

To compute the statistics, we use two different ways to aggregate the observations – *industry-level* and *firm-level* aggregation. To illustrate the difference between these two approaches, let us consider a simple example of industry with 2 firms in it. We would like to calculate the mean duration between consecutive investments. Durations for each firm are presented in the Table below.

Firm	Durations
A	1, 1
B	5, 5, 5, 5, 5, 5

Firm-level approach to aggregation proceeds as follows. First, we calculate mean duration for each firm – 1 and 5 for firms A and B respectively. Then, to obtain $\mathbb{E}\tau$ we calculate the mean for the numbers we've obtained, i.e. $(1+5)/2 = 3$. Thus, all firms have equal weight when we calculate $\mathbb{E}\tau$.

Industry-level aggregation works a bit different. First, we create the vector of durations by concatenating all the durations we observe for the industry, i.e. $(1, 1, 5, 5, 5, 5, 5, 5)$. Then, we calculate mean value for this vector, i.e. $\mathbb{E}\tau = (1\cdot2+5\cdot6)/8 = 4$. Essentially, this approach uses for a firm the weight proportional to the share of observations belonging to this firm out of total number of observations in the industry.

Baseline results in Table 5 rely on the industry-level aggregation, i.e. we weigh the firms based on the number of observations for the firm. We also show the unweighted (firm-level approach) results presented in Table 9.

We also present the histograms of durations for the whole dataset in Figure A.33 and separately for every industry in Figure A.34. We exclude financial industry from both graphs because of the low number of observations We observe the exponential decay in the distribution of durations between two consecutive investments.

Table 5: Summary statistics of duration by industry

Industry	# Firms	$\mathbb{E}\tau$	$SD(\tau)$	$Cov(\tau, I/A)$	Drift
Mining & Quarrying	609	1.039	0.238	-0.007	0.127
Chemicals	4950	1.043	0.246	-0.004	0.157
Metal & Machinery	12224	1.039	0.232	-0.006	0.167
Food & Beverages	14807	1.052	0.272	-0.006	0.154
Construction	5689	1.093	0.401	-0.009	0.101
Retail	21173	1.074	0.330	-0.005	0.133
Transportation	2973	1.072	0.342	-0.007	0.137
Insurance	2857	1.082	0.409	-0.009	0.121
Health & Beauty	1858	1.047	0.260	-0.005	0.136
Total	67227	1.059	0.297	-0.006	0.145

Notes: Duration τ are calculated as the difference in years of two consecutive investments and is measured in years. The firm-level statistics are weighted proportional to the number of observation for each firm. Investment is considered zero if the net investment to assets ratio is less than 1% in absolute value. $\mathbb{E}\tau$ is the sample mean of a pause between two consecutive non-zero investments (τ). $SD(\tau)$ is the sample standard deviation of the pause between two consecutive non-zero investments. $Cov(\tau, I/A)$ is the sample covariance between investment to assets ratio and the pause between two consecutive non-zero investments. Drift is calculated as $\dfrac{\mathbb{E}(I/A|I/A > 0)}{\mathbb{E}(\tau)}$.

H.4 Robustness

Figure A.35 shows the effects of the alternative threshold choices for the trimming of outliers. The benchmark we choose, following the literature, is to trim the top and bottom 1% of the distribution of I/A. Alternative thresholds to the benchmark of 1% are considered on the horizontal axis. The vertical axis shows the fraction of observations that are dropped as the threshold increases.

