Exit dynamics of start-up firms: Structural estimation using indirect inference

Golombek, Rolf, Arvid Raknerud

Exit dynamics of start-up firms: Structural estimation using indirect inference

by

Rolf Golombek[a)] and Arvid Raknerud[b)]

Abstract

We estimate by means of indirect inference a structural economic model where firms' exit and investment decisions are the solution to a discrete-continuous stochastic dynamic programming problem. Our method solves the main difficulty of simulation-based inference in structural discrete-continuous choice models, namely that the simulated trajectories are discontinuous functions of the structural parameters. Estimating the model on all start-up firms in the Norwegian manufacturing sector, we find that if the expected value of continuing production is persistently low relative to the expected value of exit, the firm has a high probability to exit.

JEL classification: C33, C51, C61, C72, D21

Keywords: Indirect inference, auxiliary model, continuous-discrete choice, Markovian decision model, investment, cost of capital adjustment, firm exit

---

a) The Ragnar Frisch Center for Economic Research, University of Oslo. E-mail: rolf.golombek@frisch.uio.no

b) Corresponding author: Statistics Norway, Research Department. E-mail: rak@ssb.no

1 Introduction

Reallocation of resources from old, inefficient firms to new firms with superior technology is often considered to be a dynamo in market economies; through creative destruction the exit of firms is a means to ensure growth and prosperity. New firms have to invest to build up an optimal stock of capital, but new firms are also characterized by a high exit rate. In our data set, which covers firms in Norwegian manufacturing industries over the period 1994–2012, the average share of 1-year old firms that exited during the next 3 years was 17 percent, compared with 7–8 percent for 10-year old firms. For a rational (new) firm, choosing the investment profile over time is interrelated with the decision of whether to exit or continue production. Still, most empirical studies solely examine either exit or investment. This paper derives a theory-based model of exit and investment under uncertainty that is structurally estimated on Norwegian data for start-up firms.

In our dynamic model, the firm's investment decision is determined simultaneously with the decision of whether to exit. This is in contrast to the literature. First, several models of investment under uncertainty rule out the possibility to exit; for example, see Dixit and Pindyck (1994; Chapter 7), Abel and Eberly (1994; 1996), and Bloom, Bond and van Reenen (2007). Second, other contributions consider the value of exit – the "scrap value" – as exogenous, for example, see Olley and Pakes (1996), Levinsohn and Petrin (2003), Dunne et al. (2013), and Ryan (2012). Finally, some studies on exit have no explicit investment decision; for example, see Pakes et al. (2007) and Aguirregabiria and Ho (2012).

While modeling of exit may seem simple – according to standard economic theory negative profitability is the key reason for firms to exit – accounts data indicate that the exit behavior of Norwegian manufacturing firms may be more complicated: for the period 1994–2012, the data reveal that i) 27 percent of firms that exited had positive *profit* (here defined as operating surplus less capital costs) *every year* before they exited; ii) there is no clustering of negative profitability shocks just prior to exit – around 65 percent of the firms that exited had positive profit in the last year prior to exit; and iii) firms may continue production even though they repeatedly experience negative profit; 30 percent of the firm-year observations for the non-exiting firms – one observation for each firm in each year – had negative profit. These observations raise the following questions:

Is profitability of key importance for explaining firm exit? What cause firms to exit? What are the characteristics that distinguish firms that exit from those that continue production? Thus, one purpose of the present paper is to address these questions by estimating a dynamic structural model.

Estimation of parameters in dynamic structural models is challenging. Maximum-likelihood estimation is computationally demanding if firms are assumed to take the strategy of their competitors into account; for example, see Ryan (2012). Even if strategic interaction is not considered, it still may be difficult to estimate dynamic structural models of firms' behavior because there may be no numerically tractable criterion function that can form the basis for estimation. Some papers therefore apply the simulated method of moments to estimate structural parameters; for example, see Cooper and Haltiwanger (2006) and Asphjell et al. (2014). Here the econometrician selects a set of moments ad hoc and let the parameters be determined such that the distance between the data moments and the corresponding model-based (simulated) moments are minimized according to some metric.

Similar to the papers referred to in the previous paragraph, we cannot derive a likelihood function that is numerically tractable. However, instead of using the simulated method of moments (with its ad hoc elements) we introduce an auxiliary model that closely mimics the properties of the underlying structural model. In our auxiliary model, a latent variable in the structural model is replaced with a statistic. The likelihood function of the auxiliary model – henceforth referred to as the quasi-likelihood function – can be derived and quasi-maximum likelihood estimates are then combined with the structural model through simulations to estimate structural parameters.

The idea of combining estimation of an auxiliary model with simulations from an underlying "true" model is called indirect inference. This method was proposed by Smith (1993) and developed further into a general methodology by Gourieroux et al. (1993). In Gallant and Tauchen (1996), a related method was proposed, namely to use simulation techniques to match scores of a quasi maximum-likelihood procedure with a Generalized Method of Moments (GMM) approach when the scores are difficult to calculate. The method of Gallant and Tauchen has been referred to as the Efficient Method of Moments (EMM). In the present, paper we draw on Gallant and Tauchen (1996). Indirect inference seems appropriate for our study because it is not possible to compute the exact likelihood, whereas simulation of the model is feasible.

Indirect inference is commonly used in financial econometrics; some examples include stochastic volatility-, exchange rate-, asset price- and interest rate modeling; for example, see Andersen and Lund (1997), Andersen et al. (1999), Bansal et. al. (2007), and Raknerud and Skare (2012). Other examples of application of indirect inference include Magnac et al. (1995) and An and Liu (2000) on labor market transitions, Nagypál (2007) on learning by employees, Collard–Wexler (2013) on the role of demand shocks in the US ready-mix concrete industry, and Li and Zhang (2015) on bidding by heterogeneous actors. Typically, these studies draw on indirect inference because the likelihood function cannot be derived.

We make four contributions to the literature. First, we present a novel theory-consistent econometric model that determines both exit and investment within the framework of stochastic dynamic programming. We do so by extending the Markovian discrete choice model of Rust (1994) by allowing for a continuous decision variable – investment – in addition to a discrete decision variable – whether or not to exit. In particular, we replace the standard simplifying assumption of a state-independent scrap value, see discussion above, by modeling a trade-off between the value of installed capital if production is continued and the value of installed capital if the firm exits – this is how we make the decision to exit truly endogenous.

Second, we present a solution to the main difficulty in applying indirect inference (and more generally, simulation-based inference) in discrete or discrete-continuous choice models, namely that the simulated trajectories are discontinuous functions of the structural parameters. Our solution does not rely on smoothing-functions such as the Generalized Indirect Inference (GII) method of Bruins et al. (2015), but utilizes the "smoothing properties" of the conditional expectation operator, given the simulated state variables.

Third, we contribute to the empirical literature on the relationship between profitability and exit. There is surprisingly little evidence about this relationship. Some studies provide descriptive statistics on exit rates, see Dunne et al. (1988) for US manufacturing industries and Disney et al. (2003) for UK manufacturing. Others use reduced form probit models to examine how profit components influence firm exit; for example, see Olley and Pakes (1996) and Foster et al. (2008). We draw on a rich data set for Norwegian manufacturing firms, and show that it is the cumulated effect over several years of a high risk to exit that distinguishes firms that exit from those that continue production. If, over a long period of time, the expected value of continuing production is low relative to the expected value of exit, the firm has a high probability to exit.

Fourth, we contribute to the literature on the cost of capital adjustment. In the empirical literature, the degree of cost of adjustment has been highly debated; two prominent examples are Hall (2004), who finds small adjustment-cost parameters, and Cooper and Haltiwanger (2006), who conclude that there is significant cost of adjustment. We find significant, but moderate, cost of capital adjustment.

The remainder of this paper is organized as follows: In Section 2, we identify stylized facts about the firms in the data set; these are start-up firms in Norwegian manufacturing industries (1994–2012). Our choice of firms reflects that the exit probability of an incumbent firm may differ systematically from that of a new firm due to self-selection; *surviving* firms are not a random sample of the population of all firms. In the literature, this selection problem has largely been ignored.

In Section 3 we introduce a production model – production requires input of labor, materials (including energy), and capital – and in Section 4 we explain how stochastic dynamic programming can be used to simultaneously determine (in each period) whether the firm will exit and how much the firm will invest if it does not exit. In Section 5, we discuss the stochastic specification of the auxiliary econometric model – it has exactly the same parameters as the data-generating model – and we derive the quasi-likelihood function. In Section 6, we show how to simulate data from the structural model. The parameter estimates are reported in Section 7; all parameter estimates have the expected sign and are statistically significant. In Section 7, we also investigate simplifying model assumptions. Finally, in Section 8 we present our conclusions.

2 Data

Our main data source is a database from Statistics Norway based on register data – the Capital database – which covers the entire population of Norwegian limited liability companies in the manufacturing industry. The main statistical unit in this database is the firm: A firm is defined as "the smallest legal unit comprising all economic activities engaged in by one and the same owner". We use data from the Capital database for the period 1993–2012.

A firm is defined to have exited in year t if it is not recorded in the Capital database in $t+1$ (or later) *and* the firm is registered by the end of $t+2$ as either bankrupt, compulsory liquidated[1] or having closed down for an unspecified reason according to the Norwegian Central Register of Establishments and Enterprises.[2] We limit attention to new firms that were operative in at least 2 years. For each firm, we use the first observation year solely to obtain information about the initial stock of capital. Note that a firm is removed from the Capital database if it is no longer classified as belonging to the manufacturing sector.

We only include firms that are single-plant firms in the start-up year because newly established multi-plant firms are likely to be a continuation of existing establishments under a new organization number (the firm identifier). In the period 2004–12, about 90 percent of the start-up manufacturing firms were single-plant units. These firms accounted for about two-thirds of total employment of all start-ups in their first year. Finally, if a (single-plant) firm A acquires a (single-plant) firm B, then the new multi-plant firm A is kept in the data (whereas B is of course removed).

The Capital database contains annual observations on revenue, wage costs, intermediate expenses (including energy), fixed capital (tangible fixed assets) and many other variables for all Norwegian limited liability manufacturing firms for the period 1993–2012 (see Raknerud, Rønning and Skjerpen, 2004). The database combines information from two sources: (i) accounts statistics for all Norwegian limited liability companies; and (ii) structural statistics for the manufacturing sector.

Table 1 presents summary statistics for three large manufacturing industries and for total manufacturing. The three industries we examine are Wood products (NACE 16), Metal products (NACE 25), and Machinery (NACE 28). In the table, the first and second columns show the numbers of firms and exits by industry for the period 1994–2012. Column three depicts annual exit frequencies; these lie between 3.6 and 3.9 percent. The fourth column in Table 1 shows both the average and median number of person-years in the

[1] This will happen if the firm fails to file an approved account within a deadline.

[2] Mergers and fusions (the activity continues under a new firm identifier) are not included, but are considered as exogenous right-censoring. There is typically a lag of 1 or 2 years between the date of the last approved business account and the official date of bankruptcy or liquidation.

entry year of firms. For the three individual industries, as well as for total manufacturing, the mean is either 5 or 6. Therefore, most firms are small – this is a typical feature of Norwegian manufacturing.

Table 1: **Descriptive statistics for 1994-2012**

Industry (NACE)	No. of firms	No. of exits	Average exit-frequency*	Mean/median person-years
Wood products (16)	1022	250	.036	16/5
Metal products (25)	1504	317	.037	17/6
Machinery (28)	1108	178	.039	17/5
Total manufacturing	10548	2265	.037	23/6

*Number of exits divided by number of firm-years

We have examined how the use of labor (measured as person-hours), materials (intermediate inputs, including energy), and capital changes over time. For each factor of production and each firm in each year, we first calculate the use of a factor in year t relative to the use of this factor in year $t-1$. The graphs for person-hours and materials are almost identical and resemble a normal distribution (see Figure A.1 in Online Appendix A). In contrast, the graph for changes in the stock of capital has somewhat thicker tails than those for person-hours and materials. The thicker tails mean that observations with large (negative or positive) changes are more frequent. In particular, a thicker right tail – the graph is skewed to the right – reflects the intermittent and lumpy nature of investment in Norwegian manufacturing. The distinct pattern of investment calls for another modeling of capital than that of labor and materials (see Section 3).

3 Short-run factor demand

In this section we present our model for price decisions by firms. Because Norwegian firms in the three manufacturing sectors of wood products, metal products, and machinery compete extensively in international markets, we follow the standard in the international trade literature and assume imperfect competition, here specified as monopolistic competition. Hence, each producer (in a sector) faces a demand function of the following form:

$$Q_{it} = \Phi_t P_{it}^{-e} \tag{1}$$

where Q_{it} is the output from firm i at time t, P_{it} is the output price, and Φ_t is an exogenous demand-shift parameter characterizing the size of the market. Furthermore, $e > 1$ is the absolute value of the direct price elasticity. The price elasticity is common to all firms and constant over time.

Let M_{it} denote materials, L_{it} labor, and K_{it} capital. In Section 2 we argued that the modeling of materials and labor should be similar, but this modeling should differ from that for capital. We now assume that the use of materials and labor are determined at the beginning of a time period (variable inputs), whereas capital services in year t are determined by the capital stock at the end of $t-1$, $K_{i,t-1}$. The production function of producer i is assumed to be:

$$Q_{it} = A_{it} K_{i,t-1}^{\chi} \left[M_{it}^{\rho} + w_t L_{it}^{\rho} \right]^{\frac{\varepsilon}{\rho}} , \ \rho < 1 \tag{2}$$

where the elasticity of scale is equal to $\varepsilon + \chi$, the elasticity of substitution between materials and labor is $1/(1-\rho)$, and w_t is a time-varying distribution parameter. Our production function is a nested Cobb-Douglas function defined over capital and a CES aggregate over labor and materials. The specification (2) allows for heterogeneity in productivity across firms: Hicks-neutral changes in efficiency are picked up by A_{it}, which may shift over time and vary across firms, whereas a positive change in w_t can be interpreted as a labor-augmenting innovation. Thus, w_t captures that the efficiency of labor typically changes over time.

The skill composition of labor may differ across firms, and hence, labor productivity may also differ across firms. In a perfect labor market, differences in labor productivity should mirror relative wages. To capture heterogeneity in labor input, we measure L_{it} in efficiency units by dividing the employees of each firm into skill categories based on educational attainment. Following Nilsen et al. (2011), we construct skill-adjusted person-hours, L_{it}, by multiplying the number of person-hours in each skill category by an efficiency factor reflecting the relative wage of that skill category and then summing all categories.[3] The *firm-specific* wage q_{Lit} is measured as y_{Lit}/L_{it}, where y_{Lit} is the firm's

[3]Formally, skill-adjusted labor for firm i is equal to $L_{it} = \sum_{j=1}^{n} \left(q_t^{(j)} / q_{Lt} \right) L_{it}^{(j)}$, where $q_t^{(j)}$ is the average wage in skill category j across *all* firms in the industry (in year t), $q_{Lt} = \sum_j q_t^{(j)}/n$ is the overall average wage, and $L_{it}^{(j)}$ is number of person-hours in skill category j in firm i. We use $n = 3$ skill categories corresponding to primary, secondary, and tertiary (or higher) levels of education.

total wage bill. The firm-specific wage reflects that wages within the same skill category differ across firms, e.g., reflecting differences in local labor market conditions.

Let $\mathbf{q}_{it} = (q_{Mt}, q_{Lit})$ be the vector of the *real* unit price of materials and labor, respectively. All prices have been deflated by the same price index so that in any time period, one dollar of any cost component has the same value as one dollar of a revenue component. We use the price index of capital, q_{Kt}, as the deflator, implying that the real unit price of capital is one.

Assuming that producers are price takers in all factor markets, from Shephard's lemma we find that the short-run cost function is

$$C(\mathbf{q}_{it}, K_{i,t-1}, Q_{it}) = c_{it} \left(\frac{Q_{it}}{A_{it} K_{i,t-1}^{\chi}} \right)^{\frac{1}{\varepsilon}} \tag{3}$$

where

$$c_{it} = [q_{Mt}^{\varrho} + q_{Lit}^{\varrho}/w_t]^{\frac{1}{\varrho}}, \ \varrho = \frac{\rho}{\rho - 1}. \tag{4}$$

Here, c_{it} is a firm-specific price index of variable inputs, i.e., it is derived from the CES-aggregate of materials and labor.

The short-run optimization problem of firm i in the beginning of period t, when the producer knows $\mathbf{q}_{it}$, Φ_t, A_{it} and w_t (and also e, χ, ρ and ε), is to choose – for a given stock of capital – the price that maximizes operating surplus (revenue minus costs of materials and labor). Solving the resulting first-order condition gives the following equations for revenue $R_{it} = P_{it}Q_{it}$ and short-run factor costs $q_{Mt}M_{it}$ and $q_{Lit}L_{it}$:

$$\begin{aligned} \begin{bmatrix} \ln R_{it} \\ \ln(q_{Mt}M_{it}) \\ \ln(q_{Lit}L_{it}) \end{bmatrix} &= \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \ln(\frac{\vartheta_1}{\vartheta_1 + 1}) - \begin{bmatrix} \vartheta_1 \\ \vartheta_1 + \varrho \\ \vartheta_1 + \varrho \end{bmatrix} \ln c_{it} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ln w_t \\ &+ \begin{bmatrix} 0 & 0 \\ \varrho & 0 \\ 0 & \varrho \end{bmatrix} \begin{bmatrix} \ln q_{Mt} \\ \ln q_{Lit} \end{bmatrix} + \mathbf{1}(\vartheta_2 \ln A_{it} + d_t + \kappa \ln K_{i,t-1}) \end{aligned} \tag{5}$$

where $\mathbf{1}$ is a vector of ones,

$$d_t = \frac{1}{(\varepsilon + e - e\varepsilon)} \ln \Phi_t - \vartheta_1 \ \ln(\frac{e}{\varepsilon(e-1)}) \tag{6}$$

and

$$\vartheta_1 = \frac{\varepsilon(e-1)}{(\varepsilon + e - e\varepsilon)} > 0, \vartheta_2 = \frac{(e-1)}{(\varepsilon + e - e\varepsilon)} > 0, \ \kappa = \chi\vartheta_2. \tag{7}$$

Note that if the demand parameter is allowed to be firm-time specific (Φ_{it}), the system (5) is unaltered except that A_{it} is replaced by $A_{it}^* = \Phi_{it}^{1/(e-1)} A_{it}$; that is, A_{it}^* captures both demand shocks (Φ_{it}) and technology shocks (A_{it}). Hence, it is not possible to distinguish between these two factors in the empirical analysis.[4]

Operating surplus, Π_{it}, has the closed form[5]

$$\Pi_{it} = e^{\pi_{it}} K_{i,t-1}^{\kappa} \tag{8}$$

where

$$\pi_{it} = -\ln(1 + \vartheta_1) - \vartheta_1 \ln c_{it} + d_t + \vartheta_2 \ln A_{it}. \tag{9}$$

From (8) and (9) we see that ϑ_1 is the absolute value of the elasticity of operating surplus, Π_{it}, (revenue minus costs of materials and labor) with respect to c_{it} (the variable cost index). Further, (8) shows that κ is the elasticity of operating surplus with respect to the stock of capital.[6] Finally, from the definition of π_{it} in (9), we see that this variable depends on a number of factors that reflect short-run profitability; we will therefore refer to π_{it} as a measure of *short-run profitability*.

Measurement errors Whereas the solution to (5) corresponds to an ex ante production plan that is based on the information available to the firm at the beginning of t, the ex post realizations, i.e., the data, are also determined by other factors; for example, measurement errors and new information obtained during the year. In practice, *observed* revenue (y_R), material costs (y_M), and labor costs (y_L) will not satisfy the strong restrictions imposed by (5). Therefore, we assume that the observed short-run profit factors are equal to the corresponding structural variables except for $\mathbf{e}_{it}$, an additive white noise error term. Define

$$\mathbf{y}_{it} = [\ln(y_{Rit}), \ln(y_{Mit}), \ln(y_{Lit})]'. \tag{10}$$

[4] One might think that we also have another identification problem due to the fact that we observe only a price index for material costs, q_{Mt} (which is normalized to one in the base year). To see that this is not a problem in our model, define $q_{Mt}^* = \lambda q_{Mt}$ for an arbitrary normalization constant λ. Then define $w_t^* = w_t/\lambda^{\varrho}$, $d_t^* = (\vartheta_2/e - 1)\ln \Phi_t + \vartheta_1 \ln \lambda$, and $c_{it}^* = [q_{Lit}^{\varrho}/w_t^* + q_{Mt}^{*\varrho}]^{\frac{1}{\varrho}}$. It is easy to show that (5) still holds with $(q_{Mt}, w_t, d_t, c_{it})$ replaced by $(q_{Mt}^*, w_t^*, d_t^*, c_{it}^*)$. Thus (5) is valid for any normalization of q_{Mt}.

[5] By a straightforward calculation, we find $\Pi_{it} = (1 - \varepsilon(e-1)e^{-1})((q_{Mt}/c_{it})^{\varrho} + (q_{Lit}/c_{it})^{\varrho} w_t^{-1}) c_{it}^{-\vartheta_1} e^{d_t} A_{it}^{\vartheta_2} K_{i,t-1}^{\kappa}$. Then we use that $1 - \varepsilon(e-1)/e = (1+\vartheta_1)^{-1}$ and $((q_{Mt}/c_{it})^{\varrho} + (q_{Lit}/c_{it})^{\varrho})/w_t^{-1} = 1$.

[6] In order to ensure that the optimization with respect to capital is well-defined, we need $\kappa < 1$; our model meets this requirement.

We then assume

$$\mathbf{y}_{it} = \left[\ \ln R_{it}, \quad \ln(q_{Mt}M_{it}), \quad \ln(q_{Lit}L_{it}) \ \right]' + \mathbf{e}_{it}, \tag{11}$$

with

$$\mathbf{e}_{it} \sim \mathcal{N}(\mathbf{0},\Sigma_e). \tag{12}$$

Identification Because A_{it} is unobserved, we cannot identify ϑ_2. To see this, define $a_{it} = \ln A_{it}/\widetilde{k}$ for an arbitrary proportionality factor $\widetilde{k}$ and let $\widetilde{\vartheta}_2 = \widetilde{k}\vartheta_2$. Then

$$\widetilde{\vartheta}_2 a_{it} = \vartheta_2 \ln A_{it} \tag{13}$$

regardless of $\widetilde{k}$. The parameter $\widetilde{\vartheta}_2$ can be identified only by making arbitrary scaling assumptions about a_{it}. To obtain identification, we assume that a_{it} is a stationary AR(1) process with innovation variance equal to one:

$$\begin{aligned} a_{i1} &\sim \mathcal{N}(0, \frac{1}{1-\varphi^2}) \\ a_{it} &= \varphi a_{i,t-1} + \eta_{it}, \ \eta_{it} \sim \mathcal{N}\left(0,1\right), \ t > 1. \end{aligned} \tag{14}$$

Note that any non-zero mean in a_{it} would be absorbed into the term d_t in (5). Hence, the assumption that a_{it} has a zero mean is also a purely identifying restriction. The assumptions about a_{it} in (14) enable us to identify the loading coefficient $\widetilde{\vartheta}_2$, but not the parameter ϑ_2 (because $\widetilde{k}$ is unidentified).

The data-generating model derived from the short-run factor demand model can finally be written as:

$$\begin{aligned} \mathbf{y}_{it} = & \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \ln(\frac{\vartheta_1}{\vartheta_1+1}) - \begin{bmatrix} \vartheta_1 \\ \vartheta_1+\varrho \\ \vartheta_1+\varrho \end{bmatrix} \ln c_{it} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ln w_t \\ & + \begin{bmatrix} 0 & 0 \\ \varrho & 0 \\ 0 & \varrho \end{bmatrix} \begin{bmatrix} \ln q_{Mt} \\ \ln q_{Lit} \end{bmatrix} + \mathbf{1}(\widetilde{\vartheta}_2 a_{it} + d_t + \kappa \ln K_{i,t-1}) + \mathbf{e}_{it}. \end{aligned} \tag{15}$$

It should be noted that this equation is highly non-linear in the parameters ϱ and $\ln w_t$ because c_{it} depends on ϱ and w_t. Identification of ϑ_1 and ϱ follows because q_{Lit} (and c_{it}) varies across firms.

4 Exit and investment dynamics

Let z_t be a binary variable, which is one if the firm operates in year t and zero if the firm exits during year t. If $z_t = 1$, the firm will invest optimally and remain operative in at least one more year, earning an uncertain profit Π_{t+1} in $t+1$. If $z_t = 0$, the firm will realize the scrap value at the end of t.

We take the Markovian discrete choice model of Rust (1994) as a starting point and assume that the period t utility from the choice (I_t, z_t), given the state vector $S_t = (K_{t-1}, \pi_t)$, can be written as:

$$u(S_t, I_t, z_t) + \varepsilon_t(z_t) \tag{16}$$

where $u(S_t, I_t, z_t)$ is operating surplus minus capital expenditures, and $\varepsilon_t(z_t)$ is a random component associated with the discrete choice z_t. By definition we have

$$u(S_t, I_t, z_t) = \begin{cases} \Pi_t - c(I_t) & \text{if } z_t = 1 \text{ (continue)} \\ \Pi_t - c(-(1-\delta)K_{t-1}) & \text{if } z_t = 0 \qquad \text{(exit)} \end{cases} \tag{17}$$

where the function $c(I_t)$ denotes total cost of investment and δ is the rate of depreciation. Operating surplus Π_t follows from S_t and is therefore not affected by z_t and I_t. If $z_t = 0$, t is the terminal period. The firm then sells its remaining capital stock, $I_t = -(1-\delta)K_{t-1}$, and obtains a scrap value, $-c(-(1-\delta)K_{t-1})$, at the end of the year.

Following Rust (1994), we assume that the state vector S_t is Markovian with transition probability $g(dS_{t+1}|S_t, I_t)$ and that $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))$ has a bivariate extreme value distribution with scale parameter τ and location parameters (ξ_0, ξ_1):[7]

$$h(\varepsilon_t) = \prod_{z \in \{0,1\}} \tau \exp\{-\tau\varepsilon_t(z) + \xi_z)\} \exp\{-\exp\{-\tau\varepsilon_t(z) + \xi_z\}\}. \tag{18}$$

Further, the firm's choice of whether to continue production, and if so, how much to invest, follows from the solution of the Bellman equation:

$$V(S_t, \varepsilon_t) = \max_{z_t,\, I_t} \left\{ u(S_t, I_t, z_t) + \varepsilon_t(z_t) + \frac{1}{1+r} E\left[V(S_{t+1}, \varepsilon_{t+1}) | S_t, I_t, z_t\right] \right\}. \tag{19}$$

The value function $V(S_t, \varepsilon_t)$ is characterized in Proposition 1, which is an extension of the discrete choice model of Rust (1994); that is, we allow for a discrete *and* a continuous decision variable. Without loss of generality, we may normalize ξ_1 to zero.

[7] Because $E(\tau\varepsilon_t(z) - \xi_z) = \gamma$ for $z \in \{0, 1\}$, where γ is Eulers' constant, we have $E(\varepsilon_t(z)) = (\gamma + \xi_z)/\tau$.

Proposition 1 *Assume (16)–(18) and that* $S_t = (K_{t-1}, \pi_t)$ *is Markovian with transition probability* $g(dS_{t+1}|S_t, I_t)$, *and* $\xi_1 = 0$. *Then, the expected net present value of the firm is*

$$V(S_t, \varepsilon_t) = \max_{z_t \in \{0,1\}} [\Pi_t + \nu(S_t, z_t) + \varepsilon_t(z_t)] \tag{20}$$

where

$$\nu(S_t, 0) = -c(-(1-\delta)K_{t-1}) \tag{21}$$

and

$$\begin{aligned} \nu(S_t, 1) \quad = \quad & \max_{I_t} \left\{ -c(I_t) + \frac{1}{1+r} \times \right. \\ & \left. \int \left[\Pi_{t+1} + \frac{1}{\tau} \ln \left[\exp(\tau \nu(S_{t+1}, 0) + \xi_0) + \exp(\tau \nu(S_{t+1}, 1)) \right] \right] g(dS_{t+1}|S_t, I_t) \right\}. \end{aligned} \tag{22}$$

The conditional exit probability has the closed form expression $\Pr(z_t = 0|S_t, z_{t-1} = 1) = p(S_t)$, *where*

$$p(S_t) = \frac{1}{1 + \exp\{-[-\tau c(-(1-\delta)K_{t-1}) - \tau \nu(S_t, 1) + \xi_0]\}}. \tag{23}$$

The proof of Proposition 1 is given in Online Appendix B, part I. The exit probability $p(S_t)$ is the conditional probability that $z_t = 0$, given $S_t = (K_{t-1}, \pi_t)$ and $z_{t-1} = 1$ (the firm has not already closed down). Exit is an absorbing state, so $z_t = 0$ implies $z_{t+1} = 0$. In Proposition 1, $\nu(S_t, 1)$ is the net present value of the firm if it does not exit in the current period ($z_t = 1$) and makes optimal investment decisions now (I_t) and in the future:

$$\nu(S_t, 1) = \max_{I_t} \left\{ -c(I_t) + \frac{1}{1+r} E\left[V(S_{t+1}, \varepsilon_{t+1})|S_t, I_t, z_t = 1\right] \right\}.$$

Above, we introduced the general cost of investment function $c(I_t)$. We now specify this function. Our starting point is that there is one type of capital adjustment cost, namely that the resale price of capital relative to purchaser price of capital, henceforth termed s, is less than one; for example, see Abel and Eberly (1996):

$$c_s(I) = \begin{cases} I & \text{if } I \geq 0 \\ sI & \text{if } I < 0 \end{cases} \quad s \leq 1. \tag{24}$$

According to (24), upon selling capital ($I < 0$), the firm may not obtain the purchaser price of capital: Markets for old capital may be imperfect, or there may be large transaction costs, that is, $s < 1$. For parts of the capital stock there may even be no market (i.e., zero price) because of, for example, asymmetric information. The special case $s = 1$ corresponds to the neoclassical theory of investment. With our specification (24), $c_s(I)$ is weakly convex with a kink at zero.[8]

The value function and its parameterization The value function $\nu(S_t, 1)$ is the solution to the fixed-point equation (22), but $\nu(S_t, 1)$ has no closed form; it is implicitly defined as a function of the model parameters θ (to be specified below). Under standard regularity conditions, $\nu(S_t, 1)$ will be differentiable with respect to θ.

Relation (22) contains a general profit expression Π_t and a general cost of investment function $c(I_t)$. These are now replaced with the corresponding elements in our structural model; that is, we specify Π_t using (8) and replace $c(I_t)$ by (24). Further, in (22) the transition probability $g(dS_{t+1}|S_t, I_t)$ is a function of S_t and I_t, where $S_t = (K_{t-1}, \pi_t)$; the transition probability depends on K_{t-1} and π_t. In our structural model, we assume that the change in the stock of capital follows the standard *deterministic* rule $K_t = (1-\delta)K_{t-1} + I_t$, whereas π_t is assumed to be a stationary AR(1) process:

$$\begin{aligned} \pi_t &= \mu + \varphi(\pi_{t-1} - \mu) + \zeta_t \\ \zeta_t &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{25}$$

with corresponding transition density denoted $g_{(\varphi,\mu,\sigma)}(\pi_{t+1}|\pi_t)$. Note that because of (9) and (13), the AR-coefficient in (25) must be the same as that in (14). The vector of structural parameters can now be specified as

$$\omega^{str} = (\kappa, \vartheta_1, \varrho, \varphi, \mu, \sigma, \tau, s, \xi_0).$$

In addition to the structural parameters, our model contains nuisance parameters

$$\omega^{nui} = (\widetilde{\vartheta}_2, d_1, ..., d_T, w_1, ..., w_T, vech(\Sigma_e)').$$

The nuisance parameters do not have any interesting economic interpretation, but they are needed to simulate data from the structural model. Therefore, both ω^{str} and ω^{nui}

[8] Figure A.2 in online Appendix A illustrates how the net value of continuing, $v(S_t, 1) - v(S_t, 0)$, depends both on s and on our measure of short-run profitability, π_t.

must be estimated simultaneously. Hence, define $\theta = (\omega^{nui}, \omega^{str})$ as the vector of all data-generating model parameters.[9]

The functional operator corresponding to the right-hand side in (22) can now be specified as:

$$\begin{aligned}\Psi_\theta(\nu)(K,\pi) &= \max_I \Bigg\{-c_s(I) + \frac{1}{1+r}\times \\ &\int \left[\exp(\pi')K'^{\kappa} + \frac{1}{\tau}\ln\left[\exp(\tau s(1-\delta)K' + \xi_0) + \exp(\tau\nu(K',\pi'))\right]\right] g_{(\varphi,\mu,\sigma)}(\pi'|\pi)d\pi'\Bigg\} \\ \text{s.t. } K' &= (1-\delta)K + I. \end{aligned} \tag{26}$$

The value function $\nu(S_t, 1)$ is then equal to $\nu_\theta(K_{t-1}, \pi_t)$, where ν_θ is the solution to the fixed-point equation

$$\nu_\theta = \Psi_\theta(\nu_\theta). \tag{27}$$

Furthermore, the conditional exit probability (23) can be specified as:

$$p_\theta(K_{t-1}, \pi_t) = \frac{1}{1 + \exp\{-[\tau s(1-\delta)K_{t-1} - \tau\nu_\theta(K_{t-1}, \pi_t) + \xi_0]\}}. \tag{28}$$

We will later need to differentiate $p_\theta(\cdot)$ with respect to θ (see Section 6). For this purpose, we apply the inverse function theorem to (27) to obtain:

$$\frac{\partial \nu_\theta}{\partial \theta} = \left[Id - \frac{\partial \Psi_\theta(\nu_\theta)}{\partial \nu}\right]^{-1} \frac{\partial \Psi_\theta(\nu_\theta)}{\partial \theta} \tag{29}$$

where Id denotes the identity matrix and the derivatives of the value function $\Psi_\theta(\nu)$ are obtained by applying the envelope theorem to (26) (see Milgrom and Segal, 2002).

The structure of the solution in (26) is well known from the theory of investment under uncertainty (see Stokey, 2009, Ch. 11). Consider a firm that at the end of t – "just before" making its investment decision – has a remaining stock of capital equal to $(1-\delta)K_{t-1}$. Then there exist unique threshold values $\underline{k_\theta}(\pi_t)$ and $\overline{k_\theta}(\pi_t)$ such that optimal investment is zero in a "region of inactivity" $(1-\delta)K_{t-1} \in [\underline{k_\theta}(\pi_t), \overline{k_\theta}(\pi_t)]$. If the stock of capital is outside this region, the firm adjusts its stock of capital immediately to one of the boundaries: If $(1-\delta)K_{t-1} < \underline{k_\theta}(\pi_t)$, the firm chooses $I_t > 0$ such that $K_t = \underline{k_\theta}(\pi_t)$, where the marginal return to a unit of investment equals the acquisition price 1. If $(1-\delta)K_{t-1} > \overline{k_\theta}(\pi_t)$, the firm chooses $I_t < 0$ such that $K_t = \overline{k_\theta}(\pi_t)$, where the marginal return to a unit of investment equals the selling price s.

[9] For partitioning of row vectors, we adopt the following notation: If $u = (u_1, u_2)$, then $(u, u_3) = (u_1, u_2, u_3)$. Thus, for functions f_v with $v = (u_1, u_2, u_3)$, $f_v = f_{(u_1,u_2,u_3)} = f_{(u,u_3)}$.

To solve the fixed-point equation $\nu_\theta = \Psi_\theta(\nu_\theta)$ numerically, it is necessary to discretize the state space so that $K \in \mathcal{K} = \{K_{(i)}\}_{i=1}^{N_K}$ and $\pi \in \mathcal{A} = \{\pi_{(j)}\}_{j=1}^{N\pi}$. Then $\mathcal{K} \times \mathcal{A}$ consists of $N_K N_\pi$ grid points. The standard solution method is successive approximations (see Online Appendix B, part II, for details and formulas). To discretize the continuous π_t-process, we use the standard finite-state approximation of Tauchen (1986), with a fixed grid size, to obtain transition probabilities $g^D_{(\varphi,\mu,\sigma)}(\pi_{(j)}|\pi_{(i)})$. We first apply his method on the standardized AR(1) process $u_t = \varphi u_{t-1} + \widetilde{\varepsilon}_t$ ($\widetilde{\varepsilon}_t$ has mean zero and unit variance). Then, we set $\pi_{(j)} = \mu + \sigma u_{(j)}$, where $\{u_{(j)}\}_{j=1}^{N\pi}$ is the discretized state space of u_t. To discretize the K_t-process, we use a fixed grid size on logarithmic scale.[10]

5 Parameterization and estimation of the auxiliary model

Our estimation strategy draws on the efficient method of moments (see Gallant and Tauchen, 1996), and consists of the following steps. First, we specify an auxiliary model that approximates our structural model. Next, we derive a likelihood function for the auxiliary model. The likelihood function of the auxiliary model is referred to as the quasi-likelihood function. We use real observations to estimate the parameters in the auxiliary model, θ^a. The data-generating model and the auxiliary model have the same parameters: superscript a denotes that a given parameter (or parameter vector) enters the quasi-likelihood function – as opposed to the data-generating model. The estimator of θ^a is denoted $\widehat{\theta^a}$. Because we use maximum quasi-likelihood to estimate the auxiliary parameters, the value of the resulting score function is per definition zero.