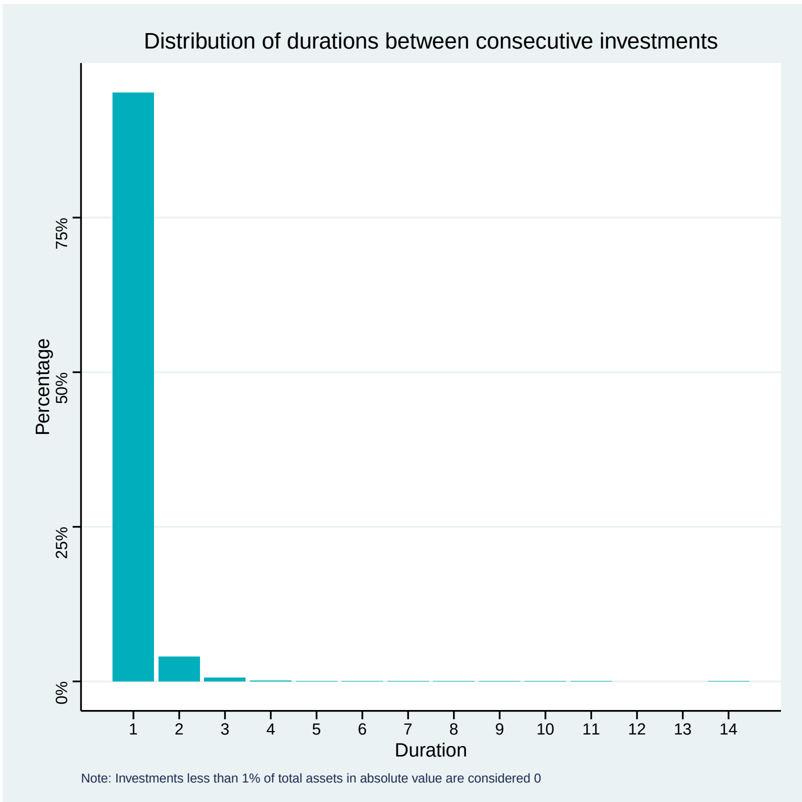


Figure A.33: Histogram of duration between two investments

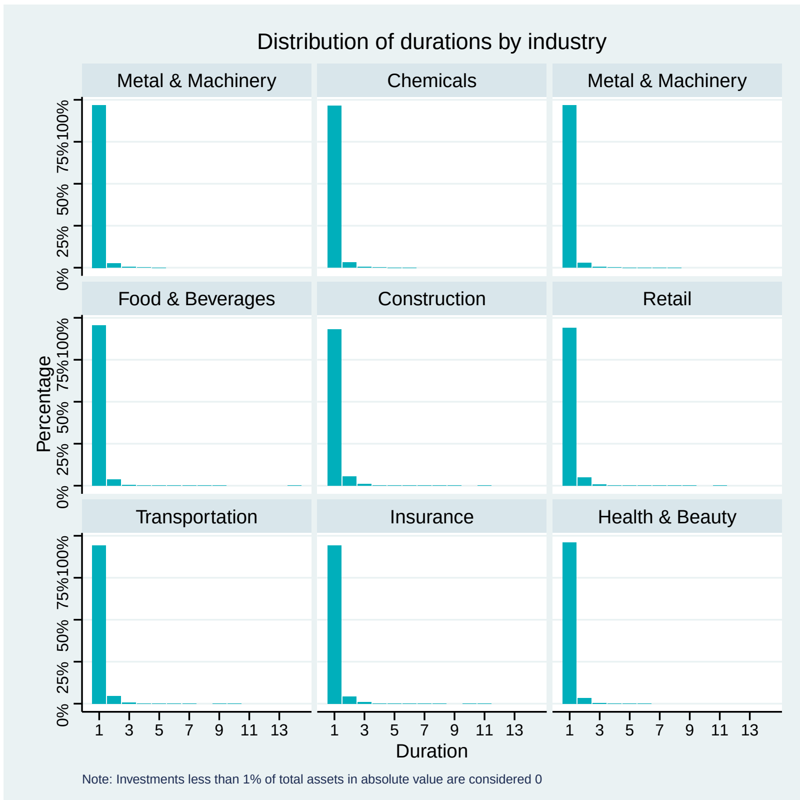


Figure A.34: Histogram of duration between two investments by industry

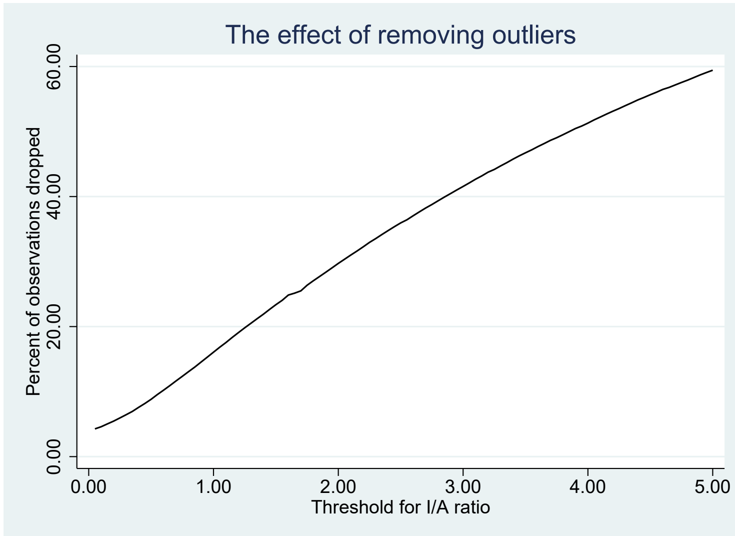


Figure A.35: The effect of alternative thresholds for the "outlier" definition

Note. Percent of observations dropped is calculated for every threshold in the investment to assets ratio with a step of 0.05%. The threshold shows the lower and upper quantiles of the I/A ratio distribution for which the firms with any observations below lower or above upper quantiles are dropped. Note that if a firm is an outlier in one year, all observations for such firm are dropped.