Next, for a given choice of the parameters in the structural model (θ), we simulate data from this model. The computer-generated data are used to recalculate the score function, with θ^a fixed at $\widehat{\theta^a}$. Since the simulated data differ from the observations, the corresponding score will in general differ from zero. The indirect inference estimator finds, through simulations of the economic model for a given θ, the value of θ that minimizes (in a weighted mean-squared error sense) the score vector evaluated at $\widehat{\theta^a}$. Note that

[10] Our approach is analogous to Tauchen (1990). We let the grid extend four standard deviations on both side of the *unconditional* mean of $\ln K^*(\pi_t)$, where $K^*(\pi_t)$ is the optimal (*steady state*) capital stock in the special case without adjustment cost ($s = 1$): $\ln K^*(\pi_t) = 1/(1-\kappa)[\ln(\kappa/(r+\delta))+\mu+\phi(\pi_t-\mu)+\sigma^2/2]$ (see also Footnote 17).

we estimate the parameters in the auxiliary model only *once*. The one-to-one relation between θ and θ^a means that our score-based indirect inference estimator is asymptotically equivalent to the more common – but in our case, infeasible – distance-based indirect inference estimators (see Gourieroux and Monfort, 1996, p. 66).[11]

Our estimation strategy is slightly modified compared with that used by Gallant and Tauchen (1996). First, the model parameters θ are partitioned into three subvectors: $\theta = (\theta_1, \theta_2, \theta_3)$. The corresponding parameters of the auxiliary model are $\theta^a = (\theta_1^a, \theta_2^a, \theta_3^a)$. The auxiliary model is estimated in three stages, where each stage corresponds to a *partial* quasi-likelihood maximization with respect to a subvector of θ^a. In the first stage, we estimate θ_1^a. The corresponding parameters of the data-generating model are:

$$\theta_1 = (\omega^{nui}, \kappa, \vartheta_1, \varrho, \varphi).$$

These are the parameters of the factor demand model (15). The difference between the structural and auxiliary factor demand model is that the latter treats investment and exit decisions as being strictly exogenous.

In the second stage, we define

$$\theta_2 = (\mu, \sigma)$$

and replace the latent state variable π_{it} with a statistic, $\widehat{\pi}_{it}$, which is calculated from the data. This approximation yields an auxiliary transition density $g_{(\varphi^a, \theta_2^a)}(\widehat{\pi}_{i,t+1}|\widehat{\pi}_{it})$ corresponding to $g_{(\varphi, \theta_2)}(\pi_{i,t+1}|\pi_{it})$. The auxiliary parameters θ_2^a are estimated in the second stage with θ_1^a fixed at its estimate from the first stage.

In the third stage, we estimate the auxiliary exit model with respect to the remaining auxiliary parameters, θ_3^a. The corresponding structural parameters are:

$$\theta_3 = (\tau, s, \xi_0)\,.$$

This stage is conditional on the estimated auxiliary parameters from the first two stages.

[11] The small sample properties of indirect inference estimators were first studied by Gourieroux, Renault and Touzi (see Gourieroux et al., 2000). Some studies have shown that score-based estimators often have poor finite sample properties relative to distance-based estimators; for example, see Michaelides and Ng (2000) and Duffee and Stanton (2008). Fuleky and Zivot (2014) propose a score-based estimator that has the same asymptotic properties as the EMM estimator.

5.1 Quasi-likelihood estimation of the factor demand model

The data on firm i can be seen as the realization of the stopped stochastic process $y_i = (K_{i0}, K_{i1}, \mathbf{y}_{i1}, ..., K_{i\tau_i}, \mathbf{y}_{i\tau_i})$, where $\mathbf{y}_{it}$ is defined in (10) and $1 \leq \tau_i \leq T$ is the stopping time. Here, T is the year of right censoring. To simplify notation, we have assumed that the firm enters at $t = 1$ and that the year of right censoring is the same for all firms.[12] The reason for stopping is either censoring or exit; in the latter case, $z_{i,\tau_i+1} = 0$. Note that $z_{it} = 1$ for $t \leq \tau_i$, while $z_{i,\tau_i+1} = 1$ (the firm is not observed) or $z_{i,\tau_i+1} = 0$ (the firm has exited). Formally,

$$\tau_i = \min(T, \max t : z_{it} = 1). \tag{30}$$

The last observed value of $\mathbf{y}_{it}$ is at τ_i, whereas z_{it} is observed at $\tau_i + 1$.

Define $z_i = (z_{i2}, ..., z_{i,\tau_i+1})$. Then, under the assumption that z_{it} and K_{it} are strictly exogenous variables, we obtain a simple log-likelihood function of θ_1^a given (y_i, z_i):

$$l^1(\theta_1^a; y_i, z_i) = \sum_{t=2}^{T} z_{it} \ln f_{\theta_1^a}(\mathbf{y}_{it}|\mathbf{y}_{i,t-1}, K_{i,t-1}, ..., \mathbf{y}_{i1}, K_{i1}) + \ln f_{\theta_1^a}(\mathbf{y}_{i1}|K_{i0}) \tag{31}$$

where $f_{\theta_1^a}(\mathbf{y}_{it}|\cdot)$ is the normal density implied by (12), (14), and (15). Note that $z_{it} = 0$ when $\mathbf{y}_{it}$ is unobserved.

The assumption that z_{it} and K_{it} are strictly exogenous variables violates the structural model. Hence, (31) is a log-likelihood function of an auxiliary model – a *quasi* log-likelihood function. The auxiliary model is straightforward to cast in a state-space form, where $\mathbf{y}_{it}$ is the observation vector with normally distributed measurement errors (12), and a_{it} is the state variable with transition equation (14). One-step ahead predictions and prediction error covariance matrices are readily available using the Kalman filter (see Shumway and Stoffer, 2000). Hence, it is straightforward to calculate $l^1(\theta_1^a; y_i, z_i)$. To obtain analytical derivatives, we use a decomposition of $l^1(\theta_1^a; y_i, z_i)$, which is well-known from the EM-algorithm (see Koopman and Shephard, 1992). The partial quasi-likelihood estimator $\widehat{\theta}_1^a$ is obtained by maximizing $l^1(\theta_1^a; y_i, z_i)$ with respect to θ_1^a.

5.2 Quasi-likelihood estimation of the exit probability model

There are two problems related to the exit probability (28). The first is to solve the functional fixed-point equation (27). This problem is difficult, but tractable, as we show

[12] There is no loss of generality in assuming that all firms enter at $t = 1$. Furthermore, T can be replaced by a firm-specific (exogenous) year of right censoring, T_i, in all formulas below.

in Online Appendix B, part II. The second is that π_{it} is a latent state variable. To handle the second problem, we use that $\pi_{it} = \ln(\Pi_{it}/K^{\kappa}_{i,t-1})$ (see (8)) and approximate π_{it} by the statistic

$$\widehat{\pi}_{it} = \ln\left(\max\{\widehat{\Pi}_{it}/K^{\widehat{\kappa^a}}_{i,t-1}, e^{\pi_{(1)}}\}\right) \text{ for } 1 \leq t \leq \tau_i. \tag{32}$$

Here, $\widehat{\Pi}_{it}$ is observed operating surplus:

$$\widehat{\Pi}_{it} = y_{Rit} - y_{Mit} - y_{Lit} \tag{33}$$

and $\pi_{(1)}$ is the lower threshold of π_t in the finite state space.

Next, we replace the latent variable π_{it} by $\widehat{\pi}_{it}$ in the transition density $g_{(\widehat{\varphi^a},\theta_2^a)}(\cdot|\cdot)$ and define $\widehat{\pi}_i = (\widehat{\pi}_{i1}, ..., \widehat{\pi}_{i\tau_i})$. This gives us a simple partial quasi-likelihood of θ_2^a, given $\widehat{\theta}_1^a$ and $(z_i, \widehat{\pi}_i)$:

$$l^2(\theta_2^a|\widehat{\theta}_1^a; z_i, \widehat{\pi}_i) = \sum_{t=2}^{T} z_{it} \ln g_{(\widehat{\varphi^a},\theta_2^a)}(\widehat{\pi}_{it}|\widehat{\pi}_{i,t-1}) \tag{34}$$

(recall that $\widehat{\varphi^a}$ is included in the vector $\widehat{\theta}_1^a$). The partial quasi-likelihood estimator $\widehat{\theta}_2^a$ is the maximizer of (34) with respect to θ_2^a.

In the last step of the specification of the auxiliary model, we obtain a partial quasi-likelihood estimate of θ_3^a. We first replace π_{it} by $\widehat{\pi}_{i,t\wedge\tau_i}$ in (28), where $t \wedge s \equiv \min(t, s)$, to reflect that $\widehat{\pi}_{it}$ is not observed when $z_{it} = 0$. We then approximate the structural exit probability, $p_\theta(K_{i,t-1}, \pi_{it})$, see (28), by the auxiliary model

$$p_{\theta^a}(K_{i,t-1}, \widehat{\pi}_{i,t\wedge\tau_i}) \equiv \frac{1}{1 + \exp\left\{-\left(\tau^a s^a(1-\delta)K_{i,t-1} - \tau^a \nu_{\theta^a}(K_{i,t-1}, \widehat{\pi}_{i,t\wedge\tau_i}) + \xi_0^a\right)\right\}}. \tag{35}$$

If $\theta = \theta^a$, then $\nu_{\theta^a}(K, \pi) = \nu_\theta(K, \pi)$ – the true value function defined in (27). The difference between ν_θ and ν_{θ^a} is that ν_θ enters the true structural model (28) (through $\nu_\theta(K_{i,t-1}, \pi_{it})$), whereas ν_{θ^a} enters the auxiliary model (through $\nu_{\theta^a}(K_{i,t-1}, \widehat{\pi}_{i,t\wedge\tau_i})$).

Let $p_{(\widehat{\theta_1^a},\widehat{\theta_2^a},\theta_3^a)}$ denote p_{θ^a} as a function of θ_3^a with (θ_1^a, θ_2^a) fixed at $(\widehat{\theta}_1^a, \widehat{\theta}_2^a)$. The corresponding partial quasi log-likelihood function of θ_3^a given $(\widehat{\theta}_1^a, \widehat{\theta}_2^a)$ and $(y_i, z_i, \widehat{\pi}_i)$ is:

$$\begin{aligned} l^3(\theta_3^a|\widehat{\theta}_1^a, \widehat{\theta}_2^a; y_i, z_i, \widehat{\pi}_i) = {} & \textstyle\sum_{t=1}^{T} z_{i,t+1} \ln\left(1 - p_{(\widehat{\theta_1^a},\widehat{\theta_2^a},\theta_3^a)}(K_{it}, \widehat{\pi}_{i,t+1\wedge T})\right) \\ & + \textstyle\sum_{t=1}^{T} (z_{it} - z_{i,t+1}) \ln p_{(\widehat{\theta_1^a},\widehat{\theta_2^a},\theta_3^a)}(K_{it}, \widehat{\pi}_{it}). \end{aligned} \tag{36}$$

To explain (36), note that there are three possibilities for the "weights" $z_{i,t+1}$ and $z_{it} - z_{i,t+1}$ in (36). Either i) the firm remains operative in $t+1$ ($z_{i,t+1} = 1$ and $z_{it} - z_{i,t+1} = 0$), or ii) an exit decision is made in $t + 1$ ($z_{i,t+1} = 0$ and $z_{it} - z_{i,t+1} = 1$), or iii) the firm has already exited ($z_{i,t+1} = 0$ and $z_{it} - z_{i,t+1} = 0$). The respective quasi log-likelihood contributions are (the log of): $1 - p_{\theta^a}(\ K_{it}, \widehat{\pi}_{i,t+1\wedge T})$ in case i) (where $t+1 \wedge \tau_i = t+1 \wedge T$), and $p_{\theta^a}(K_{it}, \widehat{\pi}_{it})$ in case ii) (where $t + 1 \wedge \tau_i = t$). In case iii), the quasi log-likelihood contribution is obviously 0.

We obtain the partial quasi-likelihood estimator of $\widehat{\theta}_3^a$ by maximizing (36) with respect to θ_3^a. This optimization problem is computationally demanding as it requires reevaluation of the value function ν_{θ^a} for each trial value θ^a, which means that the functional fixed-point equation (27) has to be solved each time a trial value is tested. Algorithmic details are given in Online Appendix B, part II.

6 Indirect inference

The partial quasi-likelihood estimator $\widehat{\theta}^a = (\widehat{\theta}_1^a, \widehat{\theta}_2^a, \widehat{\theta}_3^a)$ satisfies a score moment condition. To see this, let $Y_i = (y_i, z_i, \widehat{\pi}_i)$ and define[13]

$$\begin{aligned} l(\theta^a|Y_i) &= l^1(\theta_1^a; y_i, z_i) + l^2(\theta_2^a|\theta_1^a; z_i, \widehat{\pi}_i) + l^3(\theta_3^a|\theta_1^a, \theta_2^a; y_i, z_i, \widehat{\pi}_i) \\ \frac{\partial l(\theta^a|Y_i)}{\partial \theta^a} &= \left[\frac{\partial l^1(\theta_1^a; y_i, z_i)}{\partial \theta_1^a}', \frac{\partial l^2(\theta_2^a|\theta_1^a; z_i, \widehat{\pi}_i)}{\partial \theta_2^a}', \frac{\partial l^3(\theta_3^a|\theta_1^a, \theta_2^a; y_i, z_i, \widehat{\pi}_i)}{\partial \theta_3^a}' \right]'. \end{aligned} \tag{37}$$

Then $\widehat{\theta}^a$ satisfies the score condition

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial l(\widehat{\theta}^a|Y_i)}{\partial \theta^a} = 0. \tag{38}$$

Next, define the binding function

$$b(\theta, \theta^a) = E_\theta \left(\frac{\partial l(\theta^a|Y_i)}{\partial \theta^a} \right) \tag{39}$$

where $E_\theta(\cdot)$ means that the expected value is evaluated at the parameter vector θ. Let θ^0 denote the true values of θ. Further, let θ^{a*} be the vector of pseudo-true parameters, i.e., the probability limit of $\widehat{\theta}^a$ when $N \to \infty$ (for now, we suppress the dependence of $\widehat{\theta}^a$ on N in the notation). Relations (38)–(39) imply that θ^{a*} is determined by the moment condition

$$b(\theta^0, \theta^{a*}) = 0. \tag{40}$$

[13] Throughout this paper, gradient vectors are column vectors. Consequently, the j'th column of a Jacobian matrix contains the gradient of the j'th component of a (row- or column) vector function.

Equation (40) implicitly defines a mapping from θ^0 to θ^{a*}. From the implicit function theorem, we obtain:

$$\frac{\partial \theta^{a*}}{\partial \theta^0} = -\frac{\partial b(\theta^0, \theta^{a*})}{\partial \theta}\left[\frac{\partial b(\theta^0, \theta^{a*})}{\partial \theta^a}\right]^{-1}. \tag{41}$$

The purpose of indirect inference is to establish a link between θ^0 and θ^{a*} through simulations, which enables estimation of θ^0 from $\widehat{\theta^a}$ without knowing the binding function. To this end, we can simulate S trajectories Y_i for each of the N firms, i.e., SN trajectories in total. Let $Y_i^{(s)}(\theta) = (y_i^{(s)}(\theta), z_i^{(s)}(\theta), \widehat{\pi}_i^{(s)}(\theta))$ denote an arbitrary simulated trajectory for firm i for a given θ (see Section 6.1 for details). The binding function $b(\theta, \theta^a)$ can then be estimated through simulations as follows:

$$\frac{1}{NS}\sum_{i=1}^{N}\sum_{s=1}^{S}\frac{\partial l(\theta^a | Y_i^{(s)}(\theta))}{\partial \theta^a}. \tag{42}$$

Ideally, the estimate of θ^0, $\widehat{\theta}$, is the value of θ that makes (42) equal to 0 when $\theta^a = \widehat{\theta^a}$. This is not possible to achieve in practice for two reasons: First, the simulation of the discrete choice will change discontinuously from 0 to 1 or vice versa as θ varies. Second – and regardless of the continuity issue – we may not be able to find a θ that makes (42) equal to zero (with $\theta^a = \widehat{\theta^a}$). Our efficient method of moments estimator therefore i) replaces $Y_i^{(s)}(\theta)$ with a smooth trajectory $Y_i^{*(s)}(\theta)$ (see Section 6.1), and ii) minimizes the deviation of the average simulated score from 0 according to a specific metric ($|| \cdot ||_\Omega$):

$$\widehat{\theta} = \arg\min_{\theta} \left\| \sum_{i=1}^{N}\sum_{s=1}^{S}\frac{\partial l(\widehat{\theta^a} | Y_i^{*(s)}(\theta))}{\partial \theta^a} \right\|_\Omega \tag{43}$$

(see Section 6.2 for a definition of the metric $|| \cdot ||_\Omega$). The smoothing consists in replacing the simulated z_{it}, $z_{it}^{(s)}(\theta)$, with its conditional expectation given the simulated state variables, $\widehat{z}_{it}^{(s)}(\theta)$. That is, $\widehat{z}_{it}^{(s)}(\theta)$ is a conditional survival probability.

6.1 Simulation of trajectories from the structural model

In this Subsection, we will i) show how to calculate $Y_i^{*(s)}(\theta)$, ii) demonstrate that $E_\theta(\partial l(\theta^a | Y_i^{*(s)}(\theta))/\partial \theta^a) = b(\theta, \theta^a)$, and iii) show that $Y_i^{*(s)}(\theta)$ is a continuous function of θ. It follows that

$$\frac{1}{NS}\sum_{i=1}^{N}\sum_{s=1}^{S}\frac{\partial l(\widehat{\theta^a} | Y_i^{*(s)}(\theta))}{\partial \theta^a} \tag{44}$$

converges in probability towards $b(\theta, \theta^{a*})$ as $N \to \infty$ and $\widehat{\theta}$ is a consistent estimator of θ^0.

First, the algorithm for generating the smooth trajectory $Y_i^{*(s)}(\theta)$ is as follows:

Let θ be given (We use $\widehat{\theta^a}$ as the initial value).

A.1 Solve (27) to obtain $\nu_\theta(K,\pi)$ and the corresponding optimal investment rule $i_\theta(K,\pi)$ (the optimal I in (26)).

A.2 For given i and s: Set $t=1$ and $K_{i0}^{(s)}(\theta)=K_{i0}$ (the actual initial value of firm i).

A.3 Draw $a_{it}^{(s)}(\theta)$ from (14) and set $\pi_{it}^{(s)}(\theta)=\widetilde{\vartheta}_2 a_{it}^{(s)}(\theta)-\ln(1+\vartheta_1)-\vartheta_1\ln c_{it}+d_t$.

A.4 Draw $\mathbf{e}_{it}^{(s)}(\theta)$ from (12) and obtain $\mathbf{y}_{it}^{(s)}(\theta)$ from (15) and $\widehat{\pi}_{it}^{(s)}(\theta)$ from (32)–(33).

A.5 For $t=1$: set $\widehat{z}_{it}^{(s)}(\theta)=1$. For $t>1$: set $\widehat{z}_{it}^{(s)}(\theta)=\left(1-p_\theta(K_{i,t-1}^{(s)}(\theta),\pi_{it}^{(s)}(\theta))\right)\widehat{z}_{i,t-1}^{(s)}(\theta)$ (cf. (28)).

A.6 For $t<T+1$: set $K_{it}^{(s)}(\theta)=(1-\delta)K_{i,t-1}^{(s)}(\theta)+i_\theta(K_{i,t-1}^{(s)}(\theta),\pi_{it}^{(s)}(\theta))$. For $t=T+1$: stop.

A.7 Set $t=t+1$ and go to A.3.

The generated trajectory is $Y_i^{*(s)}(\theta)=(y_i^{(s)}(\theta),\widehat{z}_i^{(s)}(\theta),\widehat{\pi}_i^{(s)}(\theta))$ with

$$
\begin{aligned}
y_i^{(s)}(\theta) &= (K_{i0}^{(s)}(\theta),K_{i1}^{(s)}(\theta),\mathbf{y}_{i1}^{(s)}(\theta),...,K_{iT}^{(s)}(\theta),\mathbf{y}_{iT}^{(s)}(\theta))\\
\widehat{z}_i^{(s)}(\theta) &= (\widehat{z}_{i2}^{(s)}(\theta),...,\widehat{z}_{i,T+1}^{(s)}(\theta))\\
\widehat{\pi}_i^{(s)}(\theta) &= (\widehat{\pi}_{i1}^{(s)}(\theta),...,\widehat{\pi}_{iT}^{(s)}(\theta)).
\end{aligned}
$$

Second, define $W_{it}^{(s)}(\theta)=(K_{i,t-1}^{(s)}(\theta),\pi_{it}^{(s)}(\theta))$ and $W_{i,1:t}^{(s)}(\theta)=\{W_{ik}^{(s)}(\theta)\}_{k=1}^t$ and let $z_{it}^{(s)}(\theta)$ be the corresponding discrete choice. Proposition 2 establishes that $\widehat{z}_{it}^{(s)}(\theta)$ – defined recursively in A.5 – is the conditional expectation of $z_{it}^{(s)}(\theta)$ given $W_{i,1:t}^{(s)}(\theta)$. Before stating Proposition 2, we write down the components of $\partial l(\theta^a|Y_i^{*(s)}(\theta))/\partial\theta^a$ for later reference:

$$
\frac{\partial l(\theta^a|Y_i^{*(s)}(\theta))}{\partial\theta^a}=\left[\frac{\partial l^1(\theta_1^a;y_i^{(s)}(\theta),\widehat{z}_i^{(s)}(\theta))}{\partial\theta_1^a}',\ \frac{\partial l^2(\theta_2^a|\theta_1^a;\widehat{z}_i^{(s)}(\theta),\widehat{\pi}_i^{(s)}(\theta))}{\partial\theta_2^a}',\ \frac{\partial l^3(\theta_3^a|\theta_1^a,\theta_2^a;y_i^{(s)}(\theta),\widehat{z}_i^{(s)}(\theta),\widehat{\pi}_i^{(s)}(\theta))}{\partial\theta_3^a}'\right]' \tag{45}
$$

(cf. (37)), with

$$\begin{aligned}
\frac{\partial l^1(\theta_1^a; y_i^{(s)}(\theta), \widehat{z}_i^{(s)}(\theta))}{\partial \theta_1^a} &= \frac{\partial \ln f_{\theta_1^a}(\mathbf{y}_{i1}^{(s)}(\theta)|K_{i0}^{(s)}(\theta))}{\partial \theta_1^a} + \sum_{t=2}^{T} \Big[\widehat{z}_{it}^{(s)}(\theta) \times \\
&\quad \frac{\partial \ln f_{\theta_1^a}(\mathbf{y}_{it}^{(s)}(\theta)|\mathbf{y}_{i,t-1}^{(s)}(\theta), K_{i,t-1}^{(s)}(\theta), ..., \mathbf{y}_{i1}^{(s)}(\theta), K_{i1}^{(s)}(\theta))}{\partial \theta_1^a} \Big] \\
\frac{\partial l^2(\theta_2^a|\theta_1^a; \widehat{z}_i^{(s)}(\theta), \widehat{\pi}_i^{(s)}(\theta))}{\partial \theta_2^a} &= \sum_{t=2}^{T} \widehat{z}_{it}^{(s)}(\theta) \frac{\partial \ln g_{(\varphi^a, \theta_2^a)}(\widehat{\pi}_i^{(s)}(\theta)|\widehat{\pi}_{i,t-1}^{(s)}(\theta))}{\partial \theta_2^a} \\
\frac{\partial l^3(\theta_3^a|\theta_1^a, \theta_2^a; y_i^{(s)}(\theta), \widehat{z}_i^{(s)}(\theta), \widehat{\pi}_i^{(s)}(\theta))}{\partial \theta_3^a} &= \sum_{t=1}^{T} \Bigg[\widehat{z}_{i,t+1}^{(s)}(\theta) \frac{\partial \ln \Big(1 - p_{(\theta_1^a, \theta_2^a, \theta_3^a)}(K_{it}^{(s)}(\theta), \widehat{\pi}_{i,t+1\wedge T}^{(s)}(\theta))\Big)}{\partial \theta_3^a} \\
&\quad + (\widehat{z}_{it}^{(s)}(\theta) - \widehat{z}_{i,t+1}^{(s)}(\theta)) \frac{\partial \ln p_{(\theta_1^a, \theta_2^a, \theta_3^a)}(K_{it}^{(s)}(\theta), \widehat{\pi}_{it}^{(s)}(\theta))}{\partial \theta_3^a} \Bigg]
\end{aligned} \tag{46}$$

(cf. (31), (34) and (36)).

Proposition 2 *Let* $\widehat{z}_{it}^{(s)}(\theta)$ *and* $Y_i^{*(s)}(\theta)$ *be calculated recursively using the algorithm A.1–A.7. Then, with* $\partial l(\theta^a|Y_i^{*(s)}(\theta))/\partial \theta^a$ *as in (45)–(46),*

$$\widehat{z}_{it}^{(s)}(\theta) = E_\theta(z_{it}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta)) \tag{47}$$

and

$$E_\theta \left(\frac{\partial l(\theta^a|Y_i^{*(s)}(\theta))}{\partial \theta^a} \right) = b(\theta, \theta^a). \tag{48}$$

The proof is provided in the Appendix, part I. An illustration of Proposition 2 on a (very) simplified version of our model, that allows for closed-form solutions, is provided in the Appendix, part II.[14]

Third, let us examine the continuity properties of $Y_i^{*(s)}(\theta)$. We first note that the simulation of $a_{it}^{(s)}(\theta)$, $\pi_{it}^{(s)}(\theta)$ and $\mathbf{e}_{it}^{(s)}(\theta)$ in A.3–A.4 can be reduced to continuous transformations (in θ) of random draws from a $\mathcal{N}(0,1)$ distribution. The continuity of $Y_i^{*(s)}(\theta)$ then follows from that of ν_θ and i_θ (cf. (26)–(27)). However, $Y_i^{*(s)}(\theta)$ cannot be simulated on a computer exactly as in A.1–A.7. The reason is that the value function, ν_θ, and

[14] Our approach is related to – but distinctively different from – the GII method of Bruins et al. (2015). While we utilize the smoothing properties of the conditional expectation operator, Bruins et al. replace simulated discrete choices with smooth functions of underlying latent variables that determine the discrete outcomes. Our method is simpler as it does not rely on smoothing functions.

the investment function, i_θ, must be approximated. This is handled by solving a dynamic optimization problem on a discretized state space $\mathcal{K} \times \mathcal{A}$ (see Online Appendix B, part II). Although the resulting discretized solution, ν_θ^D and i_θ^D, could be extended to any (K, π) by choosing a neighboring grid point on $\mathcal{K} \times \mathcal{A}$, ν_θ^D and i_θ^D would still be discontinuous in θ.

Our solution is the following: Define $\nu_\theta^{\widetilde{D}}$ as the bilinear interpolation of ν_θ^D on the grid squares $[K_{(i)}, K_{(i+1)}] \times [\pi_{(j)}, \pi_{(j+1)}]$[15]. For any given θ, $\nu_\theta^{\widetilde{D}}(K, \pi)$ is then piecewise linear along any line parallel with a coordinate axis in $\mathbb{R}^+ \times \mathbb{R}^+$ and piecewise quadratic along any other straight line, with kinks at the border of the grid squares. Moreover, as shown in Online Appendix B, part III, $\nu_\theta^{\widetilde{D}}(K, \pi)$ is continuous in θ for any (K, π). The investment function i_θ is dealt with in a similar way. In Online Appendix B, part III, we show that by replacing ν_θ and i_θ with $\nu_\theta^{\widetilde{D}}$ and $i_\theta^{\widetilde{D}}$ in the algorithm A.1–A.7, we obtain a continuous simulated trajectory $Y_i^{*(s)}(\theta)$.

6.2 Properties of the indirect inference estimator

For any vector x and weighting matrix Ω, let $||x||_\Omega \equiv x'\Omega x$. Let $\widehat{\theta^a}_N$ denote $\widehat{\theta^a}$ as a function of the sample size, N, and $\widehat{\theta}_{NS}$ denote $\widehat{\theta}$ as a function of N and S, where S is chosen to keep the estimation uncertainty from the simulations (i.e., the Monte Carlo standard error) below a chosen tolerance level. Our indirect inference estimator, $\widehat{\theta}_{NS}$, is the solution to (43) with $\Omega = (\widehat{I}_N)^{-1}$, where $\widehat{I}_N$ is a consistent estimator of

$$I = Var_{\theta^0}\left(\frac{\partial l(\theta^{a*}|Y_i)}{\partial \theta^a}\right) \tag{49}$$

and I^{-1} is the optimal weighting matrix under the assumptions of Gourieroux et al. (1993). As is well known, the existence of a one-to-one relation between the parameters θ^a and θ implies that the weighting matrix in (43) will not affect the asymptotic distribution of $\widehat{\theta}_{NS}$.

Let $Y_i^{*(s)}(\theta)$ be generated as in the algorithm A.1–A.7, i.e., using the exact solution ν_θ of the functional equation (27). To obtain standard errors of the indirect inference estimator based on $Y_i^{*(s)}(\theta)$, we utilize a property of the "Third Version of the Indirect Estimator" in Gourieroux et al. (1993, p. S110–S111). Using this property in our model, as N becomes large, we have:

[15]See Numerical Recipes Library (Press et al., 1994, p. 123).

$$\sqrt{N}(\widehat{\theta}_{NS}-\theta^0) \overset{D}{\simeq} \left[\frac{\partial\theta^{a*}}{\partial\theta^0}\right]^{-1} J^{-1} \times$$

$$\left(N^{-1/2}\sum_{i=1}^{N}\frac{\partial l(\theta^{a*}|Y_i)}{\partial\theta^a} - S^{-1}\sum_{s=1}^{S}N^{-1/2}\sum_{i=1}^{N}\frac{\partial l(\theta^{a*}|Y_i^{*(s)}(\theta^0))}{\partial\theta^a}\right) \quad (50)$$

where

$$J = -p\lim_{N\longrightarrow\infty} N^{-1}\sum_{i=1}^{N}\frac{\partial^2 l(\theta^{a*}|Y_i)}{\partial\theta^a\partial\theta^{a\ \prime}}$$

and $\partial\theta^{a*}/\partial\theta^0$ is defined in (41). It follows that

$$Var(\widehat{\theta}_{NS}) \simeq N^{-1}\left[\frac{\partial\theta^{a*}}{\partial\theta^0}\right]^{-1} J^{-1}(I+S^{-1}I^*)J^{-1}\left[\frac{\partial\theta^{a*}}{\partial\theta^0}\right]^{-1\prime} \quad (51)$$

with

$$I^* = Var_{\theta^0}\left(\frac{\partial l(\theta^{a*}|Y_i^{*(s)}(\theta^0))}{\partial\theta^a}\right).$$

To use (51) in practice, $\partial\theta^{a*}/\partial\theta^0$ can be estimated by finite differencing using (41), whereas estimates of I, I^*, and J can be obtained from sample analog estimators (in all cases θ^0 and θ^{a*} are replaced with $\widehat{\theta}_{NS}$ and $\widehat{\theta^a}_N$, respectively).

As explained at the end of the previous Subsection, to simulate $Y_i^{*(s)}(\theta)$ on a computer, ν_θ and i_θ must be replaced (in A.1–A.7) with approximate solutions, $\nu_\theta^{\widetilde{D}}$ and $i_\theta^{\widetilde{D}}$, based on a discretization of the state space combined with bilinear interpolation. The resulting $Y_i^{*(s)}(\theta)$, which is henceforth referred to as the feasible $Y_i^{*(s)}(\theta)$, is continuous in θ, but it is not differentiable because the bilinear interpolation has kink points at the border of the grid squares. Fortunately, the averaging in (44) will "smooth out" the kinks as NS increases, and the magnitude of the discretization error can be assessed by choosing increasingly finer grids. As the number of grid points increases, (50)–(51) will be applicable to the feasible $Y_i^{*(s)}(\theta)$.[16]

[16] Santos and Vigo-Aguiar (1998) show, under quite general regularity conditions, that the error of the value function approximation resulting from discretization, is of order $O(h^2)$, where h is the grid size. Tauchen (1990) assesses the discretization error in a dynamic programming model that is similar to our model: a stochastic growth model with two state variables; capital stock and technological efficiency, and one continuous decision variable; investment. He shows that, with as little as 1,800 grid points, a highly accurate approximation of the value function and the optimal investment function is obtained using a combination of discretization and local linear interpolation.

Table 2: **Interpretation of key parameters**

ϑ_1	The absolute value of the elasticity of operating surplus with respect to variable costs
ϱ	The elasticity of substitution between materials and labor is $1 - \varrho$
κ	The elasticity of operating surplus with respect to the stock of capital
s	The price of old capital relative to the price of new capital
τ	The coefficient of $\nu(S_t, 0) - \nu(S_t, 1)$ in the conditional exit probability model
φ	AR-parameter in the process of short-run profitability: $\pi_t = \mu + \varphi(\pi_{t-1} - \mu) + \zeta_t$
μ	The unconditional mean of π_t
σ	The standard deviation of the innovation term ζ_t

7 Estimation

To estimate the model, we apply a version of the conjugate direction method developed by Brent (1973, Ch. 7) in combination with the derivative-free line search algorithm from the Numerical Recipes Library (Press et al., 1994, p. 419); cf. the discussion in Subsection 6.2. The algorithm is implemented as a GAUSS program. For the results reported here, we used $S = 10$ and $N_K = N_\pi = 100$. Then quasi-likelihood estimation typically took between 45 and 60 minutes on a 64 core Linux server with a maximum clock rate of 2.5 GHz (HP BL685c G7). Indirect inference estimation, using the corresponding quasi-likelihood estimates as initial values, converged after about 24 hours.

7.1 Parameter estimates

Table 2 shows the key parameters and their interpretations. The estimated parameters are shown in Table 3, with standard errors – calculated from (51) – in parentheses. As seen from Table 3, the parameter estimate of ϑ_1 – the absolute value of the elasticity of operating surplus (Π) with respect to variable cost (c) – is roughly 0.4 for the three individual industries, as well as for total manufacturing. Thus, our results indicate that if the prices of materials and labor increase by one percent, operating surplus is reduced by around 0.4 percent. The estimates of ϱ are around 0.3. Because the elasticity of substitution between labor and materials is $1 - \varrho$, the estimates of this substitution parameter are around 0.7.

We now turn to the importance of capital. The estimates of κ – the elasticity of operating surplus (Π) with respect to the stock of capital (K) – are depicted in the fourth column of Table 3; they are all around 0.8. Hence, operating surplus increases by around 0.8 percent if the stock of capital is increased by one percent (and the variable factors of production are optimized). Our estimates of the adjustment cost parameter s lie in the range of 0.70 to 0.93, and all estimates are significantly less than one at the 5 percent level. The estimate for total manufacturing is 0.85 (with a standard error of 0.02), implying about a 15 percent "discount" in the second-hand capital market. Our results therefore indicate moderate adjustment cost of capital.

The sixth column of Table 3 shows that the net present value of continuing production relative to exiting, $\nu(S_t, 1) - \nu(S_t, 0)$, has a significant negative impact on the probability to exit: the estimated value of the parameter τ, which reflects the extent to which the producer takes profitability into account when deciding when to exit, is significant and positive in all industries. The estimates range from 0.48 in machinery to 0.67 in metal products, with an estimate of 0.63 for total manufacturing.

The last three columns in Table 3 show the results for the parameters of the AR-process of the short-run profitability measure π_t (see (25)). All estimates of the AR-coefficient φ lie around 0.9, which implies a highly persistent, but stationary, π_t-process. Our results indicate that the distribution of short-run profitability is quite similar across the investigated industries, with respect to both the mean and the dispersion.

Table 3: **Estimates of coefficients.** Standard errors in parentheses

Industry	Estimates							
	ϑ_1	ϱ	κ	s	τ	φ	μ	σ
Wood products	.42 (.13)	.32 (.09)	.80 (.05)	.70 (.05)	.52(.15)	.91 (.04)	−.99 (.11)	.11 (.03)
Metal products	.53 (.10)	.28 (.08)	.75(.04)	.91 (.03)	.67(.12)	.89 (.03)	−.88 (.08)	.17 (.03)
Machinery	.36 (.12)	.24 (.09)	.76 (.05)	.93 (.04)	.48(.16)	.88 (.05)	−.87 (.12)	.20 (.04)
Total manuf.	.44 (.04)	.29 (.03)	.80 (.02)	.85 (.02)	.63(.04)	.91 (.02)	−1.02 (.04)	.17 (.01)

Table 4: **Specification tests.** Estimates of two versions of the structural model augmented with an auxiliary parameter. Total manufacturing.

Model specification	Parameter estimates			Partial quasi	ROC-
	β	τ	s	log-likelihood	value
Basic model	0 (–)	.63 (.04)	.85 (.02)	-7575	.78
Basic model augmented with auxiliary parameter (β)	-.03 (.02)	.26 (.03)	1 (–)	-7580	.78
Model with fixed scrap value	0 (–)	.14 (.02)	.55 (.05)	-7756	.73
Model with fixed scrap value and auxiliary parameter (β)	-.21 (.03)	.28 (.02)	1 (–)	-7574	.78

7.2 The gross rate of return to capital

In this section, we report the estimated gross rate of return to capital, using different measures for this variable. First, we introduce a theoretical rate of return, defined as the rate of return in steady state. To this end, we first define the steady-state stock of capital as $K^*(\pi) = (1-\delta)K^*(\pi) + i(K^*, \pi)$, where $i(K^*, \pi)$ is the optimal investment rule (see Section 6.1). Thus, if the firm is in steady state, that is, $(K_{t-1}, \pi_t) = (K^*(\pi_t), \pi_t)$, it will not change its stock of capital from $t-1$ to t. Steady state is a theoretically useful concept as it reduces the state space from two dimensions to one by replacing an arbitrary K_{t-1} with $K^*(\pi_t)$.

The steady-state gross rate of return to capital is:

$$GR(\pi_t) \equiv E\left(\frac{\Pi_{t+1}}{K^*(\pi_t)} |\pi_t\right) = E(e^{\pi_{t+1}} K^*(\pi_t)^{\kappa-1} |\pi_t) \tag{52}$$

where we have used that $\Pi_{t+1}/K_t = e^{\pi_{t+1}} K_t^{\kappa-1}$, see (8), in the second equation.

We now compare the steady-state rate of return to capital to other measures of the rate of return to capital. First, in neoclassical theory, the average rate of return to capital is independent of the stock of capital (in steady state), and it is equal to $(r+\delta)/\kappa$.[17] Our estimate of the neoclassical gross rate of return to capital is about 24 percent.

In panel a in Figure 1, we show estimates of three alternative measures of the gross rate of return to capital in total manufacturing. The horizontal axis in panel a shows π_t measured in standard deviations of the mean μ. Hence, a firm with average short-run profitability is represented by the number 0 on the horizontal axis. The estimated steady-state gross rate of return for a firm with average short-run profitability is 25.5 percent, see the curve referred to as "rate of return in steady state" in panel a. The small difference between this estimate (25.5 percent) and the estimate of the neoclassical gross rate of return to capital (24 percent) reflects that cost of adjustment, measured by the parameter s, is moderate; the estimate of s is 0.85 for total manufacturing, which is not much lower than 1 (the neoclassical case).