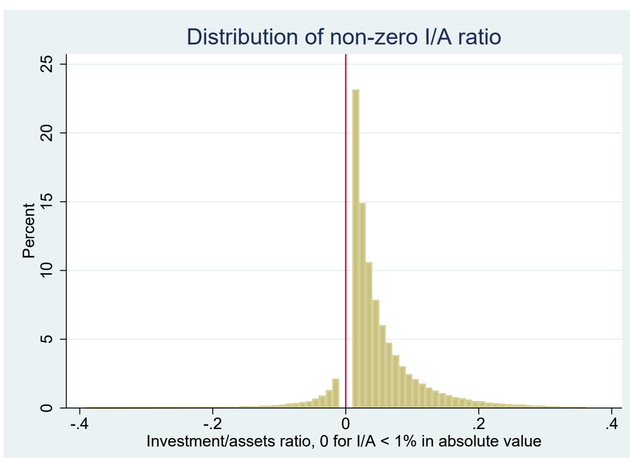


Figure A.36: Robustness: Distribution of non-zero I/A ratios using total assets.

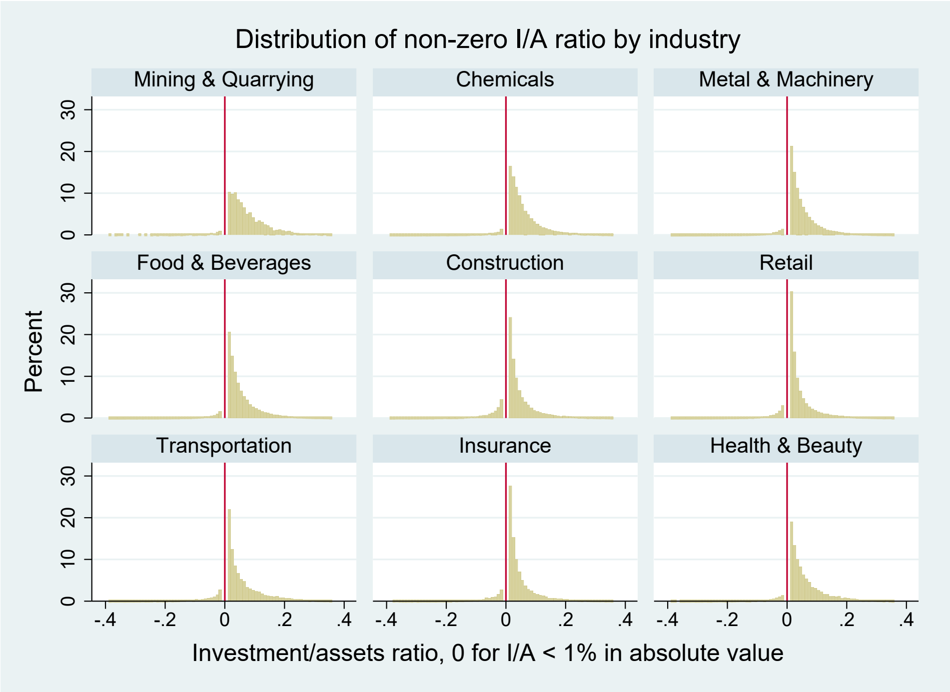


Figure A.37: Robustness: Distribution of non-zero I/A ratios by industry using total assets.