Second, by averaging the observed gross rate of return over all firms, we obtain a rate of return slightly below 24 percent for total manufacturing – see the line referred to as

[17] Without adjustment cost, it can be shown that $K^*(\pi_t)^{\kappa-1} = E(e^{\pi_{t+1}}|\pi_t)^{-1}(r+\delta)/\kappa$. In steady state, the *realized* gross rate of return is therefore $\Pi_{t+1}/K^*(\pi_t) = e^{\pi_{t+1}} E(e^{\pi_{t+1}}|\pi_t)^{-1}(r+\delta)/\kappa$. The steady-state gross rate of return, given π_t, is then $(r+\delta)/\kappa$, which is independent of π_t.

"empirical average" in panel a in Figure 1. Finally, the dotted horizontal line in panel a shows the unconditional expected rate of return to capital, defined as $E(GR(\pi_t))$ (see (52)), where the unconditional expectation is evaluated using the stationary (invariant) distribution of π_t (see (25)). For total manufacturing, the unconditional expected gross rate of return to capital is around 25.5 percent, that is, almost identical to the estimated rate of return to capital in steady state for a firm with average short-run profitability.

[Figure 1 here]

7.3 The probability to exit

We now turn to the steady-state relationship between short-run profitability and the exit probability; that is, the probability that a firm will exit during the next year. This relationship is shown as a solid curve in panel b in Figure 1 for total manufacturing. Similar to panel a, the horizontal axis shows π_t measured in standard deviations of the mean μ. The solid curve shows that when a firm is in steady state, that is, $K_t = K^*(\pi_t)$, (hypothetical) higher short-run profitability reduces the exit probability (when the stock of capital is immediately adjusted to the new steady-state level). A firm with average short-run profitability ($\pi_t = \mu$), and equipped with the corresponding steady-state stock of capital, has an exit probability of around three percent. A decrease by two standard deviations from the average short-run profitability raises the exit probability by almost one percentage point.

To illustrate how the exit probability depends on the stock of capital, in panel b we have included curves that show the exit probability when the stock of capital is fixed at 'low' and 'high' levels. The two fixed values are the 10th and 90th percentiles in the empirical distribution of capital in total manufacturing. As seen in panel b, the exit curve when the firm has a fixed 'high' stock of capital is below the exit curve when the firm has a fixed 'low' stock of capital. This is in line with a number of empirical studies that found that the larger the firm, the lower the probability to exit; some examples include Olley and Pakes (1996), Disney et al. (2003), and Foster et al. (2008). Note that the distance between the two curves reflects the estimated s (the price of old capital relative to that of new capital); if, hypothetically, $s = 1$ (the neoclassical case), there should be no difference between the two curves.

The horizontal line in panel b represents the expected exit probability. This number has been calculated analogously to $E(GR(\pi_t))$ (see (52)); that is, by taking the expectation of the steady-state exit probability with respect to the unconditional distribution of π_t. As seen in panel b, the expected exit probability is close to the exit probability of the firm with average short-run profitability and a corresponding steady-state stock of capital.

So far we have illustrated how the estimated exit probability depends on short-run profitability and the stock of capital. We now discuss how well our model predicts exits. To this end it is expedient to construct an ROC curve for total manufacturing (see panel c in Figure 1).[18] This is handled as follows. First, we use the estimated model to calculate the probability that a firm will exit during the next year. Next, we choose a threshold probability p and divide firms into two groups; those having a probability of at least p to exit, and those having a lower probability. Among firms with a probability to exit of at least p, the share that did exit the next year is termed the true positive rate, whereas the share of firms that did not exit is termed the false positive rate. For each p, the two observed exit rates represent one point on the ROC curve. By increasing p from 0 to 1, we construct the entire ROC curve.

The dotted line in panel c corresponds to the case in which the false positive rate always equals the true positive rate, in which case the model is worthless. The success of the model to predict exit is measured by the area under the ROC curve. As a rule of thumb, if the area exceeds 0.9, the test is regarded as excellent. For total manufacturing, the area under the ROC curve is slightly below 0.80.

To learn more about the ability of the model to predict exit, we now examine how the probability to exit varies between exiting firms and firms that did exit in the sample period; the latter group is termed non-exiting firms. An exiting firm is termed a closing-down firm in its last year of operation. For each closing-down firm, we calculate the estimated probability to exit during the next year i) 1 year prior to observed exit ("closing-down firms"); ii) 2 years prior to observed exit ("closing-down firms, lagged 1 year"); and iii) 4 years prior to observed exit ("closing-down firms, lagged 3 years"). Assume a firm exited in

[18] ROC (Receiver Operating Characteristic) curves are common in medicine to assess how well a decision rule, using clinical results, predicts a disease; for example, see van Erkel and Pattynama (1998). To measure the degree of predictability of the test, the area under the ROC curve is calculated; it is 1 if the test is perfect and 0.5 if the test is worthless.

2001; that is, its last full year of operation was 2000. We then calculate i) the probability in 2000 that this firm would exit in 2001 ("closing-down firms"), ii) the probability in 1999 that this firm would exit in 2000 ("closing-down firms, lagged 1 year"), and iii) the probability in 1997 that this firm would exit in 1998 ("closing-down firms, lagged 3 years"). These three probabilities are assigned to the year 2000; that is, the firm's last year of operation. This procedure is undertaken for all exiting firms. Finally, for each year and for each of the three types of exit probability, we calculate the unweighted average over firms, see panel d in Figure 1.

By construction, for each year t, the three graphs for closing-down firms contain the same firms. Hence, we can compare how the (annual) exit probability of firms evolves over time as firms approach their observed year of exit. As seen from panel d, the exit probability increases somewhat over time as firms get closer to their observed exit year; that is, in most years, the graph for the closing-down firms is above the graph for the closing-down firms lagged 1 year, which is above the graph for the closing-down firms lagged 3 years.[19] However, there is no sharp increase in the exit probability in the last year prior to exit. On the other hand, the three graphs for the closing-down firms lie above the graph for the non-exiting firms – the latter graph shows, for each year t, the average (annual) exit probability of operative firms that did not exit in the sample period. Hence, our model discriminates between closing-down firms and firms that did not exit.

Our results suggest that the main characteristic of an exiting firm is not that its annual exit probability is much higher than that of a non-exiting firm, but rather that the difference in annual exit probabilities is highly persistent. Therefore, it is the cumulated effect of higher annual exit probabilities over many years – compared with the average firm – that causes a firm to exit. In economic terms, if a firm over a long period of time has a low expected value of continuation relative to the expected value of exit, then the firm has a high probability to exit.

7.4 Capital adjustment cost

In our model, we find moderate cost of capital adjustment; the estimates of the parameter s (the resale price of capital relative to the price of new capital) are in the range of 0.70 to

[19] The increase in the exit probabilities for firms exiting after 2008 reflects the financial crises from 2008 to 2010.

0.93 (see Section 7.1). In the literature, there has been much discussion on whether there is significant cost of capital adjustment. A key contribution in the empirical literature is Cooper and Haltiwanger (2006), who use GMM to estimate structural parameters of capital adjustment costs. They specify four types of capital adjustment costs: i) strictly convex cost of adjustment; ii) a cost of adjustment that is proportional to the stock of capital; iii) a specific percentage drop in plant productivity triggered by investment; and iv) a wedge between the acquisition and resale price of capital. Using a balanced panel of approximately 7000 large US manufacturing plants that were continuously in operation between 1972 and 1988, they estimate models with different combinations of these adjustment cost types. In general, they report significant results. In the case of a specific percentage drop in plant productivity triggered by investment, that is, type iii) of capital adjustment cost, their estimate is 20 percent, which is close to what Caballero and Engel (1999) obtained for a similar specification (16 percent). For the specification which is most similar to our model – cost of adjustment of type iv) – Cooper and Haltiwanger (2006) found that the price of old capital relative to that of new capital (s) is 80 percent, which is similar to our estimate for total manufacturing (0.85), see Table 3.

Also Hall (2004) estimates adjustment cost for capital. His approach relies on estimating Euler equations for factor demand, using US annual data for factor inputs for two-digit industries for the period 1948-2001. In contrast to Cooper and Haltiwanger (2006), Caballero and Engel (1999), and us, Hall (2004) finds small capital adjustment cost.

Above, we estimated how the gross rate of return to capital, and the exit probability, depend on short-run profitability (π_t) when firms have the steady-state stock of capital, $K^*(\pi_t)$. We now test the robustness of these relationships with respect to the parameter s (the resale price of capital relative to the price of new capital). To this end, we solve the model when, *cet. par.*, s takes the value 0 or 1. The first case ($s = 0$) represents a situation in which the resale value of acquired capital is zero. The second is the neoclassical case ($s = 1$); the acquisition price of a unit of capital equals the resale price. In the latter case, the price of capital is independent of vintage.

Panel a in Figure 2 shows the relationship between short-run profitability and the steady-state stock of capital for three alternative values of s (0, 0.85, and 1) for total manufacturing (s was estimated to 0.85 for total manufacturing, see above). As seen from the figure, for all cases the steady-state stock of capital is increasing in short-run profitability. Roughly, for a given level of short-run profitability, the higher s, the higher the steady-state stock of capital. However, the differences are small; for very low levels of short-run profitability, there is de facto no difference.

[Figure 2 here]

We next examine how alternative values of s influence the following three relationships when firms have their steady-state stock of capital: i) the relationship between short-run profitability and the gross rate of return to capital; ii) the relationship between short-run profitability and the net present value of continuing production rather than exiting; and iii) the relationship between short-run profitability and the exit probability.

Panel b in Figure 2 shows the relationship between short-run profitability and the gross rate of return to capital (for total manufacturing). Without cost of capital adjustment ($s = 1$), we obtain the standard neoclassical result: the rate of return to capital is independent of short-run profitability. In contrast, with cost of capital adjustment ($s < 1$), the rate of return to capital is increasing in short-run profitability. Moreover, for any given level of short-run profitability, the rate of return to capital is higher when there is complete irreversibility ($s = 0$) than when there is partial irreversibility ($s = 0.85$). This ranking might, however, be ambiguous because more capital has two counteracting effects on the gross rate of return to capital (Π_{t+1}/K_t): more capital (K) has a direct negative effect (the denominator increases), but it has also an indirect positive effect because operating surplus (Π) increases (the elasticity of operating surplus with respect to capital is κ). Our result – higher return to capital for $s = 0$ than for $s = 0.85$ – reflects that the steady-state stock of capital is higher when $s = 0.85$ than when $s = 0$, and this (direct) effect dominates.

Panel c in Figure 2 shows the relationship between short-run profitability and the difference between the present value of continuing production and the present value to exit; henceforth, we term this difference the incremental value of continuing production. For all depicted values of s, the incremental value is increasing in short-run profitability. Further, for any level of short-run profitability exceeding the mean μ, the incremental value is highest under complete irreversibility ($s = 0$). This reflects that for $s = 0$ (old capital has no value), the value of exit does not increase in the stock of capital.

Finally, panel d in Figure 2 shows the relationship between short-run profitability and the probability to exit. Exit probabilities are decreasing in short-run profitability, and for any level of short-run profitability exceeding $\mu = 0$, the exit probability is lowest under complete irreversibility ($s = 0$). This result simply reflects that the incremental value of continuing operation is highest under complete irreversibility (see discussion above).

7.5 Alternative specification of the model

In our model, the producer compares the value of continuing production, $\nu(S_t, 1)$, see (22), to the (scrap) value he receives if the firm exits; $\nu(S_t, 0) = -c(-(1-\delta)K_{t-1})$ (see (21)). As discussed in Section 1, many papers in the literature rely on the simplifying assumption that the scrap value is exogenous; for example, see Olley and Pakes (1996). We now examine the consequences of this alternative assumption within our framework.

In the alternative model, $\nu(S_t, 0) = \Phi$, i.e., the scrap value is constant.[20] Below, we refer to this model as the "restricted structural model." We want to compare the basic structural model to the restricted one. We do so by i) running a specification test for both models, and ii) comparing their goodness-of-fit properties. Because the models are non-nested, we do not test one model against the other in the formal sense of statistical hypothesis testing.

Regardless of whether we consider our basic model or the alternative model with fixed scrap value, the conditional exit probability is influenced by short-run profitability and the stock of capital, and it is monotonically increasing in $\nu(S_t, 0) - \nu(S_t, 1)$. In particular, the partial effect of changing capital on the exit probability is captured by the shifts shown in panel b in Figure 1. If, however, either of the models is misspecified, we expect that including K_{t-1} as a separate control variable for the exit probability would (at least in large samples) lead to a rejection of the hypothesis that the coefficient of K_{t-1} is zero. Hence, we estimate the augmented exit probability model

$$p^{\text{Aug}}(S_t) = \frac{1}{1 + \exp\left\{-\left[\tau\left(\nu(S_t, 0) - \nu(S_t, 1)\right) + \beta K_{t-1} + \xi_0\right]\right\}},$$

[20]Note that a random term in the scrap value is taken into account through the error term $\varepsilon(0)$ in (16).

and test the hypothesis $\beta = 0$. If the structural model is correctly specified, see (23), the indirect inference estimator of β will tend to be zero with probability one.[21]

The results of the specification test for total manufacturing are shown in Table 4. We display the estimates for the auxiliary parameter β and the parameters τ and s in the partial quasi log-likelihood function (36). In addition, the table shows the value of the partial quasi log-likelihood function and the corresponding ROC value (i.e., the area under the ROC curve). The results are quite striking: the estimate of β in our *basic model* (augmented with the auxiliary parameter) is -0.03, with a standard error of 0.02. This gives a t-ratio of -1.5, which is not significant at conventional levels.[22] The ROC values of the basic model and the model augmented with the auxiliary parameter β are equal.

In stark contrast, the structural model with *fixed scrap value* does not capture the effect of capital on the exit probability in an adequate manner. The estimate of β is negative (-0.21), and with a t-ratio of -7, the restriction $\beta = 0$ is clearly rejected. Moreover, the model with fixed scrap value has a partial quasi log-likelihood value of about 200 points, and an ROC value of 0.05 points, lower than the basic model. We also note that the estimate of s is significantly lower in the model with fixed scrap value (0.55) than in the basic structural model (0.85). To sum up, the results in Table 4 provide substantial evidence that the model with fixed scrap value fails the specification test, and that it fits the data significantly worse than our basic structural model.

As a final specification test, we consider the limiting case when the coefficient τ approaches infinity in (22) and (23). Define

$$G_{\tau,\xi_0}(\nu(S_{t+1},0),\nu(S_{t+1},1)) = \frac{1}{\tau}\ln\left[\exp(\tau\nu(S_{t+1},0)+\xi_0)+\exp(\tau\nu(S_{t+1},1))\right]. \quad (53)$$

In the limit

$$\lim_{\tau\longrightarrow\infty} G_{\tau,\xi_0}(\nu(S_{t+1},0),\nu(S_{t+1},1)) = \max(\nu(S_{t+1},0),\nu(S_{t+1},1)). \quad (54)$$

Inserting (53) and (54) in (22), and assuming $\nu(S_{t+1},0) = \Phi$, gives a specification which is identical to that in Olley and Pakes (1996). An implication of this specification is that the exit decision becomes deterministic: the firm will exit in t if and only if $\Phi > \nu(S_t,1)$.

[21] This specification test is analogous to the one used by Rust (1987) to test a conditional independence assumption in a dynamic discrete-choice model.

[22] Although the partial quasi log-likelihood value is 5 points higher in the augmented model, the usual likelihood-ratio test does not apply here.

However, the deterministic case is clearly rejected by our estimate of τ; the exit decision is influenced by a highly important noise component, which makes the timing of exit hard to predict, given our state variables.

8 Conclusion

In the Introduction we raised three questions: "Is there a relationship between firm exit and profitability?", "What causes firms to exit?", and "What are the characteristics that distinguish exiting firms from non-exiting firms?" Using a structural econometric model we have derived explanatory variables from economic theory and estimated the model for start-up firms in three Norwegian manufacturing industries separately, as well as for total manufacturing.

We find that when exit is defined as a state in which production at the site has come to a permanent stop, increased profitability significantly lowers the exit probability; or, put differently, low profitability causes firms to exit. We have also found a clear difference in the estimated exit probabilities between firms that exited in the sample period (1994–2012) and those that did not exit. According to our results, exiting firms differ from non-exiting firms as their annual exit probabilities are *persistently* higher. Put differently: if, over a long period of time, the expected value of continuing production is low relative to the expected value of exit, the firm has a high probability to exit. We also find that exiting firms are not characterized by having a very high exit probability just prior to exit, which reflects that there are no (negative) profitability shocks in the last years prior to exit. However, the estimated probability to exit during the next year tends to increase as the firm approaches its final year of operation.

Our data suggest a rather weak relationship between profitability and exit (see the discussion in Section 1). Yet, our estimation results clearly indicate such a relationship. We believe this shows the power of econometric modeling and methods. For example, in our theory model cost of capital adjustment, which is captured by the price of old capital relative to that of new capital, is a key factor in determining exit. Although this type of cost is not included in our data, we find, through estimating a number of structural parameters by indirect inference, that for total manufacturing, the price of old capital relative to that of new capital is significantly lower than 1. Hence, cost of capital adjustment matters for the exit decision.

We have estimated a structural model where firms' exit and investment decisions are the solution to a discrete-continuous dynamic programming problem. This model has been tested against a simpler model with exogenous scrap value, which is a common assumption in empirical papers, see, for example, Olley and Pakes (1996). The result suggests that the model with fixed scrap value fits the data significantly worse than our theory-consistent econometric model.

In our model, identification is facilitated by designing an auxiliary model that has the same parameters as the structural model. In general, it may be difficult to identify parameters in dynamic structural models, see, for example, Magnac and Thesmar (2002) and Collard–Wexler (2013). Aguirregabiria and Suzuki (2014) examine identification in a dynamic structural model with fixed entry cost, fixed cost of operation, and fixed scrap value. They argue that identification requires that one of the three fixed costs must be set to zero. This restriction has implications for the interpretation of the estimates. If the restriction is specified as setting the entry cost equal to zero, the fixed cost of operation in their model is equal to the actual fixed cost of operation plus the difference between current entry cost and expected entry cost in the next period, discounted to the current period. The implications of omitting the entry decision, as we do, is that estimates are *contingent* on firms that have entered the market. This should not be mixed with the implications of including the entry decision in the model and imposing that there is no cost of entry.

Acknowledgments

This paper has benefited from numerous comments and suggestions, in particular from Anders Rygh Swensen and three anonymous referees. We also thank Daniel Bergsvik, Erik Biørn, Bernt Bratsberg, John K. Dagsvik, Torbjørn Hægeland, Jos van Ommeren, Knut Røed, Terje Skjerpen and Steinar Strøm. Earlier versions of the paper have been presented at the University of Oslo, the Norwegian School of Management, and the European Economic Association meeting in Toulouse 2014. We thank the participants for their comments. This research has been financially supported by The Norwegian Research Council through the MER Entreprenørskap program.

References

[1] Abel, A and J. Eberly, 1994, A Unified Model of Investment under Uncertainty. American Economic Review 84, 1369–1384.

[2] Abel, A. and J. Eberly, 1996, Optimal investment with costly reversibility. Review of Economic Studies 63, 581–593.

[3] Aguirregabiria, V. and C.-Y. Ho, 2012, A dynamic oligopoly game of the US airline industry: estimation and policy experiments. Journal of Econometrics 168(1), 156–173.

[4] Aguirregabiria, V. and J. Suzuki, 2014, Identification and counterfactuals in dynamic models of market entry and exit. Quantitative Marketing and Economics 12, 267–304.

[5] An, Y. M. and M. Liu, 2000, Using indirect inference to solve the initial conditions problem. The Review of Economics and Statistics 82(4), 656–667.

[6] Andersen, T.G. and J. Lund, 1997, Estimating continuous time stochastic volatility models of the short term interest rate. Journal of Econometrics 77, 343–378.

[7] Andersen, T.G., H.J. Chung and B.E. Sørensen, 1999, Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. Journal of Econometrics 91, 61–87.

[8] Asphjell, M.K., W. Letterie, Ø.A. Nilsen and G.A. Pfann, 2014, Sequentiality Versus Simultaneity: Interrelated Factor Demand. Review of Economics and Statistics 96, 986–998.

[9] Bansal, R., A.R. Gallent and G. Tauchen, 2007, Rational Pessimism, Rational Exuberance, and Asset Pricing Models. Review of Economic Studies 74, 1005–1033.

[10] Bloom, N., S. Bond and J. van Reenen, 2007, Uncertainty and Investment Dynamics. Review of Economic Studies 74, 391–415.

[11] Brent, R.P., 1973, Algorithms for minimization without derivatives. Prentice-Hall, Englewood Cliffs.

[12] Bruins, M., J.A. Duffy, M.P. Keane and A.A. Smith, 2015, Generalized indirect inference for discrete choice models. Economic working paper from Nuffield college, University of Oxford, 2015-W08.

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

[14] Collard-Wexler, A., 2013, Demand fluctuations in the ready-mix concrete industry. Econometrica 81(3), 1003–1037.

[15] Cooper, R.W. and J.C. Haltiwanger, 2006, On the nature of capital adjustment costs. Review of Economic Studies 73, 611–633.

[16] Disney, R., J. Haskel and Y. Heden, 2003, Entry, exit and establishment survival in UK manufacturing. Journal of Industrial Economics LI (1), 91–112.

[17] Dixit, A.K. and R.S. Pindyck, 1994, Investment under uncertainty. University Press, Princeton, NJ.

[18] Duffee, G. and R. Stanton, 2008, Evidence on simulation inference for near unit-root processes with implications for term structure estimation. Journal of Financial Econometrics 6, 108–42.

[19] Dunne, T., M.J. Roberts and L. Samuelson, 1988, Pattern of firm entry and exit in U.S. manufacturing industries. RAND Journal of Economics 19(4), 495–515.

[20] Dunne, T., S.D. Klimek, M.J. Roberts and D.Y. Xu, 2013, Entry, exit and the determinants of market structure. RAND Journal of Economics 44(3), 462–487.

[21] Foster, L., J. Haltiwanger and C. Syverson, 2008, Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? American Economic Review 98(1), 394–425.

[22] Fuleky, P. and E. Zivot, 2014, Indirect inference based on the score. The Econometrics Journal 17, 383–393.

[23] Gallant A.R. and G. Tauchen, 1996, Which moments to match? Econometric Theory 12, 657–681.

[24] Gourieroux C., A. Monfort and E. Renault, 1993, Indirect inference. Journal of Applied Econometrics 8, Supplement, S85–S118.

[25] Gourieroux, C. and A. Monfort, 1996, Simulation-Based Econometric Methods. Oxford University Press, Oxford.

[26] Gourieroux, C., E. Renault and N. Touzi, 2000, Calibration by simulation for small sample bias correction, in: R. Mariano, T. Schuermann and M.J. Weeks, (Eds.), Simulation-based Inference in Econometrics: Methods and Applications. Cambridge University Press, Cambridge.

[27] Hall, R. E., 2004, Measuring factor adjustment costs. Quarterly Journal of Economics 119(3), 899–927.

[28] Koopman, S. J and N. Shepard, 1992, Exact score for time series models in state space form. Biometrika 79, 823–826.

[29] Levinsohn, J. and A. Petrin, 2003, Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70, 317–342.

[30] Li, T. and B. Zhang, 2015, Affiliation and entry in first-price auctions with heterogeneous bidders: An analysis of merger effects. American Economic Journal: Microeconomics 7(2), 188–214.

[31] Magnac, T., J.-M. Robin and M. Visser, 1995, Analysing incomplete labor market histories using indirect inference. Journal of Applied Econometrics 10, S153–S169.

[32] Magnac, T. and D. Thesmar, 2002, Identifying dynamic discrete decision processes. Econometrica 70(2), 801–816.

[33] Michaelides, A. and S. Ng, 2000, Estimating the rational expectations model of speculative storage: a Monte Carlo comparison of three simulation estimators. Journal of Econometrics 96, 231–266.

[34] Milgrom, P and I. Segal, 2002, Envelope theorems for arbitrary choice sets. Econometrica 70(2), 583–601.

[35] Nagypál, E., 2007, Learning by doing vs. learning about match quality: Can we tell them apart? The Review of Economics Studies 74, 537–566.

[36] Nilsen, Ø. A, A. Raknerud, M. Rybalka and T. Skjerpen, 2011, The Importance of Skill Measurement for Growth Accounting. Review of Income and Wealth 57, 293–305.

[37] Olley, G. S. and A. Pakes, 1996, The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1297.

[38] Pakes, A., M. Ostrovsky and S. Berry, 2007, Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). RAND Journal of Economics 38(2), 373–399.

[39] Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, 1994, Numerical recipes in C. Cambridge University Press, Cambridge.

[40] Raknerud, A., D. Rønningen and T. Skjerpen, 2004, Documentation of the Capital database. A database with data for tangible fixed assets and other economic data at the firm level. Documents 2004/16, Statistics Norway. http://www.ssb.no/english/subjects/10/90/doc_200416_en/doc_200416_en.pdf

[41] Raknerud, A and Ø. Skare, 2012, Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes. Computational Statistics and Data Analysis 56, 3260–3275.

[42] Rudin, W., 1976, Principles of Mathematical Analysis. McGraw-Hill, Singapore.

[43] Rust, J., 1987, Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. Econometrica 55, 999–1033.

[44] Rust, J., 1994, Structural estimation of Markov decision processes, in: R.F. Engle and D.L. McFadden *(*Eds.*)* Handbook of Econometrics, Volume IV. Elsevier, Amsterdam.

[45] Ryan, S. P., 2012, The costs of environmental regulation in a concentrated industry. Econometrica 80(3), 1019–1061.

[46] Santos, M.S and J. Vigo-Aguiar, 1998, Analysis of a numerical dynamic programming algorithm applied to economic models, Econometrica, 66(2), 409–426.

[47] Shumway, R.H. and D.S. Stoffer, 2000, Time Series Analysis and Its Applications. Springer, New York.

[48] Smith, A.A., 1993, Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8(S1), S63–S84.

[49] Stokey, N.L., 2009, The Economics of Inaction. Stochastic Control Models with fixed costs. Princeton University Press. Princeton.

[50] Tauchen, G., 1986, Finite state Markov-chain approximations to univariate and vector autoregressions. Economics Letters 2, 177–181.

[51] Tauchen, G., 1990, Solving the Stochastic Growth Model by Using Quadrature Methods and Value-Function Iterations. Journal of Business and Economic Statistics 8, 49–51.

[52] van Erkel, A. R. and P. M. T. Pattynama, 1998, Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology. European Journal of Radiology 27, 88–94.

Appendix

Part I: Proof of Proposition 2

By definition, $Y_i^{(s)}(\theta)$ is identical to $Y_i^{*(s)}(\theta)$ except that $\widehat{z}_i^{(s)}(\theta)$ is replaced with $z_i^{(s)}(\theta) = (z_{i2}^{(s)}(\theta), ..., z_{i,T+1}^{(s)}(\theta))$. From Proposition 1:

$$E_\theta(z_{it}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta), z_{i,t-1}^{(s)}(\theta)) = \left(1 - p_\theta(K_{i,t-1}^{(s)}(\theta), \pi_{it}^{(s)}(\theta))\right) z_{i,t-1}^{(s)}(\theta)$$

(recall that $z_{i,t-1}^{(s)}(\theta) = 0$ is an absorbing state). By the rule of double expectation:

$$\begin{aligned} E_\theta(z_{it}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta)) &= \left(1 - p_\theta(K_{i,t-1}^{(s)}(\theta), \pi_{it}^{(s)}(\theta))\right) E_\theta(z_{i,t-1}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta)) \\ &= \left(1 - p_\theta(K_{i,t-1}^{(s)}(\theta), \pi_{it}^{(s)}(\theta))\right) E_\theta(z_{i,t-1}^{(s)}(\theta)|W_{i,1:t-1}^{(s)}(\theta)) \end{aligned} \quad (55)$$

where the last equality follows from conditional independence of $z_{i,t-1}^{(s)}(\theta)$ and $W_{it}^{(s)}(\theta)$ given $W_{i,1:t-1}^{(s)}(\theta)$. Since $E_\theta(z_{it}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta))$ and $\widehat{z}_{it}^{(s)}(\theta)$ satisfy the same difference equation (compare A.5 and (55)), (47) follows from the initial condition $\widehat{z}_{i1}^{(s)}(\theta) = z_{i1}^{(s)}(\theta) = 1$.

To prove (48), define $\widehat{W}_{i,1:t}^{(s)}(\theta) = \{W_{ik}^{(s)}(\theta), \widehat{\pi}_{ik}^{(s)}(\theta)_{k\wedge T}\}_{k=1}^t$. Then

$$\begin{aligned} & E_\theta\left[\frac{\partial l(\theta^a|Y_i^{(s)}(\theta))}{\partial\theta_3^a} - \frac{\partial l(\theta^a|Y_i^{*(s)}(\theta))}{\partial\theta_3^a}\right] \\ = & \sum_{t=1}^{T} E_\theta\left[E_\theta\left(z_{i,t+1}^{(s)}(\theta) - \widehat{z}_{i,t+1}^{(s)}(\theta)\middle|\widehat{W}_{i,1:t+1}^{(s)}(\theta)\right)\frac{\partial\ln\left(1 - p_{(\theta_1^a,\theta_2^a,\theta_3^a)}(K_{it}^{(s)}(\theta), \widehat{\pi}_{i,t+1\wedge T}^{(s)}(\theta))\right)}{\partial\theta_3^a}\right. \\ & \left. + E_\theta\left(z_{it}^{(s)}(\theta) - \widehat{z}_{it}^{(s)}(\theta) - (z_{i,t+1}^{(s)}(\theta) - \widehat{z}_{i,t+1}^{(s)}(\theta))\middle|\widehat{W}_{i,1:t+1}^{(s)}(\theta)\right)\frac{\partial\ln p_{(\theta_1^a,\theta_2^a,\theta_3^a)}(K_{it}^{(s)}(\theta), \widehat{\pi}_{it}^{(s)}(\theta))}{\partial\theta_3^a}\right] \\ = & \; 0 \end{aligned} \quad (56)$$

where the last equality follows from

$$\begin{aligned} E_\theta(z_{i,t+1}^{(s)}(\theta)|\widehat{W}_{i,1:t+1}^{(s)}(\theta)) &= E_\theta(z_{i,t+1}^{(s)}(\theta)|W_{i,1:t+1}^{(s)}(\theta)) = \widehat{z}_{i,t+1}^{(s)}(\theta) \\ E_\theta(z_{it}^{(s)}(\theta)|\widehat{W}_{i,1:t+1}^{(s)}(\theta)) &= E_\theta\left(z_{it}^{(s)}(\theta)|W_{i,1:t}^{(s)}(\theta)\right) = \widehat{z}_{it}^{(s)}(\theta). \end{aligned}$$

Equation (56) trivially extends to the other components of the score vector. Hence,

$$E_\theta\left[\frac{\partial l(\theta^a|Y_i^{*(s)}(\theta))}{\partial\theta^a}\right] = E_\theta\left[\frac{\partial l(\theta^a|Y_i^{(s)}(\theta))}{\partial\theta^a}\right] = E_\theta\left[\frac{\partial l(\theta^a|Y_i)}{\partial\theta^a}\right] = b(\theta, \theta^a)$$

and (48) follows.

Part II: Illustration of Proposition 2

We consider the simplified model

$$z_{i,t+1} = 1(\theta y_i + \varepsilon_{it} > 0) z_{it}$$

where $1(\cdot)$ is the indicator function, θ is the unknown structural parameter, $\varepsilon_{it} \sim \mathcal{N}(0,1)$ and y_i has a *known* covariate distribution. The conditional probability of exit (given y_i and $z_{it} = 1$) is $p_\theta(y_i) = \Phi(-\theta y_i)$. The auxiliary probability is assumed to be $p_{\theta^a}(y_i) = \theta^a$. Thus, the auxiliary model replaces the firm-varying exit probabilities with a fixed auxiliary parameter, θ^a. Let $Y_i = (y_i, z_{i2}, ..., z_{i,\tau_i+1})$ be the realized data. Then the quasi log-likelihood is

$$\sum_{i=1}^{N} l(\theta^a|Y_i) = \sum_{i=1}^{N}\sum_{t=1}^{T} \left[z_{i,t+1} \ln(1-\theta^a) + (z_{it} - z_{i,t+1}) \ln(\theta^a) \right]$$

($z_{i1} = 1$), and the quasi-likelihood estimator is

$$\widehat{\theta^a} = \sum_{i=1}^{N}\sum_{t=1}^{T} (z_{it} - z_{i,t+1}) / \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}.$$

For notational simplicity, we will hereafter assume that $T = \infty$ (each firm is observed until it exits). Because $E_\theta \left(\sum_{t=1}^{\infty} (z_{it} - z_{i,t+1}) | y_i \right) = \sum_{t=1}^{\infty} (1 - \Phi(-\theta y_i))^{t-1} \Phi(-\theta y_i) = 1$ and $E_\theta \left(\sum_{t=1}^{\infty} z_{it} | y_i \right) = 1/\Phi(-\theta y_i)$, we have

$$\widehat{\theta^a} \to_P \frac{1}{E_{\theta^0}(1/\Phi(-\theta^0 y_i))} \equiv \theta^{a*}$$

– the pseudo-true parameter. The binding function is (see (39)):

$$b(\theta, \theta^a) \equiv E_\theta \left(\frac{\partial l(\theta^a|Y_i)}{\partial \theta^a} \right) = -\frac{E_\theta(1/\Phi(-\theta y_i)) - 1}{1 - \theta^a} + \frac{1}{\theta^a}$$

(which satisfies $b(\theta^0, \theta^{a*}) = 0$). The simulated trajectory is: $Y_i^{*(s)}(\theta) = (y_i^{(s)}, \widehat{z}_{i2}^{(s)}(\theta), \widehat{z}_{i3}^{(s)}(\theta), ...)$, where $\widehat{z}_{it}^{(s)}(\theta) = (1 - \Phi(-\theta y_i^{(s)})) \widehat{z}_{i,t-1}^{(s)}(\theta) = \Phi(\theta y_i^{(s)}) \widehat{z}_{i,t-1}^{(s)}(\theta)$ (with $\widehat{z}_{i1}^{(s)}(\theta) = 1$). Hence, $\widehat{z}_{it}^{(s)}(\theta) = \Phi(\theta y_i^{(s)})^{t-1} = E_\theta(z_{it}^{(s)}(\theta)|y_i^{(s)})$ for $t = 1, 2, ...$. This confirms (47). Furthermore,

$$\sum_{t=1}^{\infty} \widehat{z}_{i,t+1}^{(s)}(\theta) = 1/\Phi(-\theta y_i^{(s)}) - 1 \text{ and } \sum_{t=1}^{\infty} (\widehat{z}_{it}^{(s)}(\theta) - \widehat{z}_{i,t+1}^{(s)}(\theta)) = 1.$$

Thus

$$E_\theta \left(\frac{\partial l(\theta^a|Y_i^{*(s)}(\theta))}{\partial \theta^a} \right) = -\frac{E_\theta(1/\Phi(-\theta y_i^{(s)})) - 1}{1 - \theta^a} + \frac{1}{\theta^a}$$

which is equal to the binding function $b(\theta, \theta^a)$, since $y_i^{(s)}$ has the same distribution as y_i. This confirms (48).

Online Appendix: Supplementary materials

Supplementary figures and proofs can be found online at https://doi.org/ ....