Table 6: Robustness: Summary statistics by industry – restricted durations sample

Industry	# Firms	Share inactive	Mean non-zero I/A	SD non-zero I/A	Share negativ
Mining & Quarrying	609	0.166	0.132	0.347	0.057
Chemicals	4950	0.134	0.164	0.289	0.062
Metal & Machinery	12224	0.134	0.174	0.338	0.064
Food & Beverages	14807	0.144	0.162	0.325	0.068
Construction	5689	0.219	0.110	0.478	0.134
Retail	21173	0.180	0.143	0.404	0.091
Transportation	2973	0.186	0.147	0.393	0.104
Insurance	2857	0.230	0.131	0.450	0.101
Health & Beauty	1858	0.167	0.143	0.347	0.073
Total	67227	0.162	0.153	0.369	0.081

Notes: Only firms that invest at least twice are included as for them the duration can be calculated. Inves is considered zero if the net investment to assets ratio is less than 1% in absolute value. Assets are defi illiquid assets only. *Share inactive* is a share of observations with zero investment. The mean and sta deviation of the net investment to assets ratio is presented for non-zero investments only. *Share negati* share of observations with negative investment.

Table 7: Robustness: Summary statistics by industry – normalizing by total assets

Industry	# Firms	Share inactive	Mean non-zero I/A	SD non-zero I/A	Share negativ
Mining & Quarrying	752	0.385	0.071	0.080	0.035
Chemicals	4990	0.255	0.055	0.069	0.046
Metal & Machinery	13004	0.307	0.049	0.065	0.046
Food & Beverages	15751	0.330	0.051	0.068	0.047
Construction	7776	0.540	0.035	0.074	0.077
Retail	25025	0.529	0.037	0.068	0.052
Transportation	3259	0.437	0.050	0.083	0.067
Insurance	3275	0.534	0.038	0.072	0.054
Health & Beauty	2076	0.365	0.055	0.074	0.048
Total	75931	0.410	0.046	0.069	0.052

Notes: Assets are defined as total assets instead of illiquid assets used in the baseline. Investment is cons zero if the net investment to assets ratio is less than 1% in absolute value. *Share inactive* is a share of observ with zero investment. The mean and standard deviation of the net investment to assets ratio is present non-zero investments only. *Share negative* is a share of observations with negative investment.

Table 8: Robustness: Summary statistics by industry – total assets and restricted sample

Industry	# Firms	Share inactive	Mean non-zero I/A	SD non-zero I/A	Share negativ
Mining & Quarrying	401	0.186	0.072	0.079	0.044
Chemicals	4244	0.228	0.055	0.069	0.047
Metal & Machinery	10616	0.273	0.049	0.065	0.048
Food & Beverages	12770	0.294	0.051	0.068	0.049
Construction	4154	0.415	0.036	0.073	0.094
Retail	16248	0.453	0.037	0.068	0.059
Transportation	2151	0.352	0.051	0.082	0.074
Insurance	1726	0.386	0.039	0.072	0.067
Health & Beauty	1411	0.279	0.056	0.074	0.052
Total	53728	0.343	0.047	0.069	0.056

Notes: Assets are defined as total assets instead of illiquid assets used in the baseline. Only firms that at least twice are included. Investment is considered zero if the net investment to assets ratio is less th in absolute value. *Share inactive* is a share of observations with zero investment. The mean and sta deviation of the net investment to assets ratio is presented for non-zero investments only. *Share negati* share of observations with negative investment.

Table 9: Robustness: Summary statistics of duration by industry – unweighted

Industry	# Firms	$\mathbb{E}\tau$	$SD(\tau)$	$Cov(\tau, I/A)$	Drift
Mining & Quarrying	609	1.078	0.088	-0.015	0.123
Chemicals	4950	1.060	0.119	-0.003	0.154
Metal & Machinery	12224	1.051	0.107	-0.004	0.166
Food & Beverages	14807	1.066	0.138	-0.003	0.152
Construction	5689	1.103	0.191	-0.005	0.100
Retail	21173	1.085	0.173	-0.003	0.132
Transportation	2973	1.077	0.166	-0.004	0.136
Insurance	2857	1.108	0.174	-0.001	0.118
Health & Beauty	1858	1.059	0.109	-0.002	0.135
Total	67227	1.074	0.147	-0.003	0.143

Notes: Duration τ are calculated as the difference in years of two consecutive investments and is measured in years. The firm-level statistics are not weighted. Investment is considered zero if the net investment to assets ratio is less than 1% in absolute value. $\mathbb{E}\tau$ is the sample mean of a pause between two consecutive non-zero investments (τ). $SD(\tau)$ is the sample standard deviation of the pause between two consecutive non-zero investments. $Cov(\tau, I/A)$ is the sample covariance between investment to assets ratio and the pause between two consecutive non-zero investments. Drift is calculated as $\dfrac{\mathbb{E}(I/A|I/A > 0)}{\mathbb{E}(\tau)}$.