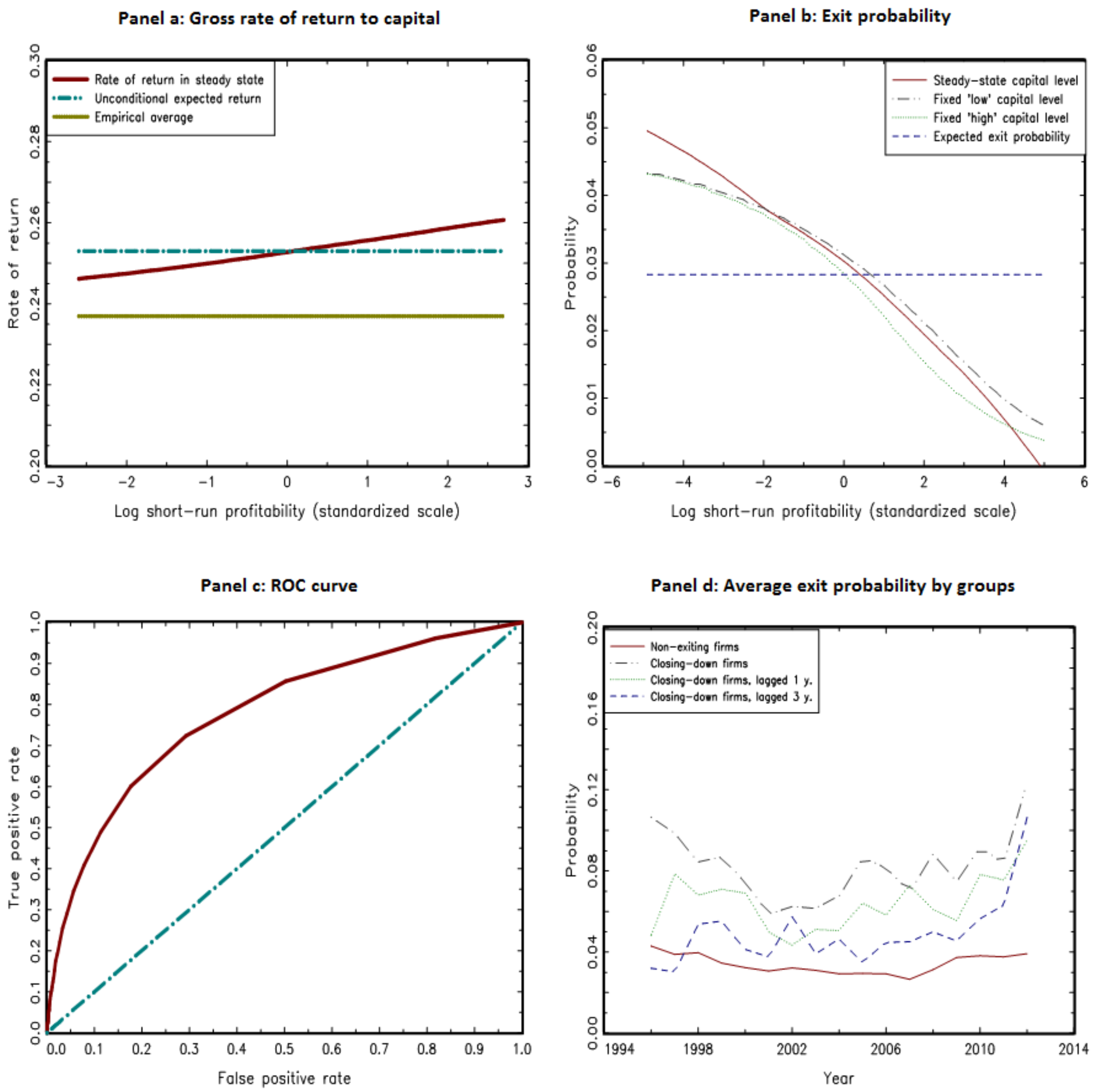

Figure 1: Estimates for total manufacturing: gross rate of return to capital vs. short-run profitability (Panel a), exit probability vs. short-run profitability (Panel b), ROC curve (Panel c), and average exit probability by groups (Panel d)

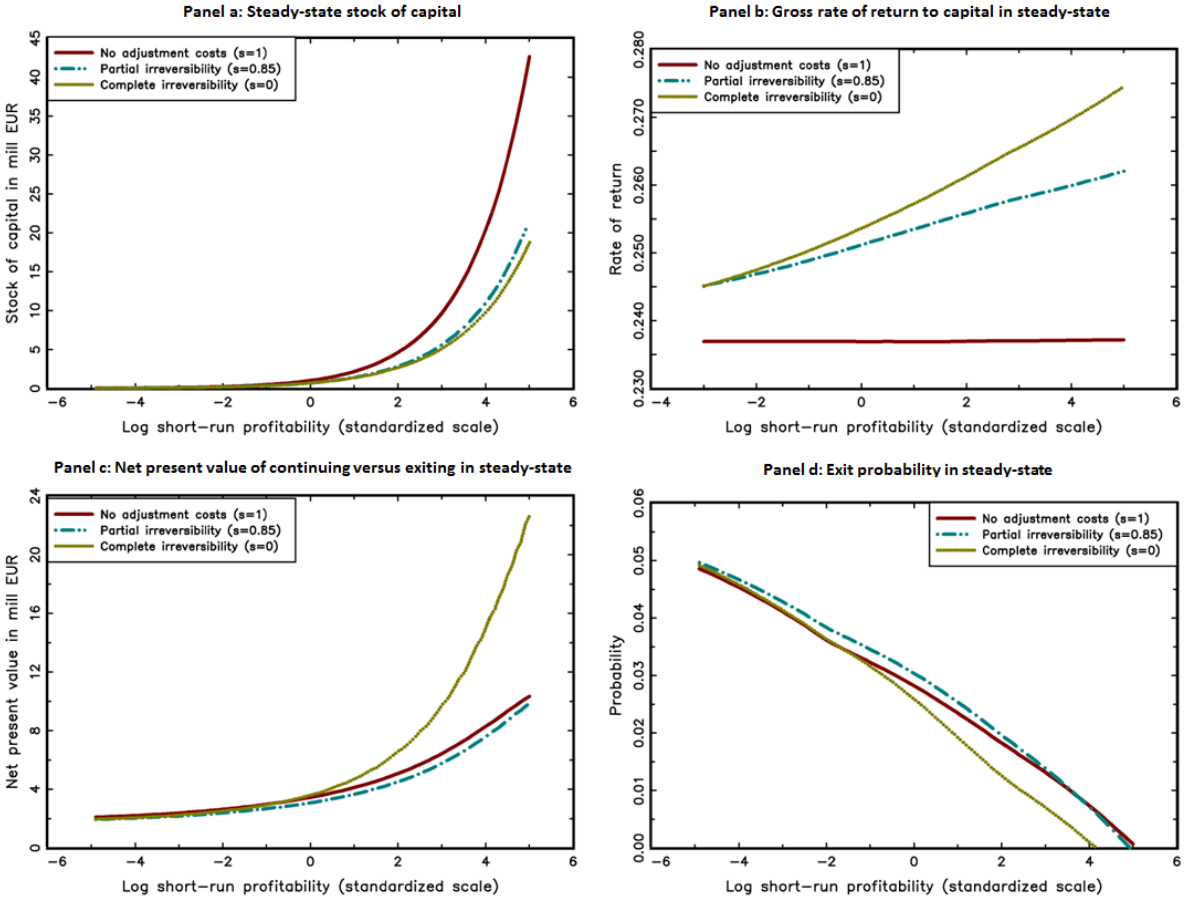

Figure 2: Estimated effects of shift in the cost of capital adjustment parameter (s). Total manufacturing

Exit dynamics of start-up firms: Structural estimation using indirect inference

Online Appendix: Supplementary materials

by

Rolf Golombek[a)] and Arvid Raknerud[b)]

a) The Ragnar Frisch Center for Economic Research, University of Oslo. E-mail: rolf.golombek@frisch.uio.no

b) Corresponding author: Statistics Norway, Research Department. E-mail: rak@ssb.no

Online Appendix A: Supplementary figures

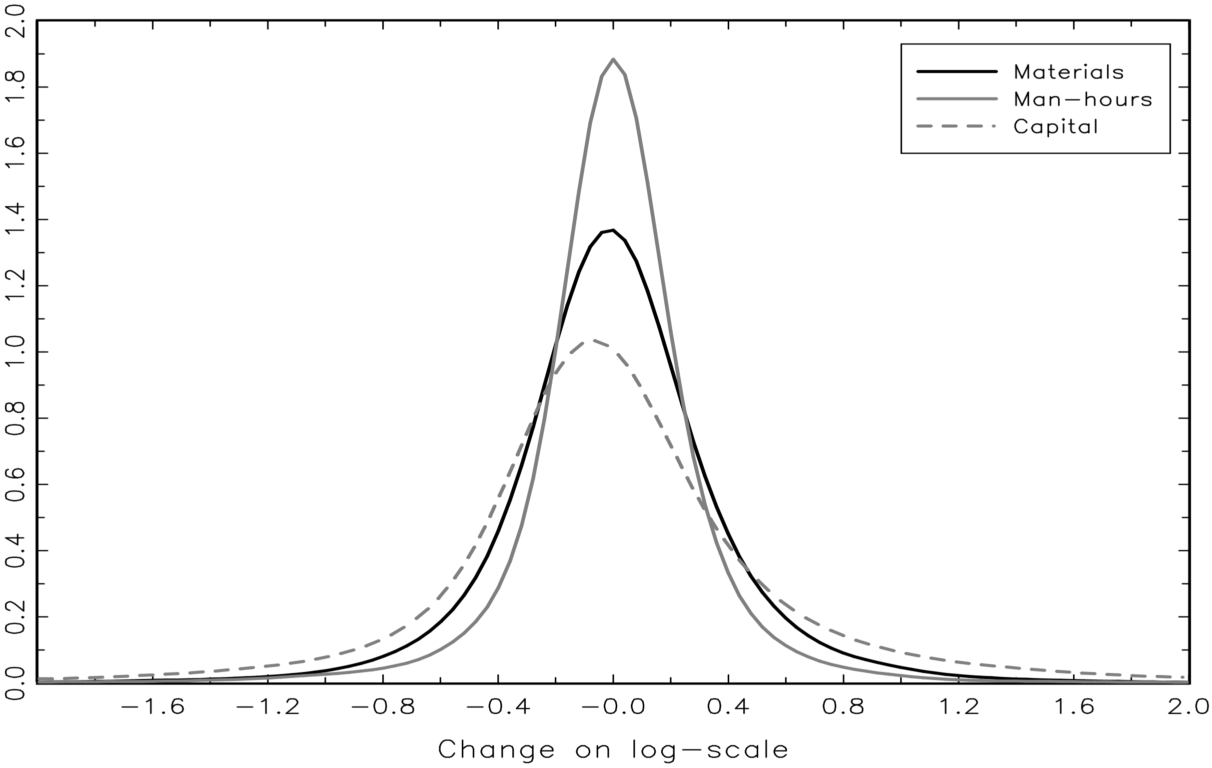

Figure A.1: Distribution of log of annual changes in capital, person-hours and materials. Kernel density estimates. Total manufacturing, 1994–2012

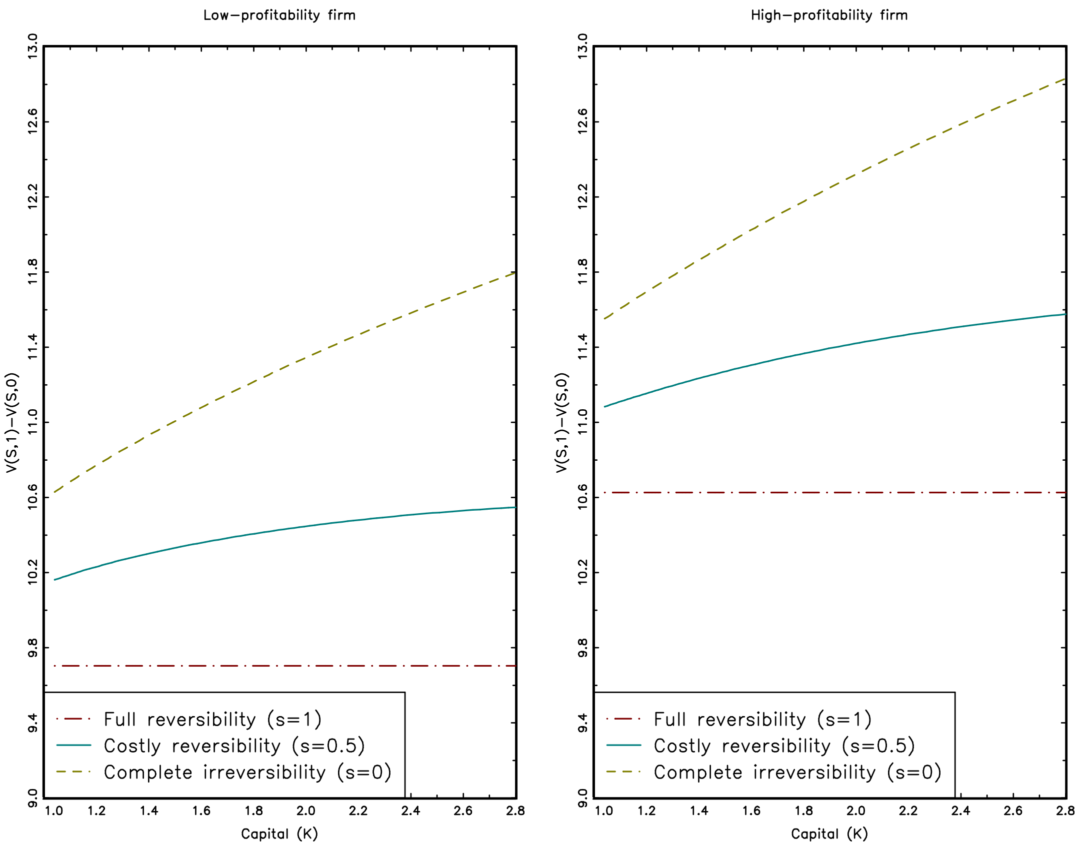

Figure A.2: The graphs show typical solutions of the difference $v(S_t, 1) - v(S_t, 0)$ (the net value of continuing production) as a function of K_{t-1} for different values of s ($s = 0, 0.5, 1$) and π_t ("low profitability" and "high profitability").

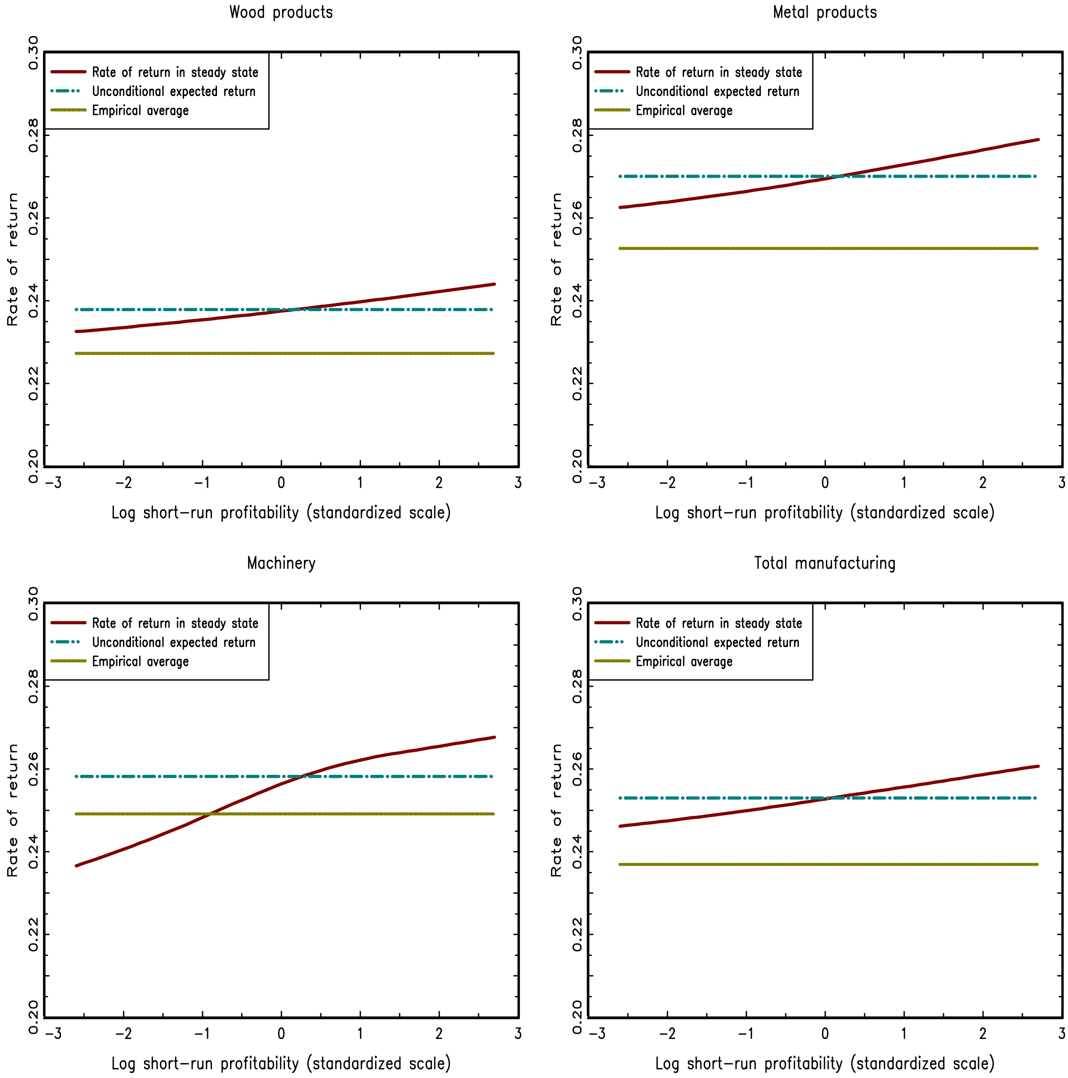

Figure A.3: Estimated gross rate of return to capital, 1994-2012

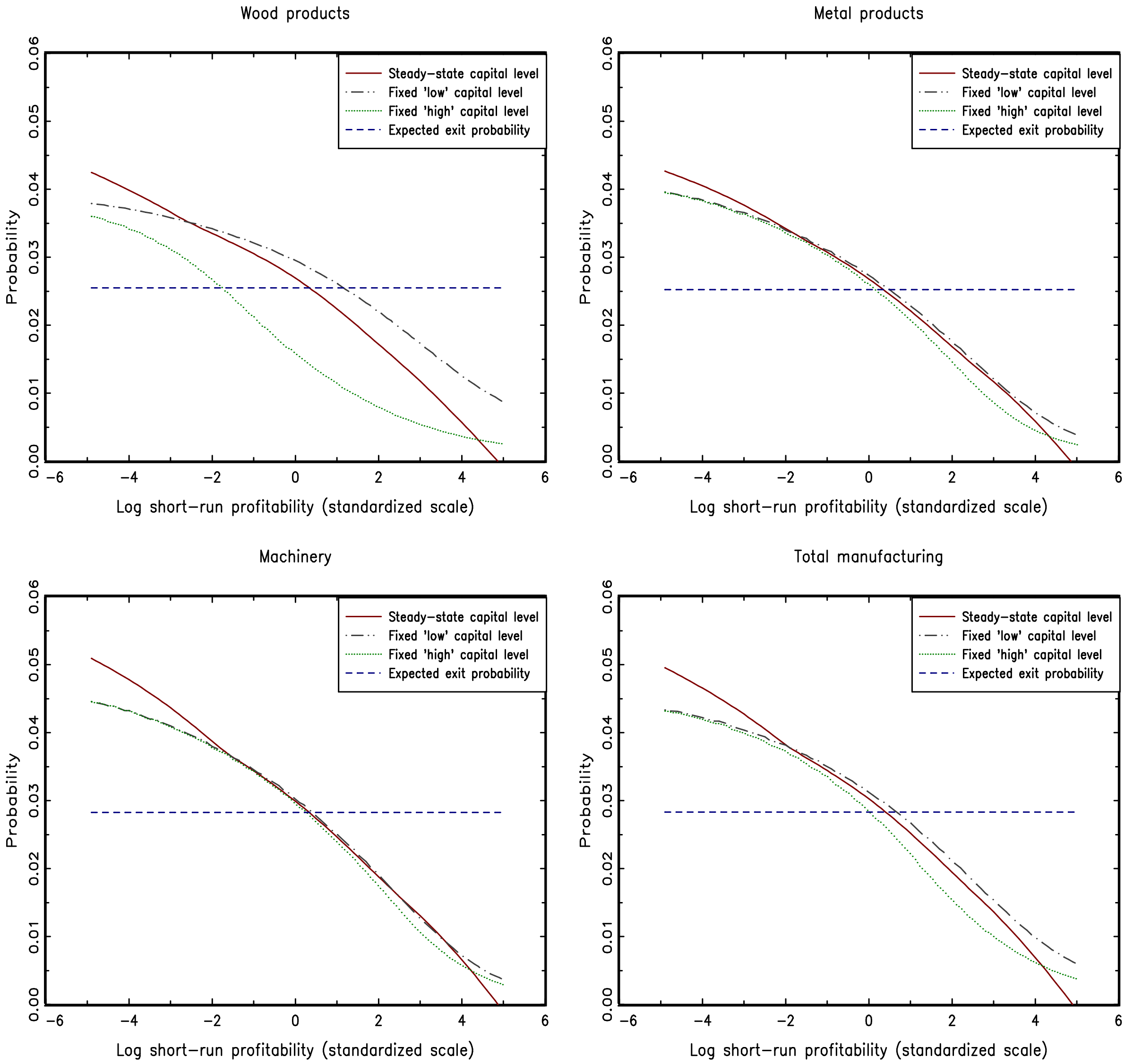

Figure A.4: Estimated exit probability, 1994-2012

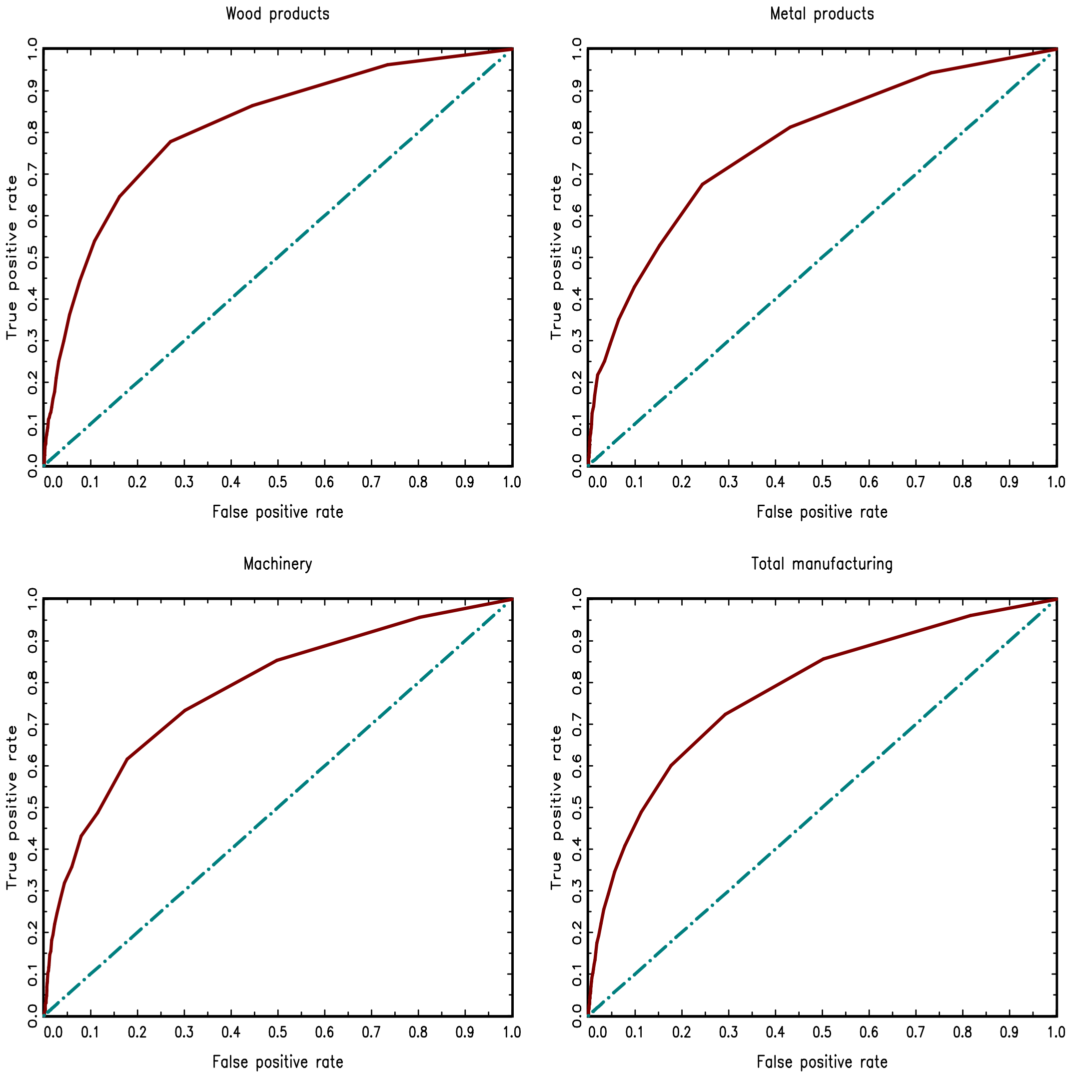

Figure A.5: ROC curves, 1994-2012

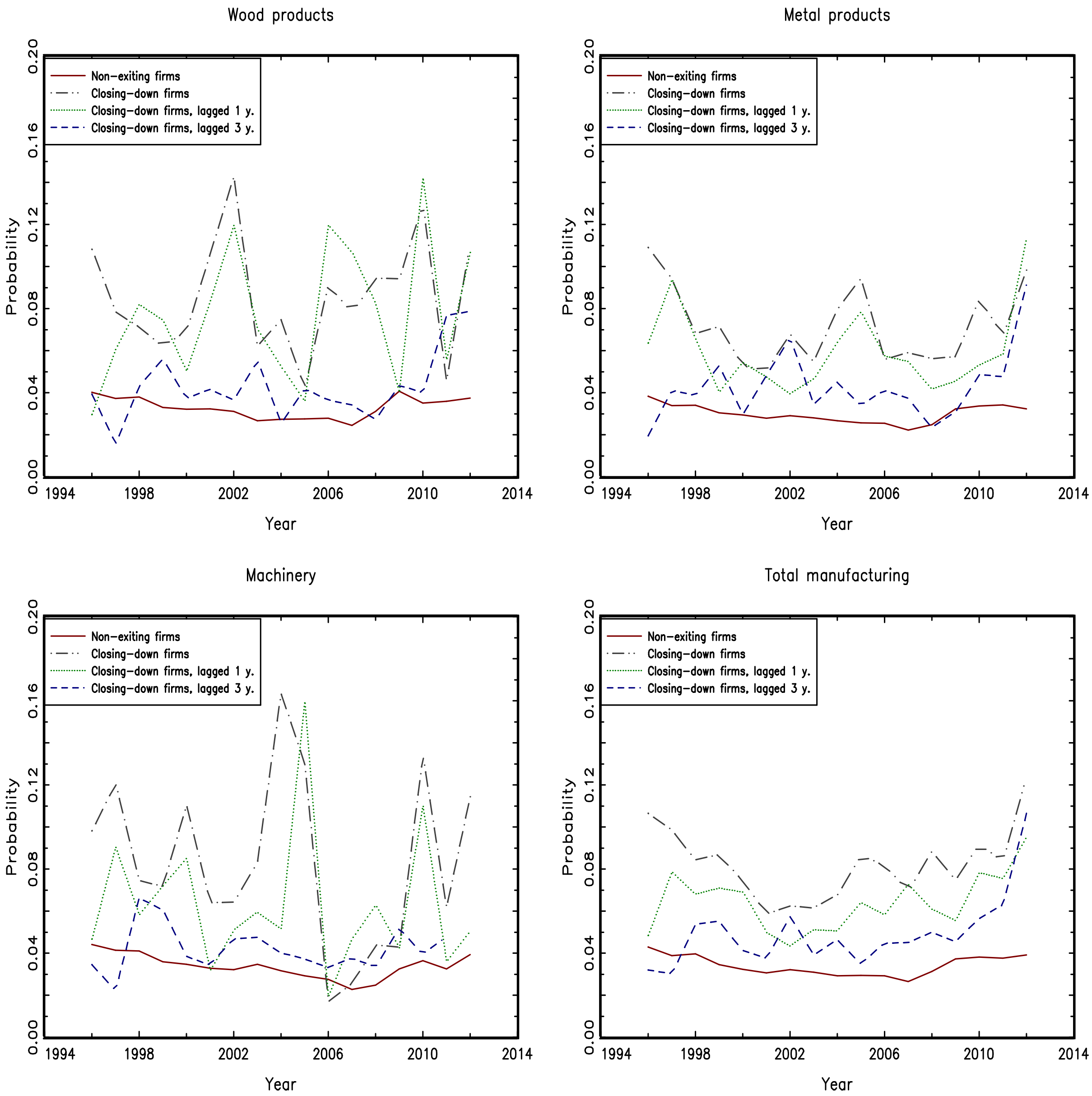

Figure A.6: Estimated average exit probability by groups, 1995–2012

Online Appendix B: Proofs

Note: The references Equation (20), Equation (21), etc. below refer to the main article, whereas (B.1), (B.2), etc. refer to the Online Appendix.

Part I: Proof of Proposition 1

Let $\nu(S_t, z_t)$ denote the net present value given S_t and z_t:

$$\nu(S_t, z_t) = \max_{I_t} \left\{ u(S_t, I_t, z_t) - \Pi_t + \frac{1}{1+r} E\left[V(S_{t+1}, \varepsilon_{t+1}) | S_t, I_t, z_t\right] \right\}. \tag{B.1}$$

Then Equation (20) follows by definition. If $z_t = 0$, t is the terminal period and $\nu(S_t, 0) = u(S_t, -(1-\delta)K_{t-1}, 1) - \Pi_t = -c(-(1-\delta)K_{t-1})$, which proves Equation (21).

To prove Equation (22), consider a T–period horizon problem where $V_T(S_t, \varepsilon_t)$ and $\nu_T(S_t, z_t)$ denote value functions for this problem:

$$V_T(S_t, \varepsilon_t) = \max_{z_t} [\Pi_t + \nu_T(S_t, z_t) + \varepsilon_t(z_t)]$$

(cf. Equation (20)). Obviously, $\nu_T(S_t, 1) = \nu(S_t, 0)$ for all T. Given the initial condition $\nu_1(S_t, 1) = \nu_1(S_t, 0)$ (in the one-period problem there is no distinction between "exiting" and "continuing", we obtain, for $T = 1, 2, 3, ...$:

$$\begin{aligned}
\nu_{T+1}(S_t, 1) &= \max_{I_t} \left\{ -c(I_t) + \frac{1}{1+r} E\left[V_T(S_{t+1}, \varepsilon_{t+1}) | S_t, I_t, z_t = 1\right] \right\} \\
&= \max_{I_t} \left\{ -c(I_t) + \frac{1}{1+r} \int \max\left\{\Pi_{t+1} + \nu_T(S_{t+1}, 0) + \varepsilon_{t+1}(0), \Pi_{t+1} + \nu_T(S_{t+1}, 1) + \varepsilon_{t+1}(1)\right\} \right. \\
&\quad \left. \times h(\varepsilon_{t+1}) d\varepsilon_{t+1} g(dS_{t+1} | S_t, I_t) \right\} \\
&= \max_{I_t} \left\{ -c(I_t) + \frac{1}{1+r} \int \left[\Pi_{t+1} + \frac{1}{\tau} \ln\left[\exp(-\tau c(-(1-\delta)K_t) + \xi_0) + \exp(\tau \nu_T(S_{t+1}, 1))\right] \right] \right. \\
&\quad \left. \times g(dS_{t+1} | S_t, I_t) \right\}
\end{aligned} \tag{B.2}$$

where the integrand after the last equality is the so-called "social surplus" function. The last equality follows from Equation (17) and a well-known property of the extreme value distribution (see Rust, 1994). Under the regularity conditions of Rust (1994), (B.2) defines a contraction mapping so that

$$\sup_S |\nu_{T+1}(S_t, 1) - \nu_T(S_t, 1)| \longrightarrow 0 \text{ as } T \longrightarrow \infty.$$

Then there exists a limiting function $\nu(S_t, 1)$ that satisfies Equation (22). Finally, from Equation (20):

$$\begin{aligned}
\Pr(z_t &= 0|S_t, z_{t-1} = 1) = \Pr\left(\nu(S_t, 0) + \varepsilon_t(0) > \nu(S_t, 1) + \varepsilon_t(1)|S_t\right) \\
&= \Pr\left(\tau\nu(S_t, 0) + \xi_0 + (\tau\varepsilon_t(0) - \xi_0) > \tau\nu(S_t, 1) + \tau\varepsilon_t(1)|S_t\right) \\
&= \frac{1}{1 + \exp\{-[\tau\nu(S_t, 0) - \tau\nu(S_t, 1) + \xi_0]\}}
\end{aligned}$$

where we in the last equation used that $\tau\varepsilon_t(z) - \xi_z$ has a standard extreme value distribution and is independent for $z = 0, 1$. Using Equation (21), Equation (23) now follows.

Q.E.D.

Part II: The method of successive approximations

Define

$$\widetilde{\Psi}_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}, I) = -c_s(I) + \frac{1}{(1+r)} \sum_{j'} \left[\frac{\exp(\pi_{(j')})[(1-\delta)K_{(i)} + I]^\kappa}{1 + \vartheta_1} + \frac{1}{\tau} \ln\left[\exp(\tau s(1-\delta)[(1-\delta)K_{(i)} + I] + \xi_0) + \exp(\tau\nu_\theta^{(n)}((1-\delta)K_{(i)} + I, \pi_{(j')}))\right]\right] g^D_{(\varphi,\mu,\sigma)}(\pi_{(j')}|\pi_{(j)}) \tag{B.3}$$

Given a value function $\nu_\theta^{(n)}(K, \pi)$ defined on the discrete state space $\mathcal{K} \times \mathcal{A}$, define the corresponding optimal investment function:

$$i_\theta^{(n)}(K_{(i)}, \pi_{(j)}) = \arg \max_{\{I:(1-\delta)K_{(i)} + I \in \mathcal{K}\}} \widetilde{\Psi}_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}, I) \tag{B.4}$$

Next, define

$$\Psi_\theta^{(n)}(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}) = -c_s(i^{(n)}(K_{(i)}, \pi_{(j)})) + \frac{1}{1+r} \times \sum_{j'} \left[\frac{\exp(\pi_{(j')})K_{(i')}{}^\kappa}{1 + \vartheta_1} + \frac{1}{\tau} \ln\left[\exp(\tau s(1-\delta)K_{(i')} + \xi_0) + \exp(\tau\nu_\theta^{(n)}(K_{(i')}, \pi_{(j')}))\right]\right] g^D_{(\varphi,\mu,\sigma)}(\pi_{(j')}|\pi_{(j)})$$

with $K_{(i')} = (1-\delta)K_{(i)} + i_\theta^{(n)}(K_{(i)}, \pi_{(j)})$. The method of successive approximations generates a sequence $\{\nu_\theta^{(n)}\}$ as follows: Let $\nu_\theta^{(1)} = 0$, and set

$$\nu_\theta^{(n+1)} = \Psi_\theta^{(n)}(\nu_\theta^{(n)}) \tag{B.5}$$

until convergence: $\nu_\theta^{(n+1)} \longrightarrow \nu_\theta^D$. The performance of the solution algorithm can be improved if we choose $\nu_\theta^{(n+1)}$ as the solution to the fixed-point problem $\nu_\theta^{(n+1)} = \Psi_\theta^{(n)}(\nu_\theta^{(n+1)})$

(instead of the simple updating (B.5)). The latter is most efficiently done by means of Newton-Kantorovich iterations (see Iskhakov et al., 2016).

Part III: Continuity issues

Define

$$\Psi_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}) = \max_{\{I:(1-\delta)K_{(i)}+I\in\mathcal{K}\}} \widetilde{\Psi}_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}, I)$$

(cf. (B.3)). The method of successive approximations (see Part II) is then equivalent to generating:

$$\nu_\theta^{(n+1)}(K_{(i)}, \pi_{(j)}) = \Psi_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}) \tag{B.6}$$

with $\nu_\theta^{(1)} = 0$. Assume that $\nu_\theta^{(n)}(K_{(i)}, \pi_{(j)})$ is continuous in θ for given $(K_{(i)}, \pi_{(j)})$. Then $\nu_\theta^{(n+1)}(K_{(i)}, \pi_{(j)})$ is also continuous in θ, being the max of a finite number of functions $\widetilde{\Psi}_\theta(\nu_\theta^{(n)})(K_{(i)}, \pi_{(j)}, I)$ that are continuous in θ.

We shall first establish the continuity of $\nu_\theta^D(K_{(i)}, \pi_{(j)})$ on a compact subset Θ of $\mathbb{R}^K$. Define

$$M = \sup_{\theta\in\Theta} ||\Psi_\theta(\nu_\theta^{(1)})|| < \infty$$

where $||\cdot||$ is the supremum norm. From the contraction mapping property of (B.6) (see Equation (2.16) in Rust, 1994):

$$\begin{aligned}\sup_{\theta\in\Theta} ||\nu_\theta^{(n+1)} - \nu_\theta^{(n)}|| &\leq \frac{1}{1+r}\sup_{\theta\in\Theta} ||\nu_\theta^{(n)} - \nu_\theta^{(n-1)}|| \\ &\leq \frac{1}{(1+r)^{n-1}} M.\end{aligned}$$

Since $\sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} M = \frac{1+r}{r} M < \infty$, it follows from Theorem 7.10 in Rudin (1976) that $\nu_\theta^{(n+1)}(K_{(i)}, \pi_{(j)}) = \sum_{i=1}^{n} (\nu_\theta^{(i+1)}(K_{(i)}, \pi_{(j)}) - \nu_\theta^{(i)}(K_{(i)}, \pi_{(j)}))$ converges uniformly on Θ towards $\nu_\theta^D(K_{(i)}, \pi_{(j)})$. Moreover, since $\nu_\theta^{(n)}(K_{(i)}, \pi_{(j)})$ is continuous in θ ($n = 1, 2, 3, ...$), it follows from Theorem 7.11 (*ibid*) that the limit function $\nu_\theta^D(K_{(i)}, \pi_{(j)})$ is also continuous in θ.

Next, define $\nu_\theta^{\widetilde{D}}(K, \pi)$ as the bilinear interpolation of $\nu_\theta^D(K_{(i)}, \pi_{(j)})$ on the grid squares $[K_{(i)}, K_{(i+1)}] \times [\pi_{(j)}, \pi_{(j+1)}]$ and

$$i_\theta^{\widetilde{D}}(K_{(i)}, \pi_{(j)}) = \arg \max_{\{I:I\geq-(1-\delta)K_{(i)}\}} \widetilde{\Psi}_\theta(\nu_\theta^{\widetilde{D}})(K_{(i)}, \pi_{(j)}, I)$$

(cf. (B.4)). For given $(K_{(i)}, \pi_{(j)})$, the maximand $\widetilde{\Psi}_\theta(\nu_\theta^{\widetilde{D}})(K_{(i)}, \pi_{(j)}, I)$ is continuous in θ and I, with a unique argmax $i_\theta^{\widetilde{D}}(K_{(i)}, \pi_{(j)})$. It follows that $i_\theta^{\widetilde{D}}(K_{(i)}, \pi_{(j)})$ is continuous in θ. We extend $i_\theta^{\widetilde{D}}(K_{(i)}, \pi_{(j)})$ to any (K, π) by bilinear interpolation.

Finally, if we in A.1–A.7 (in the main article) replace ν_θ and i_θ with $\nu_\theta^{\widetilde{D}}$ and $i_\theta^{\widetilde{D}}$, then the continuity of $K_{it}^{(s)}(\theta)$ follows recursively from A.6 (with initial condition $K_{i0}^{(s)}(\theta) = K_{i0}$), whereas the continuity of $\widehat{z}_{it}^{(s)}(\theta)$ follows from A.5 and Equation (28): since $K_{i,t-1}^{(s)}(\theta)$ and $\pi_{it}^{(s)}(\theta)$ are continuous in θ, and $\nu_\theta^{\widetilde{D}}(K, \pi)$ is continuous in both θ and (K, π), it follows that $\nu_\theta^{\widetilde{D}}(K_{i,t-1}^{(s)}(\theta), \pi_{it}^{(s)}(\theta))$ is continuous in θ.

References

[1] Iskhakov, F., L. Jinhyuk, J. Rust, B. Schjerning and K. Seo, 2016, Constrained optimization approaches to estimation of structural models: Comment. Econometrica 84(1), 365-370